ARTICLE IN PRESS

Mechanical Systems and Signal Processing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at SciVerse ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials

Luciano Sánchez [a], Inés Couso [b,*], Ana M. Palacios [a], José L. Palacios [c]

[a] Departamento de Informática, Universidad de Oviedo, 33071 Gijón, Asturias, Spain
[b] Departamento de Estadística, Universidad de Oviedo, 33071 Gijón, Asturias, Spain
[c] Department of Aerospace Engineering, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 17 December 2011
Received in revised form
13 February 2012
Accepted 18 February 2012

Keywords:
Genetic fuzzy systems
Fuzzy rule-based classifiers
Vague data
Mutual information
Ice-phobic materials
Shear adhesion strength

ABSTRACT

A methodology for obtaining fuzzy rule-based models from uncertain data is proposed. The granularity of the linguistic discretization is decided with the help of a new estimation of the mutual information between ill-known random variables, and a combination of boosting and genetic algorithms is used for discovering new rules. This methodology has been applied to predict whether the coating of an helicopter rotor blade is adequate, considering the shear adhesion strength of ice to different materials. The discovered knowledge is intended to increase the level of post-processing interpretation accuracy of experimental data obtained during the evaluation of ice-phobic materials for rotorcraft applications.

1. Introduction

Expert knowledge elicitation from experimental data is a valuable tool in the engineering design process, that aids to make agile decisions in the absence of a causal model or extensive prototype testing. Often, expert knowledge is best organized as a set of linguistic rules, describing the circumstances under which the behavior of an element is admissible or unsuitable for a given application. The benefits of computer-produced knowledge depend on the accuracy of the conclusions that might be drawn from it, and also on it being described at a level that is understandable to the design engineer [4].

It is generally agreed that, given the right amount of quality data, computer algorithms are capable of obtaining accurate and informative rule bases, however in practice this is not consistently so. Comprehensive experimental descriptions can be time consuming and expensive, and therefore there is a need for exploiting the information in low quality data, including scarce, incomplete and/or imprecise sources of information [12]. Numerous studies have been published regarding the representation of uncertain empirical information, and the difference between different types of uncertainties [5,16–18,21,23,24]. According to [16], there are two main categories of uncertainty: stochastic uncertainty, that arises from random variability related to natural processes such as the heterogeneity of population or the fluctuations

* Corresponding author. Tel.: +34 98 51 81906.
E-mail addresses: luciano@uniovi.es (L. Sánchez), couso@uniovi.es (I. Couso), palaciosana@uniovi.es (A.M. Palacios), jlp324@psu.edu (J.L. Palacios).

0888-3270/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ymssp.2012.02.009

Table 1
SAT for aluminum with a temperature of −11 °C, according to different authors.

Author	Ice	Test	SAT-kPa
Loughborough and Hass [30]	Freezer ice	Pull	558
Stallabrass and Price [50]	Impact ice	Rotating instrumented beam	97
Itagaki [28]	Impact ice	Rotating rotor	27–157
Scavuzzo and Chu [47]	Impact ice	Shear window	90–290
Reich [40]	Freezer ice	Pull	896

of a quantity with time, and epistemic uncertainty, that arises from the incomplete or imprecise nature of available information. Stochastic uncertainty can be modeled with classical probability theory, however there are different theories for handling incomplete and imprecise information [15], which often appears in engineering problems.

The purpose of this study is therefore to define a methodology for obtaining expert knowledge, comprising fuzzy classification rules, from ill-known data. This methodology will be demonstrated on a real-world problem: discovering a set of linguistic rules describing the ice accretion strength of different materials used in helicopter rotor blades, for different ambient conditions.

This problem is out of the ordinary because neither the ambient conditions nor the experimental parameters of the essays can be precisely determined on all occasions; they may change during the realization of an essay, and sometimes they cannot be directly measured. For that matter, the properties of some materials cannot be reliably estimated. This last fact will be illustrated with the help of an example: In Table 1, some estimations of the shear adhesion strength (SAT) for Aluminum at −11 °C, found in the literature of the field, were collected. These values are much different among themselves, arguably as consequence of the impossibility of accurately determining the value of some ambient or experimental parameters.

Following with the example, imagine that a SAT of 100 kPa was measured in an experiment for which the initial temperature was −14 °C and the final temperature was −10 °C. In this case, it is not correct to write "SAT=100 kPa at a temperature of −12 °C" neither it is to state "SAT=100 kPa for temperatures between −14 and −10 °C". Our knowledge is restricted to the fact "SAT=100 kPa at an unknown temperature between −14 and −10 °C". Clearly, linguistic modeling techniques are well suited for expressing this kind of information.

In view of the above, the first part of this paper contains a description of the proposed modeling methodology and their assumptions, and the second part of this study describes a practical application of this methodology to rotor blade characterization. In Section 2 the use of fuzzy sets for describing the uncertainty in the data is explained. In Section 3 it will be explained how ill-known data is discretized into linguistic values. The selection of an informative discretization is addressed in Section 4, where a mutual information measure for fuzzy discretized data is proposed. In Section 5 two algorithms for finding fuzzy classification rules from imprecise data are described, and in Section 6 the demonstration problem is detailed. The paper is finished with some concluding remarks, in Section 7.

2. Fuzzy models of uncertainty in the data

The use of stochastic techniques for describing the numerical uncertainty in experimental data is prevalent among researchers and practitioners. However there may be cases where there are better alternatives. For example, in presence of coarse digital measurements (lack of significant digits), censored data or missing values, a probabilistic model is too restrictive. Interval-valued descriptions or other characterizations of the uncertainty, based on families of probability distributions, are to be preferred [31].

The use of a possibility distribution for describing partial ignorance about a value falls in the second of these groups. Fuzzy membership functions can be derived from possibility distributions, and interval-valued descriptions are particular cases of this representation. Moreover, the same description can also be used for summarizing conflicting data, as happens for instance when a set of measurements of the same physical magnitude is produced by different sensors. Here is a case in point: suppose that these conflicting measurements are

$$X = \{2, 1, 3, 3, 2, 2, 4\}. \tag{1}$$

Their average is 2.429. Nevertheless, using this value for describing the unknown value of the physical magnitude discards information that might be relevant: there are some items as low as 1, and others as high as 4. To gain additional insight about the importance of the dispersion of the values, it can be assumed that the set of items X is a sample of a larger population, whose mean is unknown. Confidence intervals for the value of this mean can be computed for different signification levels, and the knowledge about the unknown magnitude described by a list of nested confidence intervals. Following with the same example, if a Gaussian population is assumed, these confidence intervals are

$$\tilde{X}_\alpha = 2.429 \pm 0.9759 \cdot qt_6\left(1 - \frac{\alpha}{2}\right), \tag{2}$$

where 0.9759 is the standard deviation and qt_6 is the quantile function for the t distribution. According to [7,8,46], this set of intervals contains the same information about the unknown variable that a possibility distribution defined by a fuzzy

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

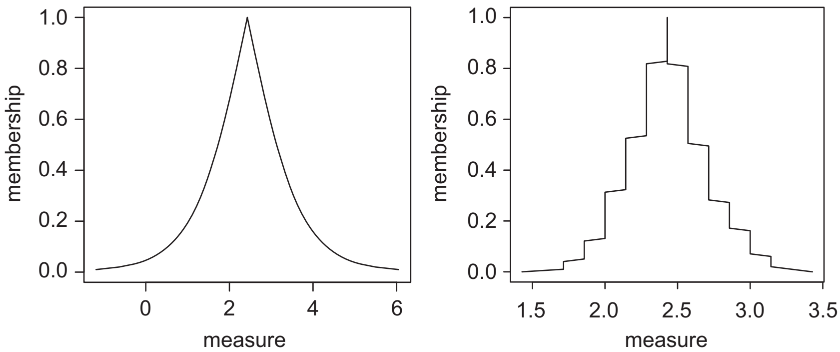

Fig. 1. Membership function of the fuzzy set that represents the sample *X*. The left one was obtained under normality assumptions, and the right one is a basic bootstrap estimation.

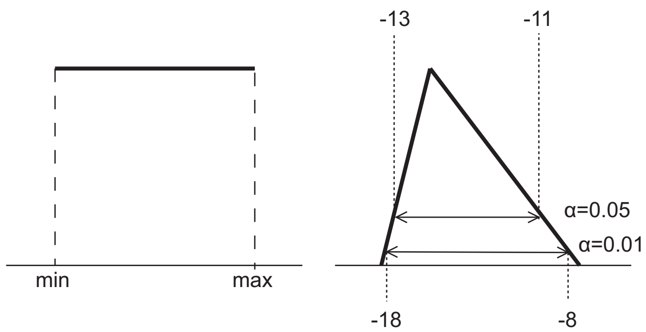

Fig. 2. Fuzzy representation of vague data. *Left*: the only information is the range of the variable. *Right*: a nested family of confidence intervals is compatible with a possibility distribution, described by means of a fuzzy membership function, whose level cuts are confidence intervals at the level $1-\alpha$. The membership function shown is compatible with the information "99% of times the temperature was between $-18\,^{\circ}$C and $-8\,^{\circ}$C, but 95% of times it was between $-13\,^{\circ}$C and $-11\,^{\circ}$C".

set whose α-cuts are confidence intervals at the levels $1-\alpha$. Hence, the membership function associated to the set *X* is shown in Fig. 1 (left part). Observe that we can approximate this shape by a triangular membership function without incurring large errors. Other techniques for computing the membership can also be applied; in the right part of Fig. 1 a bootstrap-based estimation of the membership function associated to the same data is plotted.

