

Fuzzy Sets and Systems 98 (1998) 343-354



A random sets-based method for identifying fuzzy models

Luciano Sánchez

E.T.S. Ingenieros Industriales de Gijón, Area de Lenguajes y Sistemas Informáticos, Carretera de Castiello Bernueces, Campus de Viesques, 33204 Gijón, Asturias, Spain

Received December 1995; revised November 1996

Abstract

The nature of interference sources in signal processing is a key problem in many applied disciplines. These interferences are often modelled by random processes, although it has been shown that many models can be favourably modified when some of the uncertainty sources are treated as fuzzy experiments.

Following this spirit, the objective of this paper is to build a mathematical model which explains a set of imprecise measurements taken on a physical system. Furthermore, it is assumed that the effect of unmodelled inputs to the system can be regarded as a random process, but the imprecision in the measures is better described by means of a possibility distribution. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Control theory; Linguistic modelling; Random set; Description length

1. Introduction

In this paper we will study a new method to identify fuzzy models of discrete systems. We are interested in those systems whose dynamics can be defined by means of a mapping, which relates any physically possible combination of state and input values to the values to which the state variables would evolve after a fixed amount of time. The structure of one of these systems is shown in Fig. 1.

In practical situations, we need to estimate the mentioned mapping from a sequence of sensor readings of state and input to the system. The problem is not trivial when these readings are noisy. In that case, we need to apply statistical methods to estimate the mapping. Since it is reasonable to think that these statistical methods depend on the nature of the sources of the noise, we will distinguish between two types of noise:

- 1. The noise originated when some inputs to the system are not present in the model.
- 2. The noise originated when the readings of the sensor slightly differ from the true value of the magnitude being sensed. We will call this effect as *observation error*.

Let us suppose that we want to identify a system and we only know a sequence of *observations* of state and input values, namely

$$E = \{e_k^*\},$$

where
$$e_k^* = (x_{k+1}^*, x_k^*, u_k^*)$$
 and $k = 1, 2, ..., n$ (1)

where, adopting the usual notation, the state will be named x and the input u, the asterisk * means "observed value" and it is assumed that $e_k^* \in X \times X \times U$, where X and U are finite sets. The solution of the problem should be a set $T \subset X \times X \times U$, graph of the mapping that models the physical system (see



Fig. 1. Physical system.



Fig. 2. Graph of the model of the physical system.

Fig. 2). That is, the identification consists in relating the set of observations E with the mapping described by the set T.

Besides, there are many well-known methods that perform this operation. For instance, you can guess that the observations e_k^* are outcomes of a stationary random process ξ_k . In that case, we should statistically infer the joint probability distribution of ξ_k , $p(\xi_k = (x', x, u)) = p(x', x, u)$ where $(x', x, u) \in X \times X \times U$. Then, given a pair (x, u)we can obtain a conditional probability distribution $p_{x,u}(x') = p(x'|x,u)$. The set *T* would be $\{(f(x,u),x,u): f(x,u) = \int x' dp_{x,u}\}.$

Otherwise, you can also replace the set T by a fuzzy set \mathcal{T} , the so-called *fuzzy graph* of the mapping that defines the system. This fuzzy graph is a kind of fuzzy model of the system. If we define this fuzzy graph by means of its level cuts,

$$\mathcal{T}(e) = \max\{\alpha I_{T_x}(e)\}, \quad I_{T_x}(e) = \begin{cases} 1 & e \in T_{\alpha}, \\ 0 & \text{otherwise,} \end{cases}$$
(2)

we can interpret that this kind of fuzzy model is a family of sets $\{T_{\alpha}\}_{\alpha}$ for which the assertions

"
$$e_k$$
 is in T_{α} "

have a "degree of truth" α , where α ranges in an orderable set of values. Given a pair of fuzzy – or crisp, as a limiting case – observations \mathscr{X} and \mathscr{U} of state and input to the system, we can obtain a fuzzy restriction over the values to which the state can evolve. This restriction can be obtained by applying

$$\mathscr{X}'(x') = \bigvee_{x,u} \{ \mathscr{T}(x', x, u) \land \mathscr{U}(u) \land \mathscr{X}(x) \},$$
(3)

where \land and \lor stand for a t-norm and a t-conorm, respectively.

Many different methods for selecting the set \mathcal{T} have been proposed. Most of the times, a parametric family $\{\mathcal{T}_{\theta}\}_{\theta}$ is a priori chosen, and the identification is guided to find a model \mathcal{T}_{θ_0} that minimizes the differences between model-based predictions of state and the observed values contained in the set E; cf. [1, 7, 12–14, 17].

