Uncertain Fuzzy Reasoning: A Case Study in Modelling Expert Decision Making

Jonathan M. Garibaldi and Turhan Ozen

***Abstract*—This paper presents a case study in which the introduction of vagueness or uncertainty into the membership functions of a fuzzy system was investigated in order to model the variation exhibited by experts in a medical decision-making context. A conventional (type-1) fuzzy expert system had previously been developed to assess the health of infants immediately after birth by analysis of the biochemical status of blood taken from infants' umbilical cords. Variation in decision making was introduced into the fuzzy expert system by means of membership functions which altered in small, predetermined manners over time. Three types of variation in membership functions were investigated: i) variation in the centre points, ii) variation in the widths, and iii) the addition of "white noise." Different levels (amounts) of uniformly distributed random variation were investigated for each of these types. Monte Carlo simulations were carried out to propagate the variation through the inferencing process in order to determine distributions of the conclusions reached. Interval valued type-2 fuzzy systems were also implemented to investigate the boundaries of variability in decisions. The results obtained were compared to the experts' decisions in order to determine which type and size of membership function variability best matched the experts' variability. The novel reasoning technique introduced in this study is termed *nonstationary fuzzy reasoning*.**

***Index Terms*—Interval type-2 fuzzy expert systems, medical decision making, nonstationary fuzzy reasoning, nonstationary type-1 fuzzy expert systems, umbilical acid-base assessment.**

I. Introduction

THE purpose of developing an expert system (whether based on fuzzy logic or not) is to encapsulate knowledge and expertise of human experts in the particular domain. However, human experts exhibit variation in their decision making. Variation may occur among the decisions of a panel of human experts (interexpert variability), as well as in the decisions of an individual expert over time (intraexpert variability). Up to now it has been an implicit assumption that expert systems, including fuzzy expert systems, should *not* exhibit such variation. Indeed, the property of being 100% consistent—always giving the same answer (outputs) when presented with the same data (inputs)—has been advocated as a key advantage of computerised expert systems. Understanding the dynamics of the variation in human decision making could allow the creation of expert systems that cannot be differentiated from their human counterparts. Such expert systems might be used, for example, in acceptance testing whereby an expert system could be accepted if it cannot be distinguished from human experts in its particular domain of decision making. Moreover, in application areas where it is not possible to have experts constantly available, such systems could produce the span of decisions that might have been arrived at by a panel of experts. Techniques of group decision making or consensus modelling might then be utilised to reach a single overall decision.

Modelling the variation in decision making was studied by enhancing a conventional type-1 fuzzy expert system (FES) developed in earlier work by Garibaldi *et al.* [1] which is able to assess the health of infants immediately after birth by analysis of the biochemical status of blood taken from an infant's umbilical cord (umbilical acid-base analysis). The rules of the original FES were elicited in conjunction with several experts who took part in its development. The fuzzy expert system produces a single, arbitrarily scaled index in each case which represents an overall assessment of the health of the infant. A database of fifty difficult to interpret cases had previously been collected and six experts in umbilical acid-base analysis had ranked these cases from "worst" to "best." The fuzzy expert system was run on these fifty cases and its output health index was used to obtain a similar rank order. However, when presented with the same data, although the FES produced the same output each time the same input was given, it was observed that the experts' conclusions varied both among themselves and from their previous conclusions. The source of this variation was suspected to lie in minor variations of the interpretations of linguistic terms used in the rules.

The original FES used type-1 membership functions which are precise and so do not reflect the vagueness in the terms that they represent. However, such terms may have different meanings for different experts and their interpretations may also vary depending on the environmental conditions or over time. In order to model this variation, experiments were carried out by introducing small changes in the membership functions associated with the linguistic terms in order to explore the relationship between the vagueness of these term and the variation in the overall decision making (the conclusions reached by means of the fuzzy inferencing and defuzzification processes). The term "nonstationary" was introduced to describe such time-varying fuzzy systems. The aim of this ongoing research is to investigate the process of modelling the variation in human decision making utilising a fuzzy expert system, by means of a case study in the context of medical decision making. To the best of the authors' knowledge, no other studies have been reported in the literature which explore the use of fuzzy expert systems in modelling the variability in human decision making. Preliminary investigations for determining variation in the results of fuzzy inferencing that can be obtained through variation of membership function parameters were presented in [2]–[5].

Manuscript received February 1, 2005; revised December 18, 2005, March 20, 2006, and June 16, 2006. This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant GR/R55085/02.

The authors are with the Automated Scheduling, Optimisation and Planning Research Group School of Computer Science and IT, the University of Nottingham, Nottingham NG8 1BB, U.K. (e-mail: jmg@cs.nott.ac.uk).