Because of these reasons, a possibilistic representation of ill-known data is adopted in this study. Generally speaking, those situations where one or more confidence intervals for a parameter can be determined are well suited for this technique. Conflicting values will also be aggregated and transformed into fuzzy sets. In addition to this, a triangular membership will be used for approximating those sensors whose specification comprises a broad range of extreme values, and also a smaller typical range: "99% of times the temperature was between $-18\,^{\circ}$C and $-8\,^{\circ}$C, but 95% of times it was between $-13\,^{\circ}$C and $-11\,^{\circ}$C" (see Fig. 2).

3. Linguistic discretization of ill-known data

The notation used in the rest of the paper is introduced at this point. Let $x=(x_1,\dots,x_d)$ be a vector of features, and let a fuzzy rule-based classifier system be a list of *M* "IF-THEN" rules, comprising an antecedent, a consequent and sometimes a weight:

$$\text{if } (x \text{ is } \tilde{A}_i) \text{ then class is } C_i \text{ [with weight } w_i], \tag{3}$$

where $\tilde{A}_i$ is a fuzzy subset of $\mathbb{R}^d$, and the expression "x is $\tilde{A}_i$" is a combination of asserts of the form "x_p is $\tilde{A}_{iq}$" by means of different logical connectives. The terms $\tilde{A}_{iq}$, for $q=1,\dots,n_i$, are, in turn, fuzzy subsets of $\mathbb{R}$ that have been assigned a linguistic meaning, and the membership function of $\tilde{A}_i$ models the degree of truth of this combination. C_i are the labels of the different classes. w_i are degrees of credibility that may be given to each rule, $w_i \in [0,1]$. For instance, given a rule base

$$\begin{aligned}&\text{if } x_1 \text{ is HIGH and } x_2 \text{ is LOW then class is } C_1 \text{ with weight 0.8,}\\&\text{if } x_1 \text{ is MEDIUM and } x_2 \text{ is MEDIUM then class is } C_2 \text{ with weight 0.4,}\\&\text{if } x_2 \text{ is HIGH then class is } C_1 \text{ with weight 0.9,}\end{aligned} \tag{4}$$

then $\tilde{A}_{11}=\text{HIGH}$, $\tilde{A}_{12}=\text{LOW}$, and the membership functions of these two sets define the compatibility between the values of x_1 and x_2 and the linguistic terms "HIGH" and "LOW". The fuzzy set $\tilde{A}_1=\min(\tilde{A}_{11},\tilde{A}_{12})$ defines the compatibility of each pair (x_1,x_2) and the antecedent of the first rule.

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

The output of the rule base is determined by a voting procedure, where each rule is assigned a number of votes equal to the compatibility between x and its antecedent, multiplied by its weight (the number of votes does not need to be integer-valued). The mechanisms for choosing the winner alternative are two:

1. The class of the object is given by the consequent of the most voted rule:
$$\text{class}(x) = C_k \quad \text{where } k = \arg\max_i \{w_i \mu_{\tilde{A}_i}(x)\}. \tag{5}$$

2. The votes of all rules with the same consequent are added, and the option with a higher number of votes is chosen [29]:
$$\text{class}(x) = \arg\max_i \left\{ \sum_{j=1,\ldots,M} \delta^j_{C_i} w_j \mu_{\tilde{A}_j}(x) \right\}, \tag{6}$$
where δ^b_a is Kronecker's delta, $\delta^b_a = [a=b]$.

For features that are not precisely known, a fuzzy set is used for describing the inputs: $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_d)$, as mentioned in the preceding section. In this case the output is no longer a class label but a fuzzy set of classes denoted by class$(\tilde{X})$. The membership function of this set is

$$\mu_{\text{class}(\tilde{X})}(c) = \max\{\mu_{\tilde{X}}(x) \mid c = \text{class}(x)\}. \tag{7}$$

It has been mentioned that the terms $\tilde{A}_{iq}$ are fuzzy subsets of $\mathbb{R}$ that have been assigned a linguistic meaning. This assignment is based on a fuzzy partition defined on the universe of discourse. Strong fuzzy partitions [41,51] will be used, fulfilling the strength condition

$$\forall i,\ \forall x, \quad \sum_{q=1}^{n_i} \mu_{\tilde{A}_{iq}}(x) = 1, \tag{8}$$

and also

- $\mu_{\tilde{A}_{iq}}(m_q) = 1$ where m_q is the mode of the cell $\tilde{A}_{iq}$ (unimodality).
- $\mu_{\tilde{A}_{iq}}(x)$ monotonically increases in $[m_{q-1}, m_q]$ and monotonically decreases on $[m_q, m_{q+1}]$.
- For each i, there are not more than two different memberships that are not null at the same point x.

Because of the strength condition, a real number is mapped to a vector of memberships whose sum is 1. This fact allows considering the elements of a strong fuzzy partition as likelihood functions, i.e. $\mu_{\tilde{A}_{iq}}(x)$ is the proportion of times that the value x is tagged as $\tilde{A}_{iq}$ in a random experiment [14]. Let S_{iq} be the linguistic term associated to the fuzzy set $\tilde{A}_{iq}$. The vector of memberships can then be understood as a probability distribution over the set of terms

$$L_{iq}(x) = P(S_{iq} \mid x) = \mu_{\tilde{A}_{iq}}(x). \tag{9}$$

Furthermore, we will assume that the upper probability of the qth term in the antecedent of the ith rule, conditioned on an unknown value in the interval input $[x_*, x^*]$, is

$$L_{iq}([x_*, x^*]) = \sup\{P(S_{iq} \mid x) \mid x \in [x_*, x^*]\} = \sup\{\mu_{\tilde{A}_{iq}}(x) \mid x \in [x_*, x^*]\}. \tag{10}$$

4. Selection of an informative linguistic partition

There is necessarily a loss of information in any linguistic discretization of a variable. This loss depends on the partition granularity. It is possible that a coarse discretization makes an informative variable irrelevant.

However the number of terms in a partition must be low for a good interpretability. There is a balance between the amount of information that is lost in the discretization and the degree of understandability of the knowledge base that will be ultimately produced. In this paper it is proposed to find this equilibrium by choosing the partition with the least terms that do not incur in a high information loss. The loss of information in a discretization will be measured by means of an extension of the mutual information (MI) [45] that is described now.

Let N be the number of available empirical data, and let $(X_1, X_2, \ldots, X_N)$ and $(Y_1, Y_2, \ldots, Y_N)$ be the observed values of two features (or one feature and the response variable). These will be regarded as two paired samples of two discrete random variables X and Y that can take n and m different values, respectively. Let $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_m$ be the frequencies of the respective values in both samples, and let $r_1, r_2, \ldots, r_s$, where $s = n \times m$, be the frequencies of the values of the joint sample $X \times Y$. The mutual information between the variables X and Y is estimated as follows:

$$\text{MI}((X_1, \ldots, X_N), (Y_1, \ldots, Y_N)) = -\sum_{i=1}^{n} p_i \log p_i - \sum_{i=1}^{m} q_i \log q_i + \sum_{i=1}^{s} r_i \log r_i. \tag{11}$$

A naive estimation of the MI from two samples of continuous random variables consists in discretizing these variables [10], computing the frequencies of the discretized values and applying Eq. (11) [39].

For extending this estimation to linguistic fuzzy partitions, it is proposed that these partitions are regarded as bins of Quasi-Continuous Histograms (QCH) [51], and the estimation of the MI is rewritten by replacing the empirical frequencies with possibilistic counting transfers. Let $(\overline{Z}_1,\ldots,\overline{Z}_N)$ be a sample of a random interval. Let be considered the event "the discretization of $(\overline{Z}_1,\ldots,\overline{Z}_N)$ is the vector of linguistic terms $(S_{iz_1},\ldots,S_{iz_N})$", where $1 \le z_i \le n_i$. The probability of this event (see Eq. (10)) is an unknown element of the set

$$P_Z(z_1,\ldots,z_N) = \left\{ \prod_{i=1}^{N} p_i \,\middle|\, p_i \le L_{iz_i}(\overline{Z}_i),\ \sum_{i=1}^{N} p_i = 1 \right\}. \tag{12}$$

It is proposed that the mutual information between the discretizations of the samples of two ill-known random variables $(\overline{Z}_1,\ldots,\overline{Z}_N)$ and $(\overline{T}_1,\ldots,\overline{T}_N)$ is defined by the imprecise probability distribution that follows:

$$P_{\mathrm{MI}}(m) = \left\{ \sum p_z p_t \,\middle|\, p_z \in P_Z(z_1,\ldots,z_N), p_t \in P_T(t_1,\ldots,t_N),\ m = \mathrm{MI}((z_1,\ldots,z_N),(t_1,\ldots,t_N)) \right\}. \tag{13}$$

Since there is a finite number M of different discretizations, so there are at the most M different values of the mutual information. The expected value of the estimated mutual information is therefore the set

$$\mathrm{MI}((\overline{Z}_1,\ldots,\overline{Z}_N),(\overline{T}_1,\ldots,\overline{T}_N)) = \left\{ \sum_{k=1}^{M} m_k p_k \,\middle|\, p_k \in P_{\mathrm{MI}}(m_k),\ \sum_{k=1}^{M} p_k = 1 \right\}. \tag{14}$$

This procedure can be extended to possibilistic data by applying the method proposed in [9]. In short, this method consists in regarding the fuzzy expectation as a possibility distribution over the set of all possible expected values. A possibility distribution represents a possibility measure, which is the upper probability associated to a set of probability measures, hence it can be considered that there is a set of probability measures defined on the set of expected values. For simplicity, let be supposed that there is a finite set of expected values $e_1,\ldots,e_n$ and one of the considered probabilities assigns the values $p_1,\ldots,p_n$ to these expected values. Then it is reasonable to consider the expectation as the mean of these possible expectations, with respect to this probability, $\sum_{i=1}^{n} e_i \cdot p_i$. It can be checked that the mean (in a Dubois–Prade sense [13]) of the fuzzy expectation is the interval of the means of all the expectations, for all the probabilities dominated by the considered possibility measure.