2. Deduction of a fuzzy model from numerical data

We stated that the main differences between the behaviour of a physical system and any mathematical model of it are mainly due to two factors: observation noise and unmodelled inputs to the system. One easy, classical treatment for this last factor is to assume that some random, null-mean noise is being added to the original input (see Fig. 3) and, thus, the tuple (x_{k+1}, x_k, u_k) becomes an instance of a random process. We will follow this treatment. Let



Fig. 3. Model of a system with unmodelled inputs.

 ξ be this random process, and let us assume that ξ is stationary.

Under this framework, observe that the observation error is the inability of answering "yes" or "no" to the occurrence of certain outcome of this random process. This new uncertainty could be described by a new probability distribution, i.e. $p(e|e^*)$, but we will generalize this treatment to a more general case, in which the observation error will be defined by means of a confidence interval. For instance, if we sense the speed of an object, and we obtain a sensor reading of 50 kph, we could say "the true speed is a normal distribution, whose mean is 50 and whose variance is σ " as well as "the speed is 49-51 kph with a 99% confidence level". The latter case is more general, because we are allowing that $p(e|e^*)$ is any distribution whose outcomes lie with high probability in the range 49-51. Our objective is to define a model which is not tuned to a specific probability distribution $p(e|e^*)$, but to a whole family of probability distributions.

3. Fuzzy model of a physical system

We will show that this last model is a fuzzy model. To accomplish this, we first define a confidence interval like the one mentioned by means of a set Z and a small number ε , such that

$$p(e_k - e_k^* \notin Z) \leqslant \varepsilon. \tag{4}$$

Observe that the same can be described by a consonant belief function with basic probability numbers $m(Z) = 1 - \varepsilon$ and $m(X \times X \times U) = \varepsilon$. It is well known that consonant belief functions over discrete spaces are possibility measures, which can be described by normalized fuzzy sets. Then, each observation can be described by the fuzzy set

$$\mu_k(e) = \begin{cases} 1 & \text{if } e - e_k^* \in Z, \\ \varepsilon & \text{otherwise.} \end{cases}$$
(5)

Now, if we replace every inexact measurement e_k^* in E by a fuzzy set μ_k , we obtain a set $\{\mu_1, \ldots, \mu_n\}$ which can be regarded as a sample from a fuzzy random variable \mathscr{Z} [5].

Lastly, let us define the operation

$$e \oplus Z = \{e + z : z \in Z\}$$

for any $e \in X \times X \times U$, $Z \subset X \times X \times U$ (6)

to simplify the notation. Observe that the fuzzy random variable \mathscr{Z} can be expressed in terms of the following family of nested random sets:

$$\mathscr{Z}_{\alpha} = \begin{cases} X \times X \times U & \text{if } \alpha \leqslant \varepsilon, \\ \xi \oplus Z & \text{otherwise.} \end{cases}$$
(7)

Each one of the non-trivial random sets can be related to a fuzzy set \mathscr{T}' with a membership that equals its one point coverage function: $\mathscr{T}'(e) = p(e \in \xi \oplus Z)$ [9]. This procedure is consistent with the probabilistic method, because when measurements are exact, $Z = \{0\}, \ \mathscr{T}' = p(\xi = e)$ and the membership of \mathscr{T}' reduces to the joint probability of state, input and future state.

Since \mathcal{T}' determines a fuzzy graph of compatible values of input, state and output of the system, it is a fuzzy model of this system. We will not use directly this model, but a set \mathcal{T}

$$\mathcal{T}(e) = \frac{\mathcal{T}'(e)}{\sup_{z \in X \times X \times U} \{\mathcal{T}'(z)\}}$$

for all $e \in X \times X \times U$ (8)

which is the set \mathcal{T}' normalized.

3.1. Fuzzy models derived from crisp partitions

There are many practical reasons, mainly related to complexity and memory requirements of computer implementations, which favour that the membership function of \mathcal{T} can be described by a small number of



Fig. 4. Relationships between probabilistic, random-set-based and fuzzy-set-based modelling.

numerical parameters. Most of the times, memberships are chosen triangular, trapezoidal or gaussian. Since one of these simple membership functions would not be flexible enough to describe complex systems, \mathcal{T} used to be built upon the composition of N fuzzy sets \mathcal{T}_i , making

$$\mathscr{T} = \bigvee_{i=1}^{N} \mathscr{T}_{i}.$$
(9)

We will also follow this schema. To construct these kind of models, we will divide the identification into smaller problems, just by partitioning the set *E* into *N* disjoint sets E_i and then building separately *N* different fuzzy models $\mathcal{T}_i = p(e \in \xi_i \oplus Z)$, related to *N* different random processes ξ_i .