Digital Object Identifier 10.1109/TFUZZ.2006.889755

Fuzzy systems usually employ type-1 fuzzy sets, in which the membership of any element in the set is represented by a number in the range [0,1], referred to as the degree of membership. In some ways the assignment of a single precise number in [0,1] means that there is no "fuzziness" in a type-1 fuzzy set. The concept of type-2 fuzzy sets was introduced by Zadeh [6], as an extension of type-1 fuzzy sets, in which an additional dimension is introduced that represents the uncertainty in the degrees of membership. In effect, the membership of any element in the original fuzzy set (the primary membership) is itself a fuzzy set (the secondary membership). Mizumoto and Tanaka studied the set theoretic operations of type-2 fuzzy sets and properties of membership degrees of such sets [7]; they also examined type-2 fuzzy sets under the operations of algebraic product and algebraic sum [8]. Dubois and Prade gave a formula for the composition of type-2 relations as an extension of the type-1 sup-star composition for the minimum t-norm [9]. Hisdal studied rules and interval sets for higher-than-type-1 fuzzy logic [10]. Karnik and Mendel obtained algorithms for performing union, intersection, and complement for type-2 fuzzy sets, and developed the concept of the centroid of a type-2 fuzzy set [11]. Karnik *et al.* presented a general formula for the extended sup-star composition of type-2 relations [12]. Liang and Mendel developed the theory for different kinds of fuzzifiers for interval type-2 FESs [13]. Mendel and John have developed a simple method to derive union, intersection and complement of type-2 fuzzy sets without having to use Zadeh's extension principle [14].

The organisation of this paper is as follows. Details of the proposed manner in which variation may be introduced into membership functions are given in Section II. Modifications that were necessary to work with the nonstationary membership functions are explained and details of the interval type-2 FES are given. In Section III, the details of the medical context of the medical decision-making scenario are outlined and the design and evaluation of the original type-1 FES are presented. The experiments into uncertain fuzzy reasoning that were carried out are described and the results obtained are presented. It is demonstrated that there is a direct relationship between the uncertainty in the linguistic terms used and the variation in decision making obtained, Finally, the systems that closest match the actual variation exhibited by the human experts are examined. The findings of the study, general conclusions that may be drawn and potential avenues for future work are discussed in Sections IV and V.

II. Uncertain Fuzzy Reasoning

A. Introducing Vagueness Into Membership Functions

The aim of developing expert systems is, generally, to model the decision making of experts. However, traditionally this has not extended to modelling the *variation* found in human expert decision making. Any conventional type-1 FES will always provide the same output(s) when supplied with the same input(s). In order to attempt to create an FES which exhibited the same intraexpert variation as exhibited by the human experts in this domain, a variety of methods for introducing variation into (only) the membership functions of the FES were investigated. As the rules had been elicited from human experts, it was decided that, in this study, the rules of the system should remain fixed and that small changes (or perturbations) of the membership functions would be investigated (although, clearly, fuzzy rules only have meaning in the context of the membership functions).

The effect of introducing vagueness to the membership functions was explored in two complementary methods

1) through a novel "nonstationary" FES;
2) through an interval type-2 FES.

The first method resulted in a type-1 FES in which the output fuzzy sets altered each time the system was run, such that the output centroids varied. This permitted multiple runs of the FES to be carried out in order to explore the distribution of the resulting "decisions." The second method resulted in a type-2 FES in which the output was a type-2 fuzzy set which could be type-reduced to provide the lower and upper limit of output centroid. In effect, this allowed the boundaries of the resulting "decisions" to be explored.

B. Type-1 Fuzzy Systems

An FES normally consists of four main interconnected components: a fuzzifier, a set of rules, an inference engine, and an output processor (defuzzifier). Once the rules are established, an FES can be viewed as a (non-linear) mapping from inputs to outputs. Rules are provided by experts or can be extracted from numerical data. In either case they are expressed as a collection of IF–THEN statements. A typical rule is like

IF arterial pH is *low* and venous pH is *low*;
THEN acidemia is *severe*.

In an FES, the linguistic terms in the rules, shown in italics above, (the inputs to and the outputs from the FES) are associated with fuzzy sets. Despite the notion of uncertainty that their name suggests, fuzzy expert systems, like classical expert systems, are in fact entirely deterministic in the sense that, given the same inputs, the same outputs are always obtained. Of course, in general, many different functions may be used to generate type-1 fuzzy sets, with Gaussians, for example, being a common choice. In this work, sigmoids and products of sigmoids were used. Left-edge fuzzy sets may be generated by decreasing sigmoids

$$\mu(x) = \frac{1}{1+e^{(x-c)/w}} \tag{1}$$

right-edge fuzzy sets may be generated by increasing sigmoids:

$$\mu(x) = \frac{1}{1+e^{(c-x)/w}} \tag{2}$$

where c is the "center" of the sigmoid (the value of x at which the membership value $\mu(x)$ is 0.5) and w is an arbitrary value that determines the steepness of the slope, or "width," of the sigmoid. Middle fuzzy sets may be generated by the product of two such sigmoids

$$\mu(x) = \left(\frac{1}{1+e^{(c_1-x)/w}}\right)\left(\frac{1}{1+e^{(x-c_2)/w}}\right)$$

where c_1 is the centre of the increasing sigmoid, c_2 is the centre of the decreasing sigmoid, and w is the steepness of the slopes,