Fuzzy random variables (frv) are intended for cases where the outcomes of a random experiment are modeled by fuzzy sets. A frv is a mapping that associates a fuzzy subset of the final space to each possible result of a random experiment. This association expresses the imprecise information about the relation between both universes. Thus this concept generalizes the definition of random variable. Following the model proposed in [8], let $(\tilde{Z}_1,\ldots,\tilde{Z}_N)$ and $(\tilde{T}_1,\ldots,\tilde{T}_N)$ be samples of a fuzzy random variable. Their level cuts $([\tilde{Z}_1]_\alpha,\ldots,[\tilde{Z}_N]_\alpha)$ and $([\tilde{T}_1]_\alpha,\ldots,[\tilde{T}_N]_\alpha)$ can be regarded as samples of random intervals, and therefore it makes sense to define

$$\mathrm{MI}_\alpha = \mathrm{MI}(([\tilde{Z}_1]_\alpha,\ldots,[\tilde{Z}_N]_\alpha),([\tilde{T}_1]_\alpha,\ldots,[\tilde{T}_N]_\alpha)). \tag{15}$$

According to the mentioned method, it is proposed that the expected value of the mutual information between two fuzzy discretizations of ill-known samples is the interval

$$\mathrm{MI}((\tilde{Z}_1,\ldots,\tilde{Z}_N),(\tilde{T}_1,\ldots,\tilde{T}_N)) = \left[\int_0^1 \inf \mathrm{MI}_\alpha \, d\alpha,\ \int_0^1 \sup \mathrm{MI}_\alpha \, d\alpha \right]. \tag{16}$$

A numerical algorithm for approximating this value from a sample of data comprising a mix of numerical, interval-valued and fuzzy data is given in Listing 1.

Listing 1. Simplified pseudocode of the proposed method for estimating the mutual information between two imprecisely perceived samples of random variables with Quasi-Continuous Histograms.

function Mutual Information $(\{\tilde{Z}\}_{k=1}^{N},\{\tilde{T}\}_{k=1}^{N})$
  for α in { 0, 1/LEVELS, 2/LEVELS,…,1 }
    for iter in { 1,…,NUMITER }
      Randomly select two discretizations $(z_1,\ldots,z_N)$ and $(t_1,\ldots,t_N)$ of $\tilde{Z}$ and $\tilde{T}$
      $[z_*,z^*]$ = Bounds of Probability of Discretization $(\{[\tilde{Z}]_\alpha\}_{k=1}^{N}, (z_1,\ldots,z_N))$
      $[t_*,t^*]$ = Bounds of Probability of Discretization $(\{[\tilde{T}]_\alpha\}_{k=1}^{N}, (t_1,\ldots,t_N))$
      m = mutual information between the discretizations (see Eq. (11))
      Store the tuple $([z_*,z^*], [t_*,t^*], m)$ in the list L
      $c_z = (z_* + z^*)/2$
      $c_t = (t_* + t^*)/2$
      sump = sump + $c_z c_t$
      summ = summ + $m c_z c_t$
    end for
    avm = summ/sump
    for each element of the list L
      if m > avm **then**

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

sumpupper = sumpupper + mz^*t^*
pupper = pupper + z^*t^*
else
sumplower = sumplower + mz_*t_*
plower = plower + z_*t_*
endif
end for
leftresult = leftresult + sumplower/plower
rightresult = rightresult + sumupper/pupper
end for
return [leftresult/(NLEVELS+1), rightresult/(NLEVELS+1)]

function Bounds of Probability of Discretization ($\{\overline{X}\}_{k=1}^{N}$, $(z_1,\ldots,z_N)$)
upper = $\prod_k(\sup\{\tilde{A}_{z_k}(x) \mid x \in [\overline{X}_k]_\alpha\})$
lower = $\prod_k(1-\sup\{\tilde{A}_{z_k}(x) \mid x \notin [\overline{X}_k]_\alpha\})$
return [lower, upper]

5. Discovering fuzzy classification rules from imprecise data

In this section, two different algorithms for obtaining fuzzy classification rules from empirical data will be described. Both algorithms search for a set of rules that maximize the accuracy of the model. This accuracy is understood as the expected number of misclassifications of the model. However since the data is imprecise, this number cannot always be determined, as discussed in the following section.

5.1. Assessing the misclassification rate with imprecise data

If the input to a classifier is imprecise and its output is set-valued, the same can be said about its misclassification rate. Let $\{(c^i,\text{class}(\tilde{X}^i))\}_{i=1\ldots N}$, be a list of N pairs, where c_i is the true class of the ith object, and class $(\tilde{X}^i)$ is the set valued output of the classifier. According to [44], the misclassification rate is computed as follows for a classification problem with N_c classes: let

$$p = \arg\max_{c=1,\ldots,N_c} \mu_{\text{class}(\tilde{X}^i)}(c), \tag{17}$$

be the class label with higher membership value in the classifier output at the ith object, and be

$$q = \arg\max_{c=1,\ldots,N_c, c\neq p} \mu_{\text{class}(\tilde{X}^i)}(c), \tag{18}$$

the second largest one. The number of errors that the classifier commits at the ith object is either 0 or 1, however the knowledge about this number is a fuzzy set

$$\tilde{e^i} = \begin{cases} \{1|0, \mu_{\text{class}(\tilde{X}^i)}(q)|1\} & \text{if } c^i = p, \\ \{\mu_{\text{class}(\tilde{X}^i)}(c^i)|0, 1|1\} & \text{if } c^i \neq p. \end{cases} \tag{19}$$

Finally, the total misclassification rate is a fuzzy subset of $\{0, 1/N, 2/N, \ldots, 1\}$

$$\tilde{\varepsilon} = \frac{1}{N} \cdot \bigoplus_{i=1}^{N} \tilde{e^i}, \tag{20}$$

where the meaning of the fuzzy arithmetic operators is

$$\mu_{(\tilde{A}\oplus\tilde{B})}(x) = \max\{\min(\mu_{\tilde{A}}(u), \mu_{\tilde{B}}(v)) | u+v = x\}, \tag{21}$$

$$\mu_{(k\cdot\tilde{A})}(x) = \mu_{\tilde{A}}(x/k). \tag{22}$$

The numerical procedures for obtaining rule-based models from data are designed to optimize this value under different hypotheses, as discussed in the paragraphs that follow.

5.2. Obtaining rule bases from fuzzy data

Some of the most effective numerical techniques for obtaining fuzzy rule bases from data are based on genetic algorithms, so called "genetic fuzzy systems" (GFS) [6]. A recent branch of these algorithms studies how to obtain rules from imprecise data, as happens in this study [44]. Two different state-of-the art GFSs will be used. Both are able to exploit the information in vague data for eliciting a human readable, knowledge base comprising fuzzy rules. The first of these GFSs is called Genetic Cooperative-Competitive Learning (GCCL) [35,36] and evolves a population of rules so that the misclassification rate in the experimental data (see Eq. (20)) is minimum. The second one is a generalization of the Adaboost algorithm, that determines the weights of a set of rules optimizing the expectation of the misclassification rate [37].

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

5.2.1. Genetic Cooperative-Competitive Learning

GCCL is designed for finding a set of rules optimizing the fuzzy expression described in Eq. (20). Since the search space is very large, an exhaustive search is not feasible. Unfortunately, the properties of the objective function do not allow for significant shortcuts either, thus GCCL is based in two heuristic simplifications

- The number M of rules in the knowledge base is assumed known.
- Given the antecedent part of a rule ("If (x is $\tilde{A}_i$)"), the best consequent and weight for this antecedent only depend on those experimental cases for which this last expression is true.

As a consequence of these two assumptions, the search space is reduced to the free parameters of the M fuzzy sets describing the antecedents; the consequents and weights are not part of the search, as they will be computed from the experimental cases matching their corresponding antecedents (the calculations are described later). The output of the classifier is computed as shown in Eq. (5).