To achieve this objective, we will build a classical partition Λ of E

$$A = \{E_i\}_{i \in I},$$

$$E = \bigcup_{i \in I} E_i, \quad E_i \cap E_j = \emptyset \text{ for } i \neq j, \ I = \{1, \dots, N\}$$
(10)

for which all elements in every E_i are imprecise observations of the same random variable ξ_i . Consequently, the sets $E_i \oplus Z$ can be regarded as samples of random sets

$$\Phi_i = \{\xi_i + z; z \in Z\}$$
(11)

whose coverage functions will determine the fuzzy model associated to Λ in the following

way:

$$\mathcal{M}_{A} = \bigvee_{i=1}^{N} \mathcal{T}_{i}$$
where $\mathcal{T}_{i}(e) = \frac{p(e \in \Phi_{i})}{\sup_{X \times X \times U} \{p(e \in \Phi_{i})\}}.$ (12)

Observe that this defines a family of fuzzy models and relates every member of this family of models to an classical partition of E. In fact, we will see that the selection of fuzzy model is linked to an hierarchical clustering problem.

Once we have related every fuzzy model with a classical partition, the identification will be solved when we select the partition related to the most appropriate model. A trivial method of finding it would consist in generating all possible partitions of E, testing the model related to each one of them and then choosing the best one. However, the number of partitions is too high and this solution has no practical interest. In fact, what we really need is a suboptimal algorithm that explores only a fraction of the models.

The chosen algorithm can be briefly described as follows: First, we select the partition $\Lambda_n = \{\{e_i^*\}\}_{1 \le i \le n}$, and name its associated model \mathcal{M}_n . Second, we form all possible pairs of sets in the partition and merge them, so that $\binom{n}{2}$ simpler models are built, each one of them including n-1 sets. Third, we name the best of these models \mathcal{M}_{n-1} and then we repeat the process. After $\sum_{i=0}^{n-2} \binom{n-2}{2^i}$ iterations, the set of models $\{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$ will be obtained. Finally, we select the best of these models. It lasts to define a procedure to select the best pair of sets that will be merged at each step. This will be done in the next subsection.

3.2. Quality of a model

The merging of two sets E_i always entails a decrease in the complexity of the model, and quite often also a decrease in its precision. Any definition of the quality of a fuzzy model should pay attention to the effects of these two quite different magnitudes: degree of model's fitting to data, and complexity of the model. It is obvious that a good model should balance a low complexity with a high precision. But, how can we compare these values? That is, does a common unit exist in which complexity and precision can be expressed? We think that the answer is affirmative: both magnitudes can be measured in bits.

Let $I(\mathcal{M}_A)$ be the uncertainty involved in the structure of the model \mathcal{M}_A and let $I(E | \mathcal{M}_A)$ be a measure of the deviation of the sample *E* around this model. The joint measure

$$I(\mathcal{M}_A) + I(E \mid \mathcal{M}_A) = I(\mathcal{M}_A, E)$$
(13)

will then quantify the uncertainty of the description of the sample in the context of the model \mathcal{M}_A . The terms $I(\mathcal{M}_A)$, $I(E|\mathcal{M}_A)$ and $I(\mathcal{M}_A, E)$ are usually referred to in Probabilistic Information Theory as syntactic uncertainty, semantic uncertainty and description length, respectively.

All considered models have a common structure, so its relative complexity depends only on their parameter distribution. Since we are not certain about the credibility of the values that parameters can take on, and since we know that every \mathcal{T}_i depends on a given number of these parameters, the Bernouilli indifference principle is applicable and the syntactic uncertainty could be measured by means of the Hartley entropy. For a partition Λ of N sets, this entropy equals

$$I(\mathcal{M}_A) = \log N. \tag{14}$$

Regarding $I(E | \mathcal{M}_A)$, if data are unbiased, the more precise the model is the most specific it is. To measure the nonspecificity we can follow different ways. We will choose the method suggested by Klir [6]. This method is based on the mean of Hartley's entropies of α -cuts of the fuzzy set, i.e.,

$$U(\mathcal{F}_i) = \int_0^1 \log \#((\mathcal{F}_i)_{\alpha}) \,\mathrm{d}\alpha.$$
(15)

The mean number of bits needed to measure the nonspecificity of one of the fuzzy sets in the model is

$$I(E \mid \mathcal{M}_A) = \sum_{i=1}^N \frac{\#(E_i)}{\#(E)} U(\mathcal{T}_i).$$
(16)

Consequently, the description length of the model we suggest is given by

$$I(E, \mathcal{M}_A)$$

$$= \log N + \sum_{i=1}^{N} \frac{\#(E_i)}{\#(E)} \int_0^1 \log \#((\mathscr{T}_i)_{\alpha}) \, \mathrm{d}\alpha \qquad (17)$$

and this function defines the quality of the fuzzy model related to partition Λ .