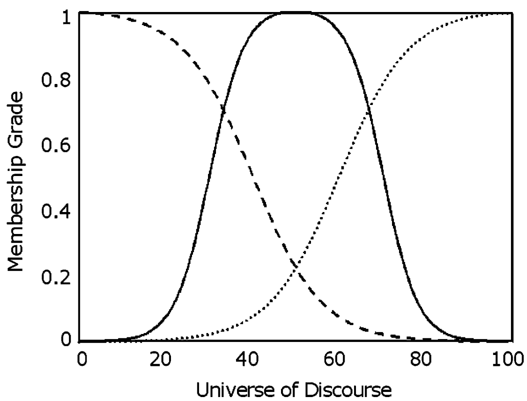


Fig. 1. Three illustrative type-1 sigmoidal membership functions generated with (1)–(3), with centers 40, 60, and 50, and widths 8, 8, and 4, respectively.

or "width," of the sigmoid (which can in general be different for the increasing and decreasing sigmoid). For symmetrical middle sigmoids, as an implementation nicety, an adjustment can be made such that only two parameters are required, for example

$$\mu(x) = \left(\frac{1}{1+e^{(c-5w-x)/w}}\right)\left(\frac{1}{1+e^{(x-c-5w)/w}}\right) \quad (3)$$

where c is the centre and w is the width of the sigmoid. Fig. 1 shows three illustrative type-1 sigmoidal functions, generated with (1)–(3), with centres 40, 60 and 50, and widths 8, 8, and 4, respectively.

C. *The Nonstationary Type-1 Fuzzy Expert System*

There are many ways imaginable in which variation in the input-output mapping of an FES can be introduced. For example, in systems that do not feature a complete set of rules, the selection of rules could be altered; rules could fire subject to a firing strength which could be randomly altered; various hedges to membership functions could be used; or the membership functions themselves could be altered arbitrarily. In this study, variation was introduced by means of minor alterations to the membership functions (only). Some explanation is required to clarify the meaning of "minor." Imagine a linguistic variable defined over a universe of discourse from 0 to 100 with two terms "low" and "high," where "low" is a Gaussian of width 10 centered at 25 and "high" is a second Gaussian centered at 75. If the center of term "low" is shifted right to 75 and the centre of term 'high' is shifted left to 25, the terms have simply swapped. This would be equivalent to swapping all occurrences of "low" and "high" in the rules of the rule base. So by "minor," we mean that although the deviations are big enough to affect the output of the FES and might have a noticeable visible effect on the membership functions if they were to be plotted, they would *not* be big enough to distort the membership functions beyond recognition.

Three mechanisms of minor variation in the membership functions were investigated

1) variation in the center points of the membership functions;
2) variation in the widths of the membership functions;
3) the addition of "white noise" to the membership functions.

The term "nonstationary" was introduced to describe such variable fuzzy systems. Preliminary investigations for determining variation in the results of fuzzy inferencing that can be obtained through variation of membership function parameters have been presented in [2]–[5].

In the first mechanism (variation in the centre points of the membership functions), new membership functions were obtained by replacing the center point (c) in (1)–(3) with $(c+\Delta\epsilon)$ where Δ is a percentage of the universe of discourse of the variable to which the membership function belongs and ϵ is a random number in $[-1,1]$. That is, left-edge centre-variation nonstationary fuzzy sets were generated by

$$\mu(x) = \frac{1}{1+e^{(x-(c+\Delta\epsilon))/w}} \quad (4)$$

right-edge center-variation nonstationary fuzzy sets were generated by

$$\mu(x) = \frac{1}{1+e^{((c+\Delta\epsilon)-x)/w}} \quad (5)$$

and middle center-variation nonstationary fuzzy sets were generated by

$$\mu(x) = \left(\frac{1}{1+e^{((c+\Delta\epsilon)-5w-x)/w}}\right) \times \left(\frac{1}{1+e^{(x-(c+\Delta\epsilon)-5w)/w}}\right) \quad (6)$$

For the second mechanism (variation in the widths of the membership functions), similar Equations were obtained by replacing w in (1)–(3) with $w+\Delta\epsilon$. Note that although parameter c and w appear more than once in (3), the new parameter (e.g., $w+\Delta\epsilon$) is generated only once before substitution.

In the third mechanism (the addition of white noise to the membership functions), $\Delta\epsilon$ is simply added to $\mu(x)$ after obtaining $\mu(x)$ using the (1)–(3), where ϵ is a random number in $[-1,1]$ for each x and Δ is a fixed percentage of the span of degree of membership (1 for normal membership functions). The value of μ is kept in the bounds $[0,1]$. Fig. 2 shows an example of a type-1 sigmoidal membership function with white noise added where $c=50, w=4$, and $\Delta = 2\%$ (0.02).