In turn, each chain can be formed as a combination of the symbols in a finite catalog of linguistic terms and connectives (for instance: "x_1 is SMALL and x_2 is LARGE and…"). Following the binary representation system described in [11], all the antecedents of a knowledge base can be represented in a bit chain, allowing for the exploration of the different rule-based classifier systems by means of a genetic algorithm.

The genetic search consists therefore in finding the bit chain that minimizes the misclassification rate defined in Eq. (20). There are some problems that must be addressed before this search can take place:

1. The concept of "minimum" depends on the definition of a total order among the fuzzy values of the misclassification rate, some of which cannot be directly compared.
2. How the best consequent and weight of a rule are produced for a given antecedent.
3. How to alleviate the computational complexity of determining the fuzzy set in Eq. (20), and in particular the intermediate step in Eq. (7).

The first issue is solved by redefining the objective of the problem: it is acknowledged that a minimum cannot be obtained, but a set of *non-dominated* classifiers can be produced. The second problem is solved by embracing and extending the definition in [27], which consists in selecting the consequent for which the number of misclassifications is the lowest, constrained to the elements that fulfill the condition "x is $\tilde{A}$". When the experimental data $\{(\tilde{X}^i, c^i)\}_{i=1}^N$ comprises fuzzy numbers, our knowledge about this number of correct classifications is

$$\widetilde{\text{hits}}(\tilde{A},c)(t) = \max\left\{ \min\{\mu_{\tilde{X}^i}(x_i)\}_{i=1}^N \,\middle|\, t = \sum_i \delta_{c^i}^c \mu_{\tilde{A}}(x_i) \right\}, \tag{23}$$

where $\delta_{c^i}^c = [c = c^i]$. The best class is selected by means of a fuzzy ranking [3] on the elements of the set $\{\widetilde{\text{hits}}(\tilde{A},c)\}_{c=0}^{N_c}$.

5.2.2. Boosting fuzzy rules from low quality data

A second technique for obtaining fuzzy rule based classifiers from data consists in regarding each fuzzy rule as a *weak learner*, and the knowledge base as an *ensemble* of learners. In other words, each rule

$$\text{if } (x \text{ is } \tilde{A}_i) \text{ then class is } C_i \text{ with weight } w_i, \tag{24}$$

is understood as a simple classifier. For an input x, the output of this classifier comprises a class label (which is always the same, C_i) and a degree of certainty in the classification, whose value is $w_i \cdot \tilde{A}_i(x)$. These classifiers are not very useful as isolated entities, but they can be combined in an *ensemble* that performs better than its constituent parts. In this case, the *boosting* [49,25] technique can be applied for obtaining the best set of weights $\{w_i\}_{i=1}^M$ for any given set of fuzzy rules. This is because the output of boosting-based ensemble is identical to the voting inference described in Eq. (6). It is expected that this procedure improves the accuracy of the preceding approach, however the results are less understandable [48]. This will be further discussed in the next section.

It is remarked that the combined output of this ensemble of classifiers is not intended to directly minimize the misclassification rate on a training set, as was the case with the GCCL algorithm. Boosting optimizes instead the exponential loss

$$\sum_{i=1}^N \exp(1-2\delta_{\text{class}(x^i)}^{c^i}), \tag{25}$$

where $\delta_a^b = [a = b]$. The ensemble for which this last value is minimum is also expected to have the least overall misclassification rate over the universe.

The Adaboost algorithm [19] solves this optimization problem iteratively. Let z be an instrumental variable, for simplifying the notation

$$z(\omega,x,w) = \omega \exp(w\mu_{\tilde{A}}(x)(1-2\delta_{C_i}^{c^j})). \tag{26}$$

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

Firstly, a weight $\omega^j = 1/N$ is assigned to each of the N instances. Secondly, a fuzzy set $\tilde{A}_i$ and a class label C_i, constituents of a weak classifier of the form

$$\text{if } (x \text{ is } \tilde{A}_i) \text{ then class is } C_i \text{ with weight } 1, \tag{27}$$

is searched, such that its exponential loss

$$Z(1) = \sum_{j=1}^{N} z(\omega^j, x^j, 1), \tag{28}$$

is minimum. A genetic algorithm is well suited for this task, using the same binary representation mentioned in the preceding section. The weight w_i of this rule is determined afterwards, as the value w_i minimizing

$$Z(w) = \sum_{j=1}^{N} z(\omega^j, x^j, w). \tag{29}$$

Thirdly, the weights of the N instances are updated according to the results of this classifier,

$$\omega^j_{(t+1)} = \frac{z(\omega^j_{(t)}, x^j, w_i)}{\sum_{j=1}^{N} z(\omega^j_{(t)}, x^j, w_i)}. \tag{30}$$

The process jumps to the second step, and loops until the desired number of fuzzy rules is eventually obtained.

When the training set comprises imprecise data, these steps are altered as follows:

1. Weight and loss of a rule: given a set of fuzzy weights $\{\tilde{\Omega}^j\}_{j=1}^{N}$, the weight of a rule is obtained by finding the minimum of the fuzzy-valued function $\tilde{Z}(w)$ defined below.

$$\mu_{\tilde{Z}(w)}(z) = \max\left\{ \min_{j=1}^{N} \{\mu_{\tilde{X}^j}(x^j), \mu_{\tilde{\Omega}^j}(\omega^j)\} \right. \tag{31}$$

$$\left. \text{such that } z = \sum_{j=1}^{N} z(\omega^j, x^j, w) \text{ and } \sum_{j=1}^{N} w^j = 1 \right\}. \tag{32}$$

   A fuzzy ranking is used to define an order among the values of $\tilde{Z}(w)$, whose minimal element w_i is found with a genetic algorithm, as done in the preceding section. The loss of a rule, used for searching the best unweighted rules, is the fuzzy set $\tilde{Z}(1)$.
2. The weights of the instances are updated after the inclusion of the ith rule as follows:

$$\tilde{\Omega}^j_{(t+1)}(\omega) = \max\{\min\{\mu_{\tilde{X}^j}(x), \mu_{\tilde{\Omega}^j_{(t)}}(u)\} \,|\, \omega = Kz(w_i, x, u)\}, \tag{33}$$

   where K is a normalization factor chosen so that the distance between the (fuzzy arithmetic-based) sum of all the weights $\Omega^j_{(t+1)}$ and the value 1 is as low as possible.

6. Numerical results: obtaining a rule-based model of the ice accretion strength in helicopter blades

In this section, the methodology proposed in this study is applied to a rotorcraft application, and the results compared to similar knowledge discovery techniques that are not based in imprecise probabilities.

Helicopter rotors are more susceptible to icing than fixed-wing vehicles. Ice accretion can be critically dangerous, as it can modify the vehicles aerodynamics, create excessive vibration, increase drag [52], and introduce ballistic concerns as thick ice layers sheds off. A passive ice-phobic coating that prevents ice formation is the ideal solution to helicopter rotor blade ice accretion, thus the search for ice-phobic materials for rotorcraft applications is ongoing.

To quantify the ice adhesion performance of novel "ice-phobic" coatings, many researchers have attempted to measure the shear adhesion strength of ice to these materials. The published data varies significantly, even for isotropic materials [38]. According to this source, the conditions governing ice accretion physics are: Liquid Water Concentration (LWC) of the cloud, Median Volume Diameter (MVD) of the super-cooled water droplets in the cloud, ambient temperature and impact velocity. There are authors that claim the surface roughness also influences the ice adhesion strength [47], but there is not a consensus about this dependence [38]. Computer models of rotorcraft icing are based on empirical data that is gathered in experiments under controlled conditions. The experiments that will be discussed in this work were developed in the Vertical Lift Research Center of Excellence at the Pennsylvania State University. This center has developed a new icing facility for rotorcraft icing research, named "Adverse Environment Rotor Test Stand" (AERTS) which is designed to generate an accurate icing cloud around test rotor [33].

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

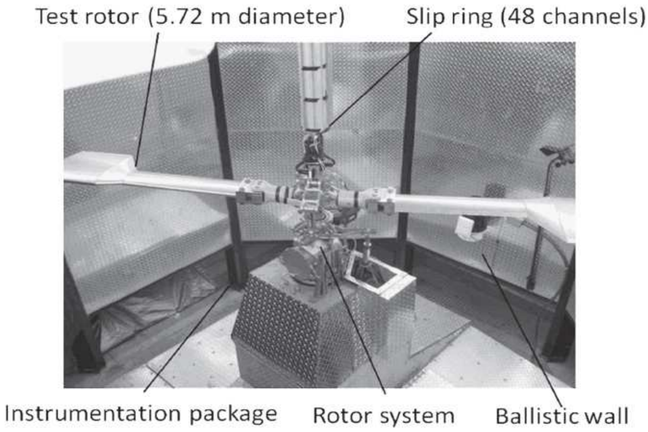

Fig. 3. Photograph of the AERTS facility. The AERTS Hub is collective and lateral cyclic capable. Max RPM:1000. Max. rotor diameter: 9 ft (2.75 m). Max. power: 120 HP.