It is worth remarking that in order to choose the populations E_i and E_j to be merged, we have to take into account that the merging involves an increase of semantic uncertainty which is proportional to

$$\Delta(\mathscr{T}_i, \mathscr{T}_j) = \#(E_i \cup E_j)U(\mathscr{T}_{E_i \cup E_j})$$
$$- \#(E_i)U(\mathscr{T}_i) - \#(E_j)U(\mathscr{T}_j).$$
(18)

Since the decrease of syntactic uncertainty does not depend on this choice, we will select populations E_i and E_i minimizing (18).

It is also remarkable that estimating a fuzzy model from a sample of size *n* only requires to compute the function $\Delta(n-1)^2$ times, because determining \mathcal{M}_i from \mathcal{M}_{i-1} only requires to recalculate i-2 values of this last function.

At last, notice that the function Δ can be regarded as a similarity measure, and therefore it is clear that the algorithm shown in Fig. 5 is an hierarchical single linkage clustering. It is easy to generalize this method by using different clustering algorithms with respect to this last similarity measure and we have conducted some tests in this respect (for instance, we compared it with the entropy minimization method NIHC [16]).

$$i \leftarrow n$$

$$A_i = \{\{e_1\}, \{e_2\}, \dots, \{e_n\}\}$$
Store $I(\mathcal{M}_{A_i}, E)$
While $i > 1$ Do
Find $V, W \in A_i$
such that $\Delta(\mathcal{T}_V, \mathcal{T}_W) = \min\{\Delta(\mathcal{T}_A, \mathcal{T}_B): A, B \in A_i\}$

$$A_{i-1} = (A_i - \{V, W\}) \cup \{V \cup W\}$$
Store $I(\mathcal{M}_{A_i}, E)$
 $i \leftarrow i - 1$
End Of While
Show $\mathcal{M}_{A_{hest}} : I(\mathcal{M}_{A_{hest}}, E) = \min_{1 \le i \le n} I(\mathcal{M}_{A_i}, E)$

Fig. 5. Simplified algorithm of the modelling procedure.

There exists some problems in which the single linkage does not find the global minimum of the description length, but a local minimum. We have observed that, in these cases, the obtained model comprises more regions than the best obtainable model. We think that this is not so severe a drawback of the single linkage algorithm as it could be if it were concerned with, for instance, pattern recognition, but, in any case, the work in this problem should continue.

4. Formulation of the method for gaussian processes

The proposed method can deal with *finite* sets of values for state and input. This limitation does not limit too much the applicability of our method (all computer implementations of fuzzy models use finite precision numbers) but it is not difficult to generalize the algorithm to some continuous cases. In particular, from now on we will assume that the unknown inputs to the system are random and gaussian, and that the system is piecewise linear, so that it can be admitted that samples e_k^* are imprecise observations of N_g different gaussian populations, N_g being unknown. This set of assumptions covers many practical situations. It is interesting to account that this particularization of our method leads to a result closely related to previous fuzzy learning algorithms [2].

4.1. Memberships from numerical data

Let F be the distribution function of an unidimensional gaussian random variable, with mean 0 and standard deviation 1. Let ξ be another random variable, normal and multidimensional, whose covariance matrix is C and whose mean vector is c.

C is symmetrical and positive definite, so there exists an orthogonal matrix B such that

$$C = B^{t} V B \tag{19}$$

where V is a diagonal matrix, whose elements are the eigenvalues of C,

$$V = \begin{pmatrix} \sigma_1^2 & \cdots & 0\\ \vdots & & \vdots\\ 0 & \cdots & \sigma_p^2 \end{pmatrix}.$$
 (20)

If we define

$$C^{-1/2} = B^{\rm t} V^{-1/2} B \tag{21}$$

and

$$A = V^{-1/2}B \tag{22}$$

so that

$$C^{-1} = A^{\mathsf{t}} \cdot A \tag{23}$$

then we have that

$$A(\xi - c) \to \mathcal{N}(\tilde{0}, I) \tag{24}$$

where I is the identity matrix.

Let

$$\lambda = \inf\{t: A(e-c) \in [-t,t]^q \text{ for all } e \in c \oplus Z\}.$$
(25)

We suggest the random set-based model of the system is

$$\Phi = \xi \oplus \{ z \colon A(z-c) \in [-\lambda, \lambda]^q \}.$$
(26)

If $A(e-c) = (t_1, ..., t_q)^t$, then the associated coverage function will be given by

$$p(e \in \Phi) = p(A(e - \xi) \in [-\lambda, \lambda]^{q})$$

= $p(A(e - c) + A(c - \xi) \in [-\lambda, \lambda]^{q})$
= $p(N(\tilde{0}, I) \in [-\lambda + t_{1}, \lambda + t_{1}]$
 $\times \cdots \times [-\lambda + t_{q}, \lambda + t_{q}])$
= $\prod_{i=1}^{q} (F(t_{i} + \lambda) - F(t_{i} - \lambda)),$ (27)

where F is the distribution function of the standard normal distribution.