Within each mechanism, there are still a number of alternative ways in which variation could be introduced into the membership functions of an FES. For example, the membership functions may be generated once when the FES is instantiated (at the beginning of each decision making process); the membership functions might be regenerated each time the rule-base is invoked (at each new input datum); or each membership function might be re-generated each time it is accessed (at each rule). A single membership function might even appear more than once in a given rule (although this is not often the case in most applications). In this study, the membership functions were updated once each time each rule is evaluated. Given that, in our application, each membership function appears at most once in each rule, these variations were introduced whenever the membership functions are evaluated (at run time), i.e., each time $\mu(x)$ is enumerated for a given x. At each enumeration the result is a single real number in [0,1], as in any conventional type-1 FES. In operation therefore, the nonstationary FES functions as a type-1 FES and so, for example, no additional inferencing facilities are required. As membership functions are still type-1 fuzzy sets,

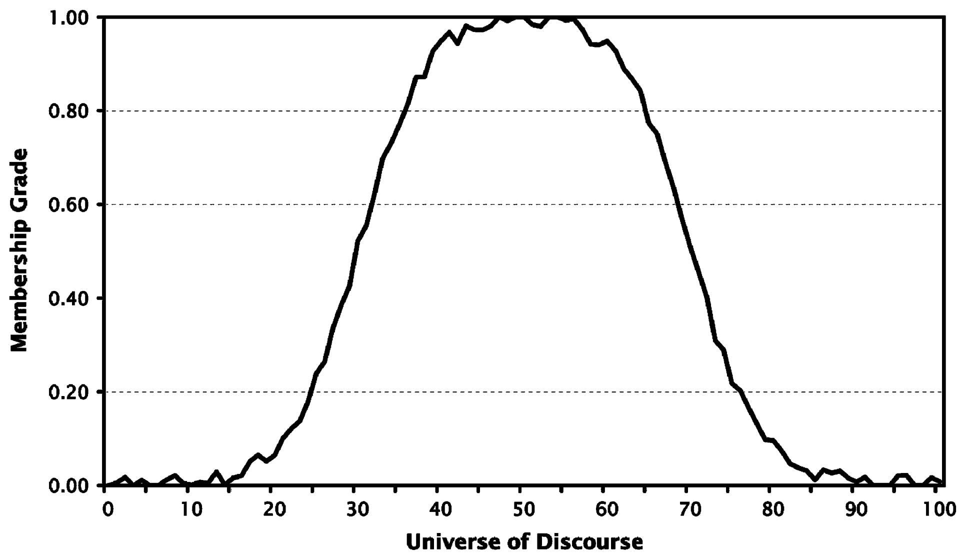


Fig. 2. Example type-1 sigmoidal membership function with white noise added where $c = 50, w30$, and $\Delta = 2\%$ (0.02).

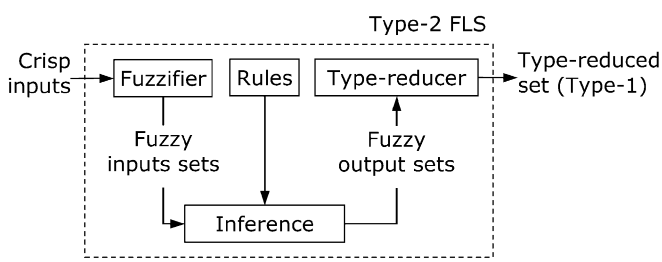


Fig. 3. Block diagram of an interval Type-2 fuzzy expert system (from Mendel).

the nonstationary FES performs as a type-1 FES, and so there are no changes in the other units that make up the FES (i.e., the rules, inferencing, or defuzzifier).

D. The Interval Type-2 Fuzzy Expert System

The purpose of the nonstationary FES described previously is to explore the distributions of decisions (outputs) produced by an FES incorporating variable membership functions when given a fixed set of inputs. Type-2 fuzzy sets are useful in circumstances where it is difficult to determine the exact membership function for a fuzzy set. Type-1 membership functions are precise in the sense that once they have been chosen all the uncertainty disappears. However, type-2 membership functions *are* fuzzy in the sense that there is uncertainty in the membership functions. A general type-2 FES is computationally too expensive—inferencing and output processing are computationally prohibitive [11]. One simplification is to use interval type-2 fuzzy sets in which the degree of membership is an interval over which the secondary membership degree is always 1.0. The mechanisms of the interval type-2 FES developed for this study are presented in Fig. 3. There are fast algorithms to compute the output of an interval type-2 FES [15]. Mendel [11] has established theoretical results to effectively determine the lower and upper bounds of the centroid of a type-2 set and has provided algorithms [15] for carrying out the necessary calculations. Hence, an interval type-2 FES was created that utilised interval type-2 membership functions with footprints of uncertainty that matched the boundaries of the random variations introduced into the nonstationary FES, using the same three mechanisms as outlined previously (that is, "center variation," "width variation," and "white noise" interval type-2 sets were used). This allowed the lower and upper *boundaries* of decision making to be determined.