6.1. Technical data of applied measuring devices

The AERTS facility is formed by an industrial 6 m $\times$ 6 m $\times$ 6 m cold chamber where temperatures between -25 °C and 0 °C can be achieved. The chamber floor is waterproofed with marine lumber covered by aluminum plating, and a drainage system in the perimeter of the room collects melted ice during the post-test defrosting process. Inside the chamber, and surrounding the rotor, there is a ballistic wall in the shape of an octagon. The ballistic wall is formed by 15.2 cm thick weather resistant lumber reinforced with 0.635 cm thick steel, and covered by aluminum plating for weather protection. A photograph of the chamber, as seen from above, is provided in Fig. 3. Convection lines and a set of fans located inside the chamber cool the facility.

A total of 15 NASA standard icing nozzles are located in the chamber ceiling to generate the icing cloud. The nozzles are similar to those used in the NASA IRT and Goodrich Icing Tunnel. The nozzles are arranged into two concentric circles located 50.8 cm and 106.6 cm from the center of the rotor shaft to distribute the cloud evenly in the chamber. The nozzles operate by aerosolizing water droplets with a combination of water and air as per nozzle calibration curves available in Ref. [26]. The water and air pressures are measured at the input of the water and air lines to the nozzles, which ensures precise readings of the pressure differential controlling the droplet size. The number of nozzles operating and the MVD of the water droplets dictate the LWC in the room. The water system is generally similar to the air system, with added complications in maintaining a constant and controllable supply of pure water. A series of pumps and a feedback control system is in place to maintain the water pressure at desired conditions.

In the center of the chamber, a 89.5 kW motor rotates the lower hub of a QH-50D DASH UAV vehicle. The configuration has the capability of reaching 1500 RPM with 1.37 m radius blades, reproducing full-scale helicopter tip speeds.

6.2. Variability of the data and control parameters

There are several parameters influencing the measurements that must be controlled in the ice system. According to [33], these are the active nozzles, temperature, MVD, LWC, icing time, water input temperature and water purity. However it is difficult to settle in a set of values for these parameters, as they are subject to slight changes during the tests. The reasons under this variability are detailed in the following paragraphs.

6.2.1. Temperature control issues

Temperature is arguably the most important parameter for icing testing [20]. In the AERTS system, three thermocouples are installed around the test chamber to monitor the chamber temperature. Two sensors are mounted near the rotor plane, and one is mounted on the rotor stand just below the rotor head. There rarely exists an agreement between the readings of the different thermocouples. Determining the actual temperature of the test chamber is a problem of information fusion [1,22]. Furthermore, the kinetic friction of the rotor and the input of warm water to the chamber alters the temperature, thus it makes sense to describe each experiment with two different magnitudes: beginning and end of test temperatures.

6.2.2. Mean volumetric diameter of the water droplets

The size of the water droplets coming from the nozzles is another important parameter. In the AERTS facility, droplet size is not directly measured. Instead, droplet size is based upon NASA standard nozzle calibration tables and experimental readings of pressure differentials between the water and the air inputs to the nozzles. To maintain the particle size it is necessary to adjust the water and air pressure during testing. A feedback control system monitors the nozzle input pressures and adjusts the water and air pressure to maintain particle size, as shown in Fig. 4. The adjustment made is

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

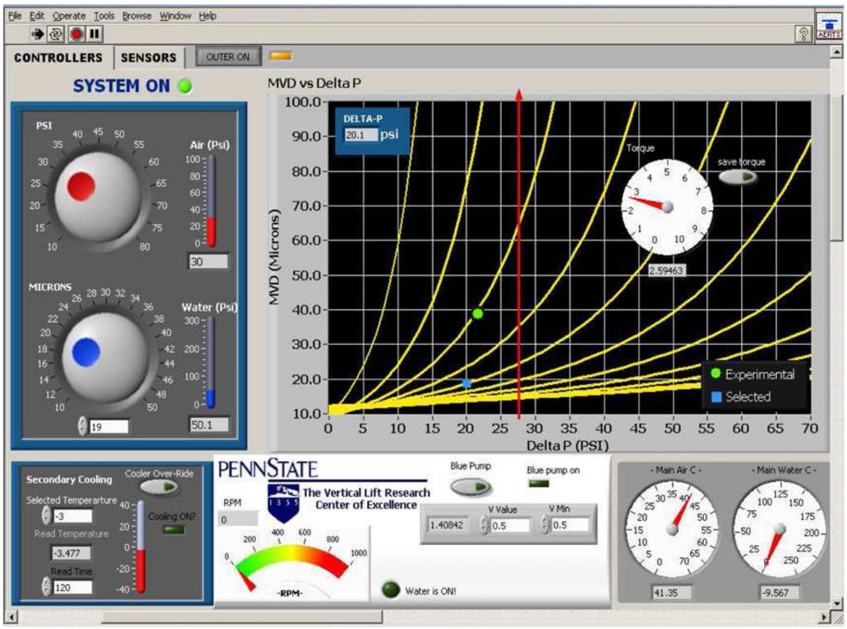

Fig. 4. Labview monitor: to adjust the water and air pressure to maintain particle size.

never exactly achieved neither maintained during the experiment and this implies that the droplet size can not be determined with an exact value and is estimated with a $\pm 15\%$ error.

6.2.3. *Liquid water content (LWC)*

LWC is the third parameter that cannot be directly measured or calculated during testing in the chamber. Static LWC sensors are not applicable to the facility because they require flow velocities of 15 m/s (49 ft/s) to determine the LWC value of a cloud. To provide these devices with proper operational velocity conditions, the LWC sensors should have been mounted on the blades. Due to the size and cost of these sensors, their rotation was not possible. Even if they could be rotated, centrifugal forces on the devices might impair their ability to accurately measure LWC [34].

Instead, this parameter is controlled by the number of active nozzles and input pressures, which are adjusted during the experiment, and its estimated value is calculated after each test based upon accreted ice thickness. LWC is calculated from total ice thickness per unit time (within $\pm 20\%$ error, see [34]). This is yet another source of variability in the models of ice adhesion shear force.

6.2.4. *Other factors*

The three mentioned parameters are responsible of most of the uncertainty in the test data, however there are many other factors with a smaller relevance, for instance the characteristics of the rotor or the rotation speed. The properties of the material can also influence the results. The material will be described by two parameters, namely the surface roughness and the Young's modulus. The surface roughness is measured by hand with a profilometer. The assumed value of this parameter is the average of the maximum and minimum of the measurements taken at the stagnation point of the coating. Lastly, the Young's modulus is the ratio between the linear stress and the linear strain for a given material. This value is not experimentally determined at AERTS facilities.

6.3. *Properties of the set of data*

The set of data describing the ice adhesion strength of helicopter rotor blades materials has been produced after repeated experiments with nickel, titanium and stainless steel. A threshold of 34.4 kPa (5PSI) has been selected for the ice adhesion strength. There are seven variables, some of which are imprecisely perceived, whose description follows:

1. Initial temperature (°C): This is the temperature measured by a thermocouple prior to the start of the icing cloud that promotes ice accretion. This temperature is read three times: when the experiment begins, when the rotor reaches the desired revolutions per minute (RPM) and before turning on the icing cloud (dry room). These three measurements of the initial temperature are combined into a fuzzy value with the procedure described in [43].
2. Final temperature (°C): This is the temperature read at the end of the test, once the ice sheds off from the rotor. The increase in temperature during testing is due to inability of the chamber to compensate for the increase in temperature

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

provided by kinetic friction of the rotor. Again, this temperature is defined by a fuzzy term provided by the aggregation of readings of several thermocouples.

3. Median Volume Diameter (MVD) of the water particle (μm): This is the size of the water droplets in the cloud (measured in micrometers). The size is calculated from calibration tables provided by NASA (NASA icing nozzles used). The nozzles work via atomization of water. The water and air pressure provided determine the water droplet size. Both water and air pressure are measured by pressure sensors located at the water and air input to the nozzles. Acceptable variability of MVD is $\pm 15\%$, thus an interval-valued variable is used. The width of these intervals is determined by a human expert.
4. Liquid Water Concentration (LWC) (g/m^3): This is the severity of the icing cloud. LWC quantifies how much water is there in an icing cloud. This quantity is not measured or directly calculated during testing. Test procedures for determining LWC are based on the measurement of ice thickness at the stagnation location of a blade for a given time interval. Acceptable variability of LWC is $\pm 20\%$ which implies again an interval-valued input whose width is determined by the expert.
5. Revolutions per minute (RPM) (cycles/min): This is the velocity of rotation of the rotor. The RPM of the rotor is measured by a Hall sensor that quantifies the rotational speed of the shaft. The measurements obtained from this sensor are combined into a fuzzy value. Shedding tests are usually conducted at 350 RPM to avoid severe imbalance issues when shedding is not symmetric, so the rotor RPM was limited to 350 (170 ft/s or 51.86 m/s tip speed) for safety reasons. The rotor was designed to achieve 800 RPM (380 ft/s or 115.82 m/s tip speed), but shedding rotor imbalance concerns limited operational RPM.
6. Surface roughness of tested material (μin): This is the surface roughness of the coating being tested. The surface roughness is measured by hand with a profilometer. The quantity provided is a fuzzy value obtained from the values measured at the stagnation point of the coating.
7. Young's modulus (Pa): This is the ratio of the linear stress to the linear strain for a given material, a crisp value.