348



Fig. 6. Approximation factor for gaussian processes.

Since the membership function of the fuzzy set \mathcal{T} associated with Φ is

$$\mathcal{F}(e) = \frac{p(e \in \Phi)}{\sup_{s \in X \times X \times U} \{ p(s \in \Phi) \}}$$

for all $e \in X \times X \times U$ (28)

then

$$\mathscr{T}(e) = \frac{1}{(2F(\lambda) - 1)^q} \prod_{i=1}^q (F(t_i + \lambda) - F(t_i - \lambda))$$
(29)

is our fuzzy model of the system.

When the observation error is low with respect to the probabilistic uncertainty in the model, the obtained model is numerically very close to other one in which the observation error vanishes and the uncertainty due to unmodelled inputs was slightly higher. Let C^* be the covariance matrix of a random process modelling this hypothetical system.

Observe that there exists a value $k(\lambda)$ such that

$$err(\lambda) = \sup_{e \in \mathbb{R}} |F(e+\lambda) - F(e-\lambda) - (2F(\lambda) - 1)e^{-(1/2)k(\lambda)e^2}| < 0.05 \quad (30)$$

if $\lambda \in [0, 2]$ (values of $k(\lambda)$ and maximum error of the approximation are represented in Fig. 6.) Then,

$$\mathcal{F}(e) \approx \frac{1}{(2F(\lambda) - 1)^q} \prod_{i=1}^q (2F(\lambda) - 1) \mathrm{e}^{-(1/2)k(\lambda)t_i^2},$$
(31)

where

$$\sum_{i=1}^{q} t_i^2 = t^{t} t = (A(e-c))^{t} (A(e-c))$$
$$= (e-c)^{t} (A^{t}A)(e-c)$$
$$= (e-c)^{t} C^{-1}(e-c), \qquad (32)$$

whence

$$\mathscr{T}(e) \approx e^{-(1/2)k(\lambda)(e-c)^{t}C^{-1}(e-c)}.$$
(33)

Since the membership of the hypothetical observation noise-free system would be

$$\mathscr{T}^{*}(e) = e^{-(1/2)(e-c)^{t}(C^{*})^{-1}(e-c)}$$
(34)

it follows that

$$C = k(\lambda)C^* \tag{35}$$

when λ is low.



Fig. 7. Coverage functions. Left: high value of λ . Right: low value of λ .

This result suggests a relationship between the unknown variance of the underlying random process and the known variance of the observations set E. This relationship has a key importance in our algorithm, because it allows us to infer the value of the parameter Cof the random process that underlies the fuzzy model.

Notice also that, for high values of λ , $\mathcal{T}(e)$ approximates to the membership of the crisp set

$$\{e: A(e-c) \in [-\lambda, \lambda]^q\}.$$
(36)

In fact, the value λ expresses the balance between randomness and fuzziness in the fuzzy random variable \mathscr{X} . Thus, if $\lambda = 0$, the variable \mathscr{X} is a nonfuzzy random one and it becomes a crisp (nonrandom) set as $\lambda \to \infty$; see Fig. 7.

The covariance C^* can be estimated from the sample $\{e_i^*\}$. However, $C = k(\lambda)C^*$ and λ depends on C, which is unknown. C will be estimated then by stating a sequence of values

$$\lambda_{i+1} = \min\{t: A_i(e-c) \in [-t,t]^q \text{ for all } e \in c \oplus Z\},$$
(37)

where

$$A_{i+1}A_{i+1}^{t} = k(\lambda_{i+1})(C^{*})^{-1}$$
(38)

and by using as starting conditions

$$\lambda_1 = 0, \quad A_1(A_1)^t = (C^*)^{-1}.$$
 (39)

The process will be repeated until λ_{i+1} is close enough to λ_i .

4.2. Semantic entropy

The expression (15) is only applicable to discrete universes. We can consider as a natural extension for this method the replacement of the cardinality of the α -cuts by their Lebesgue measure, so that Hartley entropy would be replaced by the upper limit of the Boltzmann entropy of any probability distribution defined on the α -cut.