Given that ϵ ranges over $[-1, 1]$ the upper and lower bounds of type-2 membership functions that include all possible nonstationary membership functions can be obtained by replacing the term $\Delta\epsilon$ with $\pm\Delta$ in (4)–(6). Hence, for example, the lower and upper bounds of centre variation type-2 sets are given by, for left-edge sets

$$\mu(x)_u = \max\left(\frac{1}{1+e^{(x-(c+\Delta))/w}}, \frac{1}{1+e^{(x-(c-\Delta))/w}}\right) \tag{7}$$

$$\mu(x)_l = \min\left(\frac{1}{1+e^{(x-(c+\Delta))/w}}, \frac{1}{1+e^{(x-(c-\Delta))/w}}\right) \tag{8}$$

for right-edge sets

$$\mu(x)_u = \max\left(\frac{1}{1+e^{((c+\Delta)-x)/w}}, \frac{1}{1+e^{((c-\Delta)-x)/w}}\right) \tag{9}$$

$$\mu(x)_l = \min\left(\frac{1}{1+e^{((c+\Delta)-x)/w}}, \frac{1}{1+e^{((c-\Delta)-x)/w}}\right) \tag{10}$$

and for middle-sets, as shown in the equation at the bottom of the next page, where $\mu(x)_u$ is the upper-bound, $\mu(x)_l$ is the lower-bound, and Δ is a percentage of the universe of discourse of the variable to which the membership function belongs. Similar expressions can be obtained for width variation and white noise.

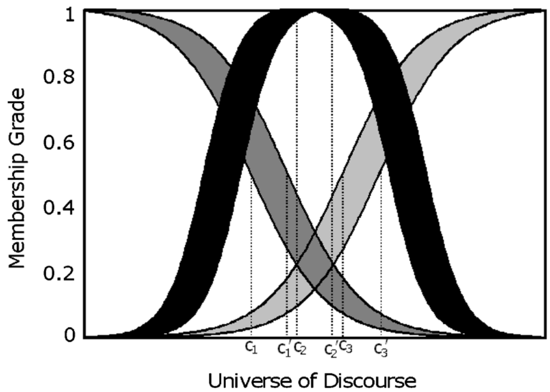


Fig. 4. Three illustrative interval type-2 sigmoidal membership functions obtained by deviating centres.

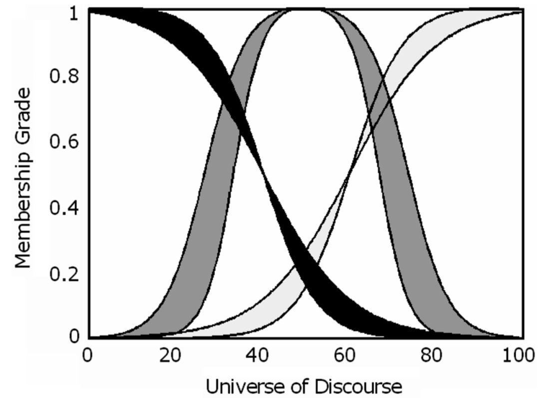


Fig. 5. Three illustrative interval type-2 sigmoidal membership functions obtained by deviating widths.

Fig. 4 shows three illustrative interval type-2 sigmoidal membership functions in which the universe of discourse ranges from 0 to 100, based on those shown in Fig. 1. The left-edge set has centre $c_1 = 45 \pm 5$ with a width of 8, the middle set has centre $c_2 = 55 \pm 5$ with a width of 4, and the right-edge set has centre $c_3 = 65 \pm 5$ with a width of 8, i.e., $\Delta = 5\%$ in each case. Fig. 5 shows three illustrative interval type-2 sigmoidal membership functions obtained by varying the width of the primary membership functions. The three centres are at 40, 50, and 60, the widths 8, 4, 8, respectively, and again $\Delta = 5\%$ in each case. A final set of interval type-2 membership functions were created by adding a uniform band around the primary membership functions, i.e., $\mu(x) = \mu(x) \pm \Delta$ where Δ is now a fixed percentage of the span of degree of membership. Fig. 6 shows three illustrative interval type-2 sigmoidal membership functions with centres at 40, 50, and 60, widths of 8, 4, 8, respectively, and all with $\Delta = 0.05$.

It is clear that the interval type-2 membership functions will contain all the possible the type-1 sets created by the equivalent nonstationary FES as *embedded* type-1 sets. However, *not* all embedded type-1 sets can be generated by the equivalent nonstationary FES.

III. A Case Study

A case study was chosen in which to investigate uncertain fuzzy reasoning. The domain of umbilical acid-base analysis was selected as it possessed three key properties

- a successful type-1 FES had previously been developed for the domain [16], [17];
- both interexpert variability (differences in opinion between different experts) and intraexpert variability (differences in opinion of one expert with themselves over time) were evident in the human expert decision making;
- experts were available to carry out repeated experiments in order to gather more data on inter- and intraexpert variability.