6.4. *Granularity of the linguistic partitions*

The first stage in the proposed methodology consists of selecting an appropriate granularity for the variables. In Table 2 the Mutual Information Matrix (MIM), computed with the software provided in reference [32], is included. The last row measures the dependence between each design parameter and the outcome of the experiment (variable "C" for "Class"). It is remarked that there is a significant degree of epistemic uncertainty in these data that is being ignored for computing this initial matrix; each fuzzy value has been replaced by its modal point, and intervals were replaced by their centerpoints.

In Table 3 the same values have been produced by the QCH-based algorithm in Section 4, partitioning each variable into five linguistic terms (see Fig. 6). Observe in the table and in Fig. 5 that this discretization does not discard a relevant amount of information. In this last figure, the white circles are the values of the crisp MIM, and the black circles, along with their corresponding intervals, are the estimations obtained after the application of the new procedure. In all cases, the bands of the new estimation contain or are near the crisp estimations, thus it can be concluded that the discretization into five terms does not discard a relevant amount of information.

6.5. *Learning fuzzy classification rules*

In Table 4 a selection of rule-based models, built upon these terms, and learnt with different algorithms, is shown for the ice accretion strength problem. These models are labelled "crisp data" when they have been obtained after removing the uncertainty of the conflicting measurements, as discussed in the preceding section.

The genetic algorithms have been run with a population size of 100, probabilities of crossover and mutation of 0.9 and 0.1, respectively, and limited to 150 generations. The fuzzy partitions of the labels are uniform and their granularity is the minimum value without a significant loss of information (measured by means of the generalized mutual information defined in the preceding sections). All the imprecise experiments were repeated 100 times with bootstrapped resamples of the training set. Each test set contains 1000 bootstrap resamples. Membership functions were not learned or tuned for

Table 2
Mutual information matrix between the modal points of all the variables in the study, using the algorithm described in [32].

Variable	Tini	Tfin	MVD	LWC	RPM	Rough	Young
Tfin	1.02						
MVD	0.25	0.33					
LWC	0.18	0.23	0.16				
RPM	0.22	0.18	0.19	0.15			
Rough	0.22	0.23	0.24	0.16	0.11		
Young	0.08	0.03	0.13	0.12	0.08	0.08	
C	0.25	0.31	0.12	0.09	0.06	0.05	0.01

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

Table 3
Mutual information matrix computed from ill-known data by means of the a QCH-based estimation and granularity five, as proposed in this paper.

Variable	Tini	Tfin	MVD	LWC	RPM	Rough	Young
Tfin	[0.83 0.93]						
MVD	[0.17 0.26]	[0.23 0.34]					
LWC	[0.12 0.18]	[0.18 0.25]	[0.11 0.18]				
RPM	[0.12 0.21]	[0.16 0.24]	[0.15 0.23]	[0.09 0.16]			
Rough	[0.13 0.20]	[0.18 0.26]	[0.20 0.27]	[0.13 0.19]	[0.14 0.22]		
Young	[0.06 0.11]	[0.09 0.15]	[0.08 0.14]	[0.09 0.13]	[0.09 0.16]	[0.18 0.26]	
C	[0.27 0.33]	[0.29 0.32]	[0.06 0.11]	[0.06 0.10]	[0.03 0.08]	[0.05 0.09]	[0.02 0.02]

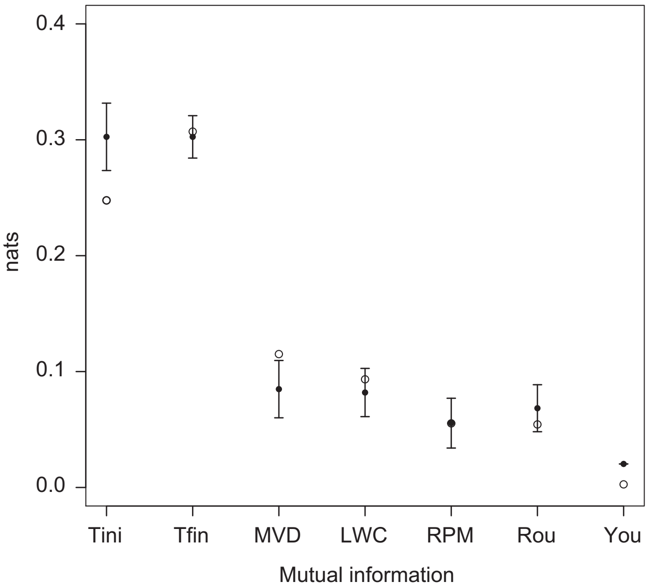

Fig. 5. Graphical comparison of the mutual information between the outcome of the experiment and the different features. White circles: estimation from the modal points of the data with the algorithm in [32]. Black circles and vertical bars: expectation of the mutual information computed with a QCH-based estimation and granularity five, as proposed in this paper, showing that the linguistic discretization is not losing a significant amount of information.

preserving their linguistic interpretability, however rules have multiplicative weights, that compensate for the absence of tuning [42].

The following conclusions can be drawn:

1. The use of boosting is preferred over Genetic Cooperative-Competitive Learning (GCCL), as the misclassification rate of the former algorithm is better for both crisp and fuzzy data and the readability of both GCCL and boosting is similar for the problem at hand.
   In Fig. 7 there is a boxplot showing that the accuracy of the model produced by the boosting algorithm for low quality data (p-LQD_Boos algorithm) is better than that of Genetic Cooperative-Competitive Learning for Low Quality Data algorithm (LQD_GCCL), in both crisp and fuzzy data.
2. The results are noticeably better when fuzzy data is used, in both accuracy (as shown in the preceding boxplot) and understandability of the model (5% less linguistic terms, on average).
   For supporting this last claim, in Table 5 two examples are included where the information provided by the fuzzy data-based approach about temperature and roughness is preferred. In these cases, the use of fuzzy data allows concluding that the nickel is a valid material when the temperature is fairly low, independently of the roughness (rules #9 of the fuzzy model and #7 of the crisp model). On the contrary, if crisp data is used, the roughness appears as a relevant factor, contradicting recent works [38]. In turn, if rules #9 of the fuzzy model and #8 of the crisp model are compared, the fuzzy model predicts that any material is valid when the final temperature is fairly low, again independently of the roughness. The crisp model wrongly predicts that this behavior only happens with nickel.

Summarizing the knowledge elicited from the experimental data, the ice shear adhesion strength grows when either the temperature decreases or the roughness is higher. Stainless steel should be discarded unless the temperature is very high and the roughness is low. Nickel is the most appropriate material, improving titanium, and therefore it should

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

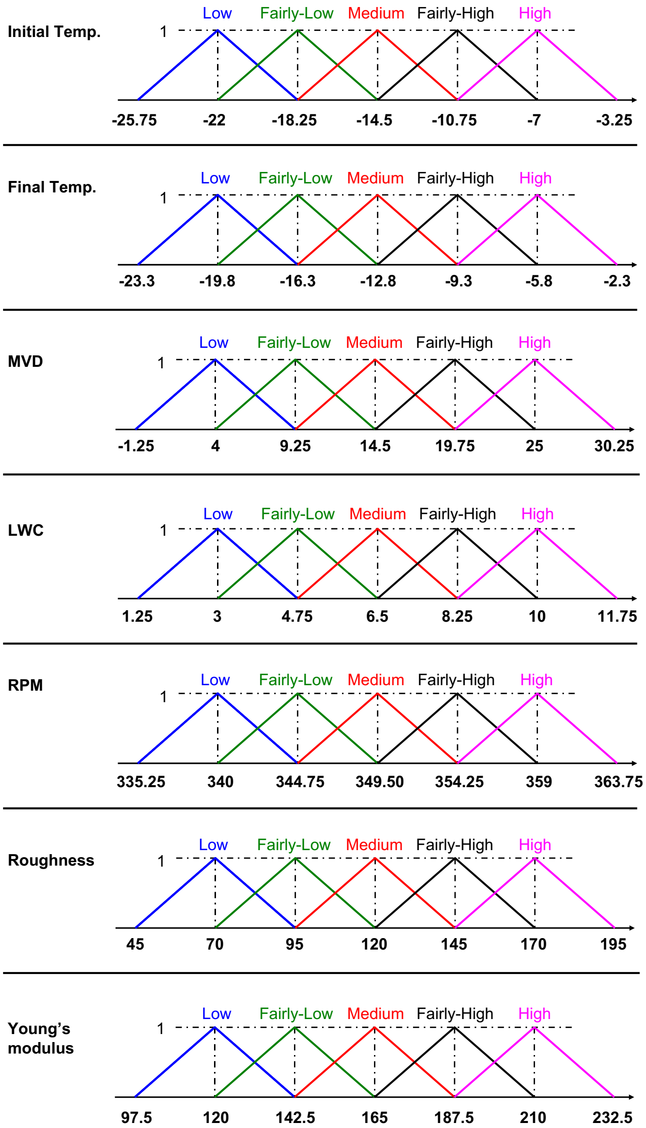

Fig. 6. Fuzzy sets defining the meaning of the linguistic terms in the knowledge base.

become the baseline for all future tests. Interestingly enough, some of the results obtained by the learning algorithm agree with previous works in the field (see [38]) and there are facts that have been discovered by the artificial intelligence-based method proposed herein that were not noticed by the experts in these studies.