The semantic entropy of the set \mathcal{T} can be computed as follows:

$$U(\mathscr{F}) = \int_0^1 \log(|\mathscr{F}_{\alpha}|) \, \mathrm{d}\alpha = \int_0^1 \log(|A|^{-1}|\mathscr{N}_{\alpha}|) \, \mathrm{d}\alpha$$
$$= \frac{1}{2} \log|C| + U(\mathscr{N}, \lambda), \tag{40}$$

where

$$\mathcal{N}(e) = \mathcal{N}(e_1, \dots, e_q) = \frac{1}{(2F(\lambda) - 1)^q} \prod_{i=1}^q F(e_i + \lambda) -F(e_i - \lambda).$$
(41)

Some values of $U(\mathcal{N}, \lambda)$ are collected in Fig. 8.

4.3. Description length of a model

Given a model

$$\mathcal{M}_{A} = \bigvee_{i=1}^{N} \mathcal{T}_{i} \tag{42}$$

λ	$k(\lambda)$	$U(\mathcal{N})$
0.00	1.00	0.75q
0.25	0.98	0.76q
0.50	0.92	0.80q
0.75	0.83	0.85q
1.00	0.73	0.91 <i>q</i>
1.25	0.62	1.00 <i>q</i>
1.50	0.52	1.09 <i>q</i>
1.75	0.42	1.20q
2.00	0.35	1.31q
>2	-	$q \log(2\lambda)$

Fig. 8. Values of $k(\lambda)$ and $U(\mathcal{N})$.

with $N = N_g$ sets, estimated from a partition $\Lambda = \{E_i\}$, its description length will be

$$I(E, \mathcal{M}_{A}) = \log N + \sum_{i=1}^{N} \frac{\#(E_{i})}{\#(E)} \left(\frac{1}{2} \log |C_{i}| + U(\mathcal{N}, \lambda_{i})\right), \quad (43)$$

where the matrix C_i can be inferred from E_i by applying (35) and (37)–(39).

It is remarkable that when there is no observation noise, the value of the suggested description length differs by an additive constant from an equivalent expression in which the semantic entropy was measured through Boltzmann's entropy of gaussian distributions fitting to subsets E_i . Thus, the proposed fuzzy minimum description length criterion reduces to the probabilistic one when all variables can be exactly observed.

5. Examples

5.1. Numerical example

The following input-output samplings of a physical system variables are available:

$e_1 = (0.619, 0.695),$	$e_2 = (1.392, 1.183),$
$e_3 = (1.276, 0.901),$	$e_4 = (0.764, 1.398),$
$e_5 = (0.704, 0.947),$	$e_6 = (0.885, 0.646),$
$e_7 = (1.472, 1.329),$	$e_8 = (0.607, 0.675),$

$e_9 = (0.867, 0.589),$	$e_{10} = (1.476, 0.792),$
$e_{11} = (-0.83, -0.912),$	$e_{12} = (-1.413, -1.265),$
$e_{13} = (-1.469, -1.129),$	$e_{14} = (-1.128, -1.179),$
$e_{15} = (-1.106, -0.829),$	$e_{16} = (-0.509, -1.191),$
$e_{17} = (-0.763, -0.755),$	$e_{18} = (-1.324, -0.665),$
$e_{19} = (-1.088, -1.259),$	$e_{20} = (-0.762, -1.28),$
$e_{21} = (2.225, 2.474),$	$e_{22} = (2.345, 1.916),$
$e_{23} = (1.924, 2.118),$	$e_{24} = (2.061, 1.874),$
$e_{25} = (1.658, 2.362),$	$e_{26} = (2.405, 2.334),$
$e_{27} = (1.822, 1.583),$	$e_{28} = (1.513, 1.596),$
$e_{29} = (1.559, 1.679),$	$e_{30} = (2.056, 1.533).$

The instrument employed to take the measurements guarantees that, 99% of the times, real values of the variables are in a circle with diameter 0.5 units, and centered in the inexact measurement. We wish to estimate a fuzzy model of the system's input-output transference relationship.

5.1.1. Solution

We will consider that all tabulated values are outcomes of a random two-dimensional gaussian process, measured with an imprecision characterized by means of the set Z

$$Z = \{e: e^{t} \mathbb{Z} e \leq 1\},\tag{44}$$

where

$$\mathbf{Z} = \begin{pmatrix} 16 & 0\\ 0 & 16 \end{pmatrix}. \tag{45}$$

If the variance of the random process is C, $\{A(z-c): z \in Z\}$ is inside a cubic cell $[-\lambda, \lambda]^2$, so that

$$\lambda = \min\{t: A(e-c) \in [-t,t]^2, \ e \in Z\}.$$
(46)

and since

$$(A(e-c))^{t}W(A(e-c)) \le 1$$

for $W = (A^{-1})^{t}ZA^{-1}$ (47)



Fig. 9. α -cuts ($\alpha = 0.1$) of M_{30} , M_{10} , M_5 , M_3 , M_2 and M_1 .