The following sections describe the medical background, the design and evaluation of the original type-1 FES, and the novel experiments carried out to investigate uncertain fuzzy reasoning.

A. Medical Context

The umbilical cord (navel-string) is the flexible cord-like structure connecting a fetus at the abdomen with the placenta. The umbilical cord contains one vein that transports nourishment to the fetus and two umbilical arteries which remove its waste products. In most developed countries, the umbilical cord is clamped at each end immediately after the birth of a baby and cut in order to allow the birth process to be more carefully managed. If the cord is double-clamped at each end, it may be cut between each pair of clamps and blood samples may be taken from the venous and arterial vessels in the isolated section. The acidity (pH), partial pressure of oxygen $(p\mathrm{O}_2)$ and partial pressure of carbon dioxide $(p\mathrm{CO}_2)$ of blood in each of the vessels may then be measured by a blood gas analysis machine. A parameter termed the base deficit of extracellular fluid $(\mathrm{BD}_{\mathrm{ecf}})$ can be derived from the pH and $p\mathrm{CO}_2$ parameters [18]. An interpretation of the state of health of the baby, termed umbilical acid-base (UAB) assessment, can be made based on

$$\mu(x)_u = \max\left(\frac{1}{(1+e^{(x-(c+\Delta)-5w)/w})(1+e^{((c+\Delta)-5w-x)/w})}, \frac{1}{(1+e^{(x-(c-\Delta)-5w)/w})(1+e^{((c-\Delta)-5w-x)/w})}\right) \tag{11}$$

$$\mu(x)_l = \min\left(\frac{1}{(1+e^{(x-(c+\Delta)-5w)/w})(1+e^{((c+\Delta)-5w-x)/w})}, \frac{1}{(1+e^{(x-(c-\Delta)-5w)/w})(1+e^{((c-\Delta)-5w-x)/w})}\right) \tag{12}$$

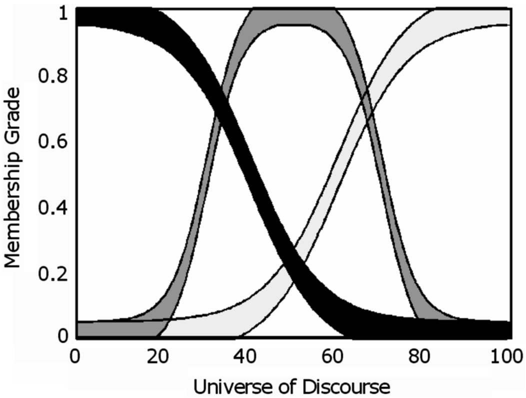


Fig. 6. Three illustrative interval type-2 sigmoidal membership functions obtained by adding white noise.

TABLE I
A SYNOPSIS OF THE RULE SET

	Input Variables				Output Variables		
Rule	pH_a	BD_a	pH_v	BD_v	*acidemia*	*component*	*duration*
1	low	high	low	high	very severe	metabolic	very chronic
2	low	high	low	mid	severe	metabolic	chronic
3	low	high	low	low	severe	metabolic	intermediate
4	low	high	mid	—	severe	metabolic	intermediate
5	low	high	normal	high	severe	metabolic	intermediate
6	low	high	normal	not high	severe	metabolic	acute
7	low	mid	low	not low	significant	mixed	chronic
8	low	mid	low	low	significant	mixed	intermediate
9	low	mid	mid	—	significant	mixed	intermediate
10	low	mid	normal	high	significant	mixed	acute
11	low	mid	normal	not high	significant	mixed	acute
12	low	low	low	not high	moderate	respiratory	chronic
13	low	low	mid	not high	moderate	respiratory	intermediate
14	low	low	normal	not high	moderate	respiratory	acute
15	mid	high	mid	not low	moderate	metabolic	chronic
16	mid	high	mid	low	moderate	metabolic	intermediate
17	mid	high	normal	—	moderate	metabolic	acute
18	mid	mid	mid	high	mild	mixed	chronic
19	mid	mid	mid	not high	mild	mixed	intermediate
20	mid	mid	normal	high	mild	mixed	intermediate
21	mid	mid	normal	not high	mild	mixed	acute
22	mid	low	mid	not high	mild	respiratory	intermediate
23	mid	low	normal	not high	mild	respiratory	—
24	normal	—	normal	—	normal	—	—
25	high	—	high	—	alkalotic	—	—

the pH and BD_{ecf} parameters from both arterial and venous blood.

UAB assessment of an infant immediately after delivery is an objective measure of labour and, as such, can be used to audit assessment of labour performance and can be used to exclude the diagnosis of birth asphyxia in approximately 80% of depressed newborns at term [19]. There are, however, a number of difficulties with the UAB assessment process which requires expertise in order to make a correct interpretation of the health of the baby. It is not possible to have an expert always available to make UAB assessments. To overcome these difficulties a type-1 FES was developed encapsulating the knowledge of leading obstetricians, neonatologists and physiologists gained over years of acid-base interpretation [1]. This FES combined knowledge of the errors likely to occur in acid-base measurement, physiological knowledge of plausible results and statistical knowledge of a large database of results.