7. Concluding remarks

Learning linguistic rules from empirical data allows the engineer to make predictions about the future performance of a product without the need of extensive experimentation. Obtaining precise experimental data is a time-consuming and costly process, thus there is interest for algorithms that do not discard data with missing or ill-perceived values. In this paper it was proposed a methodology for obtaining classification rules from low quality data, whose epistemic uncertainty is modeled by means of possibility distributions. The loss of information in the linguistic discretization of the features was measured with a novel measure of mutual information, based upon the use of Quasi-Continuous Histograms.

The outcome of this methodology, when applied to the prediction of ice accretion strength in helicopter rotors, is a human-understandable model that predicts whether a material is suitable or not for its use, as a function of the desired environmental and icing conditions. By using the imprecise data collected in 42 experiments it has been confirmed that the outcomes of the linguistic models learned from possibilistic data improve those of statistical and machine learning techniques that assume stochastic imprecision in the variables.

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

Table 4
Rule base obtained for the dataset "Ice_Shedding" with the algorithms "p-LQD_Boos" and "LQD_GCCL", when the imprecise data in all the input fields is replaced by the mean of the measurements in conflict. Interval-valued errors are understood as the margin between the prediction errors in the best and worst cases.

Algorithm and $\overline{error}$	**Rules**
p-LQD_Boos imprecise data, $\overline{error}=[\mathbf{0.228},\mathbf{0.297}]$	R1: If initial temp. is low and final temp. is low and RPM is medium then class is no R2: If final temp. is low and RPM is fairly high and roughness is fairly low then class is no R3: If LWC is fairly low and RPM is fairly high and roughness is fairly high and Young's modulus is low then class is no R4: If final temp. is low and roughness is fairly high and Young's modulus is low then class is no R5: If final temp. is high and MVD is low and LWC is fairly low and roughness is medium and Young's modulus is high then class is no R6: If initial temp. is medium and final temp. is high and LWC is medium and RPM is medium and roughness is fairly low then class is yes R7: If MVD is medium and roughness is fairly low then class is yes R8: If final temp. is fairly high and LWC is medium and RPM is medium and Young's modulus is fairly high then class is yes R9: If final temp. is fairly low and MVD is medium and LWC is medium and Young's modulus is fairly high then class is yes R10: If final temp. is high and MVD is fairly high and LWC is fairly low and Young's modulus is low then class is yes
p-LQD_Boos crisp data, $\overline{error}=[\mathbf{0.292},\mathbf{0.361}]$	R1: If final temp. is low and LWC is fairly high and RPM is fairly high and Young's modulus is high then class is no R2: If initial temp. is fairly high and final temp. is fairly high and LWC is high and RPM is high and Young's modulus is high then class is no R3: If initial temp. is low and MVD is fairly low then class is no R4: If RPM is fairly high and roughness is low and Young's modulus is high then class is no R5: If final temp. is medium and RPM is fairly high and Young's modulus is high then class is no R6: If final temp. is fairly high and roughness is fairly low then class is yes R7: If initial temp. is fairly low and LWC is high and RPM is high and roughness is high and Young's modulus is fairly high then class is yes R8: If initial temp. is fairly Low and final temp. is failry low and MVD is fairly high and RPM is and high then class is yes R9: If final temp. is medium and LWC is fairly low and RPM is high and roughness is low and Young's modulus low is then class is yes R10: If initial temp. is fairly high and final temp. is fairly high and MVD is fairly low and roughness is high then class is yes
LQD_GCCL imprecise data, $\overline{error}=[\mathbf{0.470},\mathbf{0.545}]$	R1: If initial temp. is fairly low and final temp. is fairly high and MVD is low and LWC is medium and RPM is medium and roughness is fairly low and Young's modulus is medium then class is yes R2: If initial temp. is medium and final temp. is fairly high and MVD is fairly high and LWC is medium and RPM is medium and roughness is fairly high and Young's modulus is medium then class is no R3: If initial temp. is fairly low and final temp. is medium and MVD is medium and LWC is fairly low and RPM is medium and roughness is fairly low and Young's modulus is medium then class is yes R4: If initial temp. is high and final temp. is high and MVD is fairly low and LWC is medium and RPM is fairly low and roughness is medium and Young's modulus is low then class is no R5: If initial temp. is medium and final temp. is fairly high and MVD is fairly high and LWC is medium and RPM is medium and roughness is fairly low and Young's modulus is medium then class is yes R6: If initial temp. is low and final temp. is low and MVD is low and LWC is medium and RPM is medium and roughness is fairly low and Young's modulus is medium then class is no R7: If initial temp. is fairly low and final temp. is failry low and MVD is medium and LWC is medium and RPM is medium and roughness is fairly low and Young's modulus Medium is then class is yes R8: If initial temp. is fairly low and final temp. is medium and MVD is medium and LWC is fairly low and RPM is medium and roughness is fairly high and Young's modulus is medium then class is no
LQD_GCCL crisp data, $\overline{error}=[\mathbf{0.507},\mathbf{0.576}]$	R1: If initial temp. is high and final temp. is high and MVD is fairly low and LWC is fairly high and RPM is fairly low and roughness is medium and Young's modulus is low then class is no R2: If initial temp. is fairly low and final temp. is medium and MVD is medium and LWC is fairly high and RPM is medium and roughness is fairly low and Young's modulus is medium then class is yes R3: If initial temp. is medium and final temp. is medium and MVD is medium and LWC is fairly high and RPM is failry low and roughness is high and Young's modulus is medium then class is no R4: If initial temp. is low and final temp. is low and MVD is fairly low and LWC is fairly high and RPM is medium and roughness is fairly low and Young's modulus is medium then class is no R5: If initial temp. is fairly high and final temp. is fairly high and MVD is fairly high and LWC is fairly high and RPM is fairly low and roughness is fairly low and Young's modulus is medium then class is yes R6: If initial temp. is low and final temp. is low and MVD is fairly high and LWC is fairly high and RPM is medium and roughness is fairly low and Young's modulus is medium then class is no R7: If initial temp. is fairly low and final temp. is medium and MVD is medium and LWC is fairly high and RPM is medium and roughness is low and Young's modulus is medium then class is yes

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

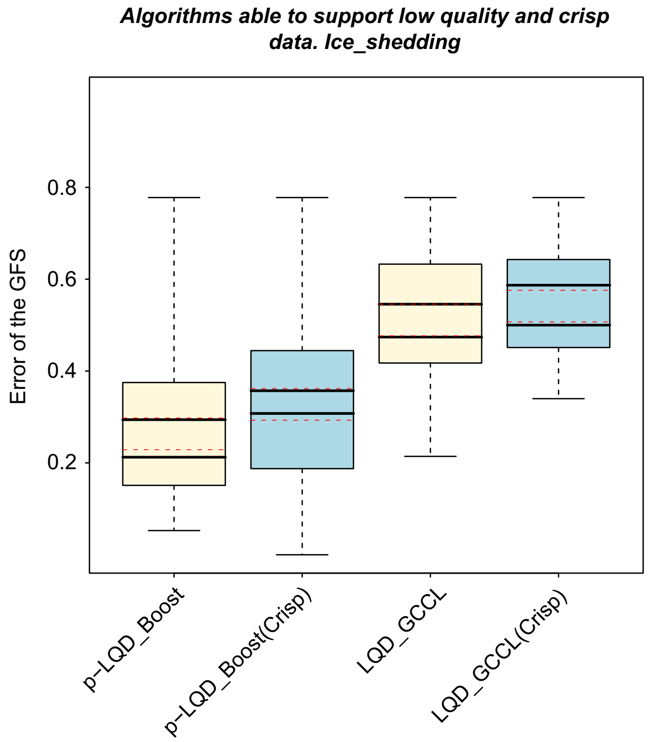

Fig. 7. Behaviour of "p-LQD_Boos" and "LQD_GCCL" in the dataset "Ice_Shedding" with fuzzy and crisp data.

Table 5
Comparison between rules obtained with "p-LQD_Boos" fuzzy and crisp data.

Imprecise data	Crisp data
R9: If final temp. is fairly low and MVD is medium and LWC is medium and Young's modulus is fairly high then class is yes	R7: If initial temp. is fairly low and LWC is high and RPM is high and roughness is high and Young's modulus is fairly high then class is yes
R9: If final temp. is fairly low and MVD is medium and LWC is medium and Young's modulus is fairly high then class is yes	R8: If initial temp. is fairly low and final temp. is failry low and MVD is fairly high and RPM is and high then class is yes

Acknowledgments

This study has been supported by the Spanish Ministry of Science and Technology, projects TIN2008-06681-C06-04 and TIN2011-24302.