then

$$\lambda = \max\{\sqrt{u_{ii}}: W^{-1} = (u_{ij})\}.$$
(48)

If $\lambda < 2$, the membership of the set \mathcal{T} will be approximated by

$$\mathscr{T}(e) \sim \mathrm{e}^{-0.5k(\lambda)e^{\mathrm{t}(C)^{-1}e}} \sim \mathrm{e}^{-0.5e^{\mathrm{t}(C^*)^{-1}e}},$$
 (49)

where C^* is the variance of the inexact sample. Otherwise, \mathcal{T} degenerates to the crisp set

$$\mathscr{T} = \{ e: A(e-c) \in [-\lambda, \lambda]^2 \}.$$
(50)

The semantic uncertainty of the set \mathcal{T} will be given by

$$U(\mathcal{T}) = 0.5 \log|C| + U(\mathcal{N}, \lambda)$$
(51)

and $C^* = C$ if $\lambda > 2$.

5.1.2. Numerical results

First, the model \mathcal{M}_{30} is calculated. In this model, every region is estimated from one sample. We obtain that

$$U(\mathcal{F}_i) = -1.39\tag{52}$$

N. of sets	Description L.	Semantic. U.	Syntactic U.	
1	21.9483	21.9483	0	
2	6.80813	-13.9863	20.7944	
3	16.3077	-16.6507	32.9584	
4	22.0352	-19.5536	41.5888	
5	27.418	-20.8651	48.2831	
6	29.6316	-24.1212	53.7528	
7	32.4765	-25.9008	58.3773	
8	34.1291	-28.2542	62.3832	
9	36.4016	-29.5152	65.9167	
10	37.8247	-31.2529	69.0776	
15	48.2461	-32.9954	81.2415	
20	52.1016	-37.7704	89.872	
25	60.2566	-36.3097	96.5663	
30	60.4471	-41.5888	102.036	

Fig. 10. Uncertainties of the models.

for all sets. The model reduces to 30 square-shaped crisp sets (see Fig. 9) centered on every sample. The description length of this model is given by

$$LD = \sum_{i=1}^{30} 1 \cdot (-1.39) + 30 \log 30 = 102.04.$$
 (53)

In the following step, we examine how the semantic uncertainty increases as two sets in the partition are merged. For instance, when sets 1 and 2 are merged,



Fig. 11. Point estimates based on the suggested method.

the fuzzy set $\mathscr{T}_{E_1 \cup E_2}$ has a nonspecificity

$$U(\mathcal{F}) = 2.44\tag{54}$$

and the increasing of syntactic uncertainty equals

$$\Delta_{1,2} = 4.87 + 1.39 + 1.39 = 7.64. \tag{55}$$

If the process is repeated for all pairs of sets, the most favourable result is achieved when sets 1 and 8 are merged. The lowest gain of syntactic uncertainty is $\Delta_{1,8} = 0.86$. Then, the sets \mathcal{T}_1 and \mathcal{T}_8 of the model \mathcal{M}_{30} are discarded and $\mathcal{T}_{1,8}$ is added to form the model \mathcal{M}_{29} . The whole process is repeated until \mathcal{M}_1 is obtained.

Syntactic entropy, semantic entropy and description length of the successive models are tabulated in Fig. 10. The chosen model of the system is formed by two fuzzy sets, which are

$$\mathcal{F}_{1}(x, y) = \exp\left(0.5(x, y)^{t} \begin{pmatrix} 0.33 & 0.29\\ 0.29 & 0.33 \end{pmatrix} (x, y)\right)$$
(56)

and

$$\mathcal{F}_{2}(x, y) = \exp\left(0.5(x, y)^{t} \begin{pmatrix} 0.09 & -0.002 \\ -0.02 & 0.05 \end{pmatrix} (x, y)\right).$$
(57)

Point estimates based on this model were plotted in Fig. 11. These values were obtained by assessing crisp values to the input variable, to get fuzzy restrictions over the output space by applying (3). Then, the points with maximum membership for each of the obtained fuzzy restrictions were related to their associated crisp values. The so-formed triplets lie on a different plane (which is a linear model of the system) for every set \mathcal{F}_i . Relating again these planes to the projections of their original fuzzy regions over the input space, we obtain a bank of Sugeno-type rules, whose output was solved in the usual way.

5.2. Comparison with other methods

Box-Jenkins gas furnace data set is part of a classical problem which is widely used to test performances of system identification procedures. This set has been used to compare some fuzzy-set-based schemes [13]. The performance comparison is based on a point estimate of the output of the system. In this case, this estimate was obtained as described in the preceding example.