The expert system was developed in three main incremental stages. Initially, a crisp expert system was developed incorporating conventional forward-chaining logic. Next, this crisp system was extended by deriving a preliminary type-1 FES for interpretation of error-free cases (only) in which the crisp rules were converted directly into a rule set whose terms were associated with type-1 fuzzy sets. This preliminary model was automatically tuned to match expert opinion using an algorithm based on simulated annealing [16]. Finally, the limitations of the preliminary fuzzy expert system were overcome through the creation of an integrated fuzzy expert system for the interpretation of all cases. Fresh knowledge elicitation resulted in a new rule set for identification of vessel origin and a revised rule set for the interpretation of results [20]. The performance of both aspects of this integrated system was validated in a further comparison with expert opinion [17].

B. *The Original Fuzzy Expert System*

The fuzzy system for umbilical acid-base assessment was based on four main input variables; arterial pH (pH_a), arterial base deficit (BD_a), venous pH pH_v, and venous base deficit (BD_v). Each of the four input parameters was assigned a linguistic variable: each of the pH variables $(pH_a$ and $pH_v)$ had four fuzzy terms corresponding to meanings of *low*, *mid*, *normal* and *high*; each of the BD_{ecf} variables $(BD_a$ and $BD_v)$ had three fuzzy terms corresponding to meanings of *low*, *mid* and *high*. Three output fuzzy variables were used, severity of acidemia (*acidemia*), duration of acidemia (*duration*), and component of the acidemia (*component*). The acidemia variable had six terms in its term-set: *severe*, *significant*, *moderate*, *mild*, *normal* and *alkalotic*; the duration variable had three terms: *chronic*, *intermediate* and *acute*; and the component variable had three terms in its term-set: *metabolic*, *mixed* and *respiratory*. Although it might seem somewhat odd that only four input variables were used to derive three output variables, it should be stressed that each of the three output variables correspond to actual linguistic terms used by the clinicians when assessing umbilical acid-base in clinical practice. All three output variables are complex nonlinear combinations of the four input variables.

The rules for the FES were obtained as a result of knowledge elicitation sessions with several leading clinicians skilled in umbilical cord blood acid-base analysis, and had been carefully refined to form a complete and consistent set of classifiers, in the sense that all possible values of the original input parameters were classified in accordance with the experts' opinion. A synopsis of the fuzzy rule set is shown in Table I. Note that the rules did not form a complete set in the sense that not every possible combination of input terms had an associated fuzzy rule. This was not necessary due to the inter-relationships between the various parameters. For example, rule 24 deals with all situations where the arterial and venous pH are *normal*—the values of the arterial and venous base deficit variable are not (clinically) relevant in this situation. The symbol "—" indicates any value of the parameter in the context of input variables, and no consequence result in the context of output variables, i.e., the variable is not utilised in the rule. The *very* hedge was taken as the square operator.

The Mamdani model of inference was used. The probabilistic family of operators was chosen, in which conjunction is defined as (ab) and negation as $1 - a$. No disjunction was used within the FES. Center-of-gravity (centroid) defuzzification was used

TABLE II
UNIVERSE OF DISCOURSE FOR FUZZY INPUT AND OUTPUT VARIABLES

Variable	Universe of Discourse
pH_a	6.60 ... 7.60
BD_a	0 ... 20
pH_v	6.60 ... 7.60
BD_v	0 ... 20
acidemia	0 ... 100
component	0 ... 100
duration	0 ... 100

TABLE III
CENTER (AND WIDTH) PARAMETERS USED TO GENERATE SIGMOID MEMBERSHIP FUNCTIONS FOR FUZZY INPUT AND OUTPUT VARIABLES

Inputs	Fuzzy Terms					
	low	*mid*	*normal*	*high*		
	left	middle	middle	right		
pH_a	7.05 (0.3)	7.10 (0.3)	7.30 (0.6)	7.45 (0.15)		
	low	*mid*	*high*			
	left	middle	right			
BD_a	10 (10)	11 (6)	12 (4)			
	low	*mid*	*normal*	*high*		
	left	middle	middle	right		
pH_v	7.10 (0.3)	7.15 (0.3)	7.35 (0.6)	7.50 (0.15)		
	low	*mid*	*high*			
	left	middle	right			
BD_v	8 (10)	9 (6)	10 (4)			
Outputs						
	severe	*significant*	*moderate*	*mild*	*normal*	*alkalotic*
	left	middle	middle	middle	right	right
acidemia	10 (10)	15 (20)	30 (20)	50 (40)	60 (40)	90 (10)
	metabolic	*mixed*	*respiratory*			
	left	middle	right			
component	25 (30)	50 (80)	75 (30)			
	chronic	*intermediate*	*acute*			
	left	middle	right			
duration	25 (30)	50 (80)	75 (30)			