References

[1] C.L. Bowman, A.N. Steinberg, Systems engineering approach for implementing data fusion systems, in: M.E. Liggins, D.L. Hall, J. Llinas (Eds.), Handbook of Multisensor Data Fusion: Theory and Practice, 2nd ed., CRC Press, Boca Raton, 2009.
[3] J.J. Buckley, Ranking alternatives using fuzzy numbers, Fuzzy Sets Syst. 15 (1) (1985) 21–31.
[4] J. Casillas, O. Cordon, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy Modeling, Springer, Berlin, 2003.
[5] M. Colyvan, Is probability the only coherent approach to uncertainty? Risk Anal 28 (3) (2008) 645–652.
[6] O. Cordón, F. Herrera, F. Hoffmann, L. Magdalena (Eds.), Genetic Fuzzy Systems: Evolutionary Tuning and Leaning of Fuzzy Knowledge Bases, World Scientific Publishing Company, Singapore, 2001.
[7] I. Couso, S. Montes, P. Gil, The necessity of the strong α-cuts of a fuzzy set, Int. J. Knowl. Based Syst. 9 (2) (2001) 249–262.
[8] I. Couso, L. Sánchez, Higher order models for fuzzy random variables, Fuzzy Sets Syst. 159 (2008) 237–258.
[9] I. Couso, E. Miranda, G. de Cooman, A possibilistic interpretation of the expectation of a fuzzy random variable, in: M. López Díaz, M.A. Gil, P. Grzegorzewski, O. Hryniewicz, J. Lawry (Eds.), Soft Methodology and Random Information Systems, Springer, 2004, pp. 133–140.
[10] P. Domingos, M.J. Pazzani, On the optimality of the simple Bayesian classifier under zero-one loss, Mach. Learn. 29 (2–3) (1997) 103–130.
[11] A. González, R. Pérez, Selection of relevant features in a fuzzy genetic learning algorithm, IEEE Trans. Syst. Man Cybern. 31 (3) (2001) 417–425.
[12] E. Hüllermeier, Fuzzy methods in machine learning and data mining: status and prospects, Fuzzy Sets Syst. 156 (3) (2005) 387–406.
[13] D. Dubois, H. Prade, The mean value of a fuzzy number, Fuzzy Sets Syst. 24 (3) (1987) 279–300.
[14] D. Dubois, H. Prade, The three semantics of fuzzy sets, Fuzzy Sets Syst. 90 (2007) 141–150.
[15] D. Dubois, H. Prade, Formal representations of uncertainty, in: D. Bouyssou, D. Dubois, M. Pirlot, H. Prade (Eds.), Decision-making Process-concepts and Methods, ISTE/Wiley, London, 2009, pp. 85–156. (Chapter 3).
[16] D. Dubois, D. Guyonnet, Risk-informed decision-making in the presence of epistemic uncertainty, Int. J. Gen. Syst. 40 (2) (2011) 145–167.

Please cite this article as: L. Sánchez, et al., A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials, Mech. Syst. Signal Process. (2012), doi:10.1016/j.ymssp.2012.02.009

[17] S. Ferson, L.R. Ginzburg, Different methods are needed to propagate ignorance and variability, Reliab. Eng. Syst. Saf. 54 (1996) 133–144.
[18] S. Ferson, L. Ginzburg, V. Kreinovich, D.M. Myers, K. Sentz, Construction of Probability Boxes and Dempster–Shafer Structures, Technical Report SANDD2002-4015, Sandia National Laboratories, 2003.
[19] Y. Freund, R. Schapire, Experiments with a new boosting algorithm, in: Proceedings of 13th International Conference Machine Learning, 1996, pp. 148–156.
[20] R.W. Gent, N.P. Dart, J.T. Candsdale, Aircraft Icing, Philos. Trans. R. Soc. London Ser. A 358 (2000) 2873–2911.
[21] D. Guyonnet, B. Côme, P. Perrochet, A. Parriaux, Comparing two methods for addressing uncertainty in risk assessments, J. Environ. Eng. 125 (7) (1999) 660–666.
[22] D.L. Hall, S.A.H. McMullen, Mathematical Techniques in Multisensor Data Fusion, 2nd ed., Artech House, Northwood, 2004.
[23] J.C. Helton, J.D. Johnson, W.L. Oberkampf, An exploration of alternative approaches to the representation of uncertainty in model predictions, Reliab. Eng. Syst. Saf. 85 (1–3) (2004) 39–71.
[24] F.O. Hoffman, J.S. Hammonds, Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability, Risk Anal. 14 (1994) 707–712.
[25] F. Hoffman, Boosting a genetic fuzzy classifier, in: Proceedings of Joint 9th IFSA World Congr. 20th NAFIPS Int. Conf., vol. 3, 2001, pp. 1564–1564.
[26] R. Ide, J. Oldenburg, Icing Cloud Calibration of the NASA Glenn Icing Research Tunnel, AIAA-2001-0234, 2001.
[27] H. Ishibuchi, T. Nakashima, T. Murata, A fuzzy classifier system that generates fuzzy if then rules for pattern classification problems, in: Proceedings of 2nd IEEE International Conference on Evolutionary Computation, 1995, pp. 759–764.
[28] K. Itagaki, Self-shedding of Accreted Ice from High Speed Rotors, ASME 83-WA/HT-68, US Army Cold Regions Research and Engineering Laboratory, 1983.
[29] L.I. Kuncheva, Fuzzy Classifier Design, Springer-Verlag, NY, 2000.
[30] D. Loughborough, E. Hass, Reduction of adhesion of ice to de-icer surfaces, J. Aeronaut. Sci. 13 (1946) 126–134.
[31] G. Mauris, V. Lasserre, L. Foulloy, Fuzzy modeling of measurement data acquired from physical sensors, IEEE Trans. Instrum. Meas. 49 (6) (2000) 1201–1205.
[32] P.E. Meyer, Information-theoretic Variable Selection and Network Inference from Microarray Data, PhD thesis of the Universite Libre de Bruxelles, 2008.
[33] J.L. Palacios, E.W. Brouwers, Y. Han, E.C. Smith, Adverse environment rotor tests stand calibration and ice shape correlation, in: 66th Annual Forum Proceedings—American Helicopter Society, 2010.
[34] Y. Han, J. Palacios, E. Smith, An experimental correlation between rotor test and wind tunnel ice shapes on NACA 0012 airfoils, SAE Technical Paper 2011-38-0092, (2011), doi:10.4271/2011-38-0092.
[35] A. Palacios, L. Sánchez, I. Couso, Extending a simple cooperative-competitive learning fuzzy classifier to low quality datasets, Evol. Intell. 2 (1–2) (2010) 73–84.
[36] A. Palacios, L. Sánchez, I. Couso, Diagnosis of dyslexia with low quality data with genetic fuzzy systems, Int. J. Approx. Reason. 51 (2010) 993–1009.
[37] A. Palacios, L. Sánchez, I. Couso, Boosting of fuzzy rules with low quality data, J. Multiple Valued Logic Soft Comput., in press.
[38] E.W. Brouwers, A.A. Peterson, J.L. Palacios, L. Centolanza, Ice adhesion strength measurements for rotor blade edge materials, in: 67th Annual Forum Proceedings, American Helicopter Society, 2011.
[39] L. Paninski, Estimation of entropy and mutual information, Neural Comput. 15 (2003) 1191–1253.
[40] A. Reich, Interface influences upon ice adhesion to airfoil materials, in: 32nd Aerospace Sciences Meeting and Exhibit, Paper AIAA 94-0714, 1994.
[41] E. Ruspini, New experimental results in fuzzy clustering, Inf. Sci. 6 (1973) 273–284.
[42] L. Sánchez, J. Otero, Tuning fuzzy partitions or assigning weights to fuzzy rules: which is better? in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Accuracy Improvements in Linguistic Fuzzy Modeling, Physica-Verlag, 2003, pp. 366–386.
[43] L. Sánchez, I. Couso, J. Casillas, Modelling vague data with genetic fuzzy systems under a combination of crisp and imprecise criteria, in: Proceedings of 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, 2007, pp. 30–37.
[44] L. Sánchez, I. Couso, Advocating the use of imprecisely observed data in genetic fuzzy systems, IEEE Trans. Fuzzy Syst. 15 (4) (2007) 551–562.
[45] L. Sánchez, M.R. Suarez, J.R. Villar, I. Couso, Mutual information-based feature selection and partition design in fuzzy rule-based classifiers from vague data, Int. J. Approx. Reason. 49 (3) (2008) 607–622.
[46] L. Sánchez, I. Couso, J. Casillas, Genetic learning of fuzzy rules based on low quality data, Fuzzy Sets Syst. 160 (17) (2009) 2524–2552.
[47] R.J. Scavuzzo, M.L. Chu, Structural Properties of Impact Ices Accreted on Aircraft Structures, NASA CR 179580, 1987.
[48] R. Schapire, Y. Freund, P. Bartlett, W.S. Lee, Boosting the margin: a new explanation for the effectiveness of voting methods, Ann. Stat. 26 (5) (1998) 1651–1686.
[49] R. Schapire, The boosting approach to machine learning: an overview, in: MSRI Workshop on Nonlinear Estimation and Classification, 2001.
[50] J.R. Stallabrass, R.D. Price, On the Adhesion of Ice to Various Materials, LR-350, National Research Laboratory, 1962.
[51] O. Strauss, Quasi-continuous histograms, Fuzzy Sets Syst. 160 (17) (2009) 2442–2465.
[52] T. Withington, Bad vibrations, Aerosaf. World Mag. Helicopter Saf. Flight Saf. Found. (2010) 26–28.