Assuming that this data set comes from several inexactly measured gaussian stationary processes, and after applying our method the results gathered in Fig. 12 were obtained. We only find a population,

q = 3	N. of Rules	q = 4	N. of Rules
0.469	19	_	_
0.320	81	-	_
0.328	25	-	-
0.359	2	-	_
-	-	0.190	6
0.181		0.150	_
0.181	1	0.150	1
	q = 3 0.469 0.320 0.328 0.359 - 0.181 0.181	q = 3 N. of Rules 0.469 19 0.320 81 0.328 25 0.359 2 - - 0.181 - 0.181 1	q=3N. of Rules $q=4$ 0.46919-0.32081-0.32825-0.35920.1900.181-0.1500.18110.150

Fig.	12.	Numerical	results,	Box'	S	gas	furnace
------	-----	-----------	----------	------	---	-----	---------

so the obtained bank of Sugeno rules degenerates to a classical linear model. We have assumed that the tolerance of a measurement was defined by an ellipsoid, whose axes are the admissible margins (0.5 and 1 for inputs and outputs, respectively). Values of errors achieved with other methods, collected in the same Fig. 12, have been taken from [12, 13]. The values labelled "Linear" were obtained by fitting a plane by least squares method to the whole set of samples.

Results in Fig. 12 should be carefully interpreted, because the sets of state variables chosen are sometimes different. The tabulated values of error are the medium-squared differences between samples and model.

6. Conclusions

When measurements are inexact, successive observations of variables in certain physical systems are experiments that can be modelled in terms of fuzzy random variables. These fuzzy variables can be expressed by means of their level cuts, which are random sets. Since one-point coverages of random sets are compatible with memberships of fuzzy sets, these sets were used to build a fuzzy model.

Then, a procedure to adjust parametric memberships to subsets of the initial samples set has been devised. The volumes of the subsets were selected to simultaneously achieve the required precision and maintain a simple model, by using an uncertaintybased quality measure of the fuzzy models.

The final algorithm can be applied to qualitatively model a system, generating rules from examples, and also to design nonlinear purely numerical fuzzy filters.

Acknowledgements

The research in this paper has been supported in part by DGICYT Grant No. PB95-1049-A. Their financial support is gratefully acknowledged.

Reference

- H.R. Berenji, Fuzzy systems that can learn, in: Computational Intelligence: Imitating Life, IEEE Press, New York, 1994, pp. 23-30.
- [2] J.A. Dickerson, B. Kosko, Ellipsoidal learning and fuzzy throttle control for platoons of smart cars, in: Yager, Zadeh (Eds.), Fuzzy Sets, Neural Networks and Soft Computing, Van Nostrand Reinhold, New York, 1994, pp. 63-84.
- [3] M.A. Gil, Sufficiency and fuzziness in random experiments, Ann. Inst. Statist. Math. 44(3) (1992) 451-462.
- [4] M.M. Gupta, J. Qi, Theory of t-norms and fuzzy inference methods, Fuzzy Sets and Systems 40 (1991) 431-450.
- [5] H. Kwakernaak, Fuzzy random variables: I. Definitions and theorems, Inform. Sci. 15 (1978) 1–29.
- [6] G. Klir, T. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [7] P.S. Khedkar, Learning as adaptive interpolation in neural fuzzy systems, in: Yager, Zadeh (Eds.), Computational Intelligence: Imitating Life, IEEE Press, New York, 1994, pp. 31-42.
- [8] M. Laviolette, A probabilistic and statistical view of fuzzy methods, Technometrics 37(3) (1995) 249-262.
- [9] G. Nguyen, Uncertainty Models for Knowledge-based Systems, North-Holland, Amsterdam, 1985.
- [10] T.J. Procyk, E.H. Mamdani, A linguistic self-organizing process controller, Automatica 15 (1979) 15–30.
- [11] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
- [12] M. Sugeno, K. Tanaka, Successive identification of a fuzzy model and its applications to prediction of a complex system, Fuzzy Sets and Systems 42 (1991) 315–334.
- [13] M. Sugeno, T. Yasukawa, A fuzzy logic based approach to qualitative modelling, IEEE Trans. Fuzzy Systems 4 (1993) 7-31.
- [14] T. Tagaki, M. Sugeno, Fuzzy identification of systems and its application to modelling and control, IEEE Trans. Systems Man Cybernet. 15 (1985) 116-132.
- [15] E. Trillas, C. Alsina, J.M. Terricabras, Introducción a la lógica borrosa, Ariel, 1995.
- [16] R.S. Wallace, Finding natural clusters through entropy minimization, Ph.D. Thesis, School of Computer Science, Carnegie Mellon University, 1989.
- [17] M. Wang, Generating fuzzy rules by learning from examples, IEEE Trans. Systems Man Cybernet. 22(6) (1992) 1414-1427.