to produce crisp values for each fuzzy output variable. All fuzzy sets were modelled with sigmoid membership functions as given by (1)–(3) although a scaling factor was introduced (purely as an implementation nicety)—thus left-edge sets were given by

$$\mu(x) = \frac{1}{1 + e^{5(x-c)/w}} \tag{13}$$

right-edge sets were were given by

$$\mu(x) = \frac{1}{1 + e^{5(c-x)/w}} \tag{14}$$

and middle sets were given by

$$\mu(x) = \frac{1}{(1 + e^{15(x-c-w/3)/w})(1 + e^{15(c-w/3-x)/w})}. \tag{15}$$

The universe of discourse of each variable used in the expert system is given in Table II, and the parameters used to generate the membership functions of each term in each linguistic variable are given in Table III. The parameters of the membership functions were determined by a tuning process [16].

The FES generated three fuzzy output variables, each of which were centroid defuzzified to produce three crisp outputs. In order to permit comparison with human expert opinion (there is no other absolute measure against which umbilical acid-base assessment can be compared) a single numeric "health index" was obtained, which represented how severe the overall results were for the infant. This health index was derived from the three outputs of the FES by combining their centroids into a single index by

$$\text{condition} = \text{acidemia} + \frac{\text{component}}{20} + \frac{\text{duration}}{10} \tag{16}$$

where the relative weightings of the three terms were determined within the tuning process [16]. Given that the three output variables were designed in such a way that low scores indicated a worsening condition for the infant, to the extreme *severe, metabolic, chronic acidemia*, this index can be thought of as indicating the *health* of the infant as represented by its acid-base balance at birth.

TABLE IV
EXAMPLE OF SPEARMAN RANK ORDER CORRELATION

Raw Data		Corresponding Ranked Data	
d_1	d_2	r_1	r_2
0.2	0.3	1	1.5
0.4	0.3	2	1.5
0.5	0.5	3	4
0.6	0.4	4	3
0.7	0.6	5	5
0.8	0.7	6	6
1.1	0.8	7	7.5
1.2	0.8	8	7.5
1.5	1.9	9	9
5.2	2.2	10	10
Standard Correlation 0.856		Spearman Correlation 0.982	

C. Spearman Rank Order Correlation

This single output ("health index") of the fuzzy system is arbitrarily scaled, making an absolute comparison between the fuzzy system output and expert opinion difficult. It is similarly not possible for human experts to attribute any meaningful absolute score to their assessment of the health of an infant on the basis of umbilical acid-base information alone. In such a situation, a statistic which ignores the absolute values, but takes into account the ordering of results can be used. Spearman rank order correlation can determine the degree of association between two sets of rank ordered data [21]. Spearman rank order correlation is simply the standard (Pearson) correlation applied to rank ordered data. The data are sorted in ascending order; the lowest datum is assigned a rank of 1, the next 2, and so on. In the case of ties (more than one data of the same numeric value), each tying datum is assigned the mean of their corresponding ranks. For example, if three data are tied in positions 4, 5 and 6, they are each assigned the mean rank (5). An example of Spearman rank order correlation is given in Table IV.

D. Evaluation of the Original Fuzzy Expert System

A target set of rankings was obtained by seeking the opinion of six human experts (clinicians). The clinicians were given the pH and $\mathrm{BD}_{\mathrm{ecf}}$ readings of 50 cases which were chosen from a database of more than 10000 cases by an independent expert both to cover a wide range of categories and to be "difficult to interpret." The clinicians were asked to rank these 50 results in order from "worst" to "best" in terms of likelihood of the infant having suffered intrapartum asphyxial damage. The interexpert

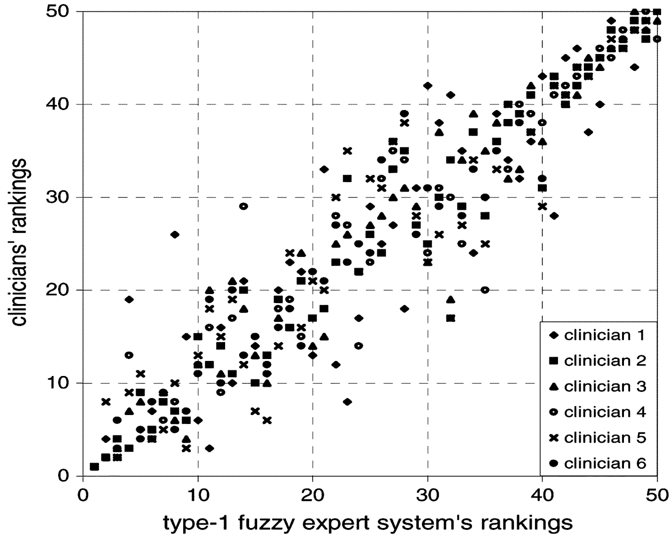


Fig. 7. Interexpert variation: each of the six expert's rankings plotted against the ranking given by the original type-1 FES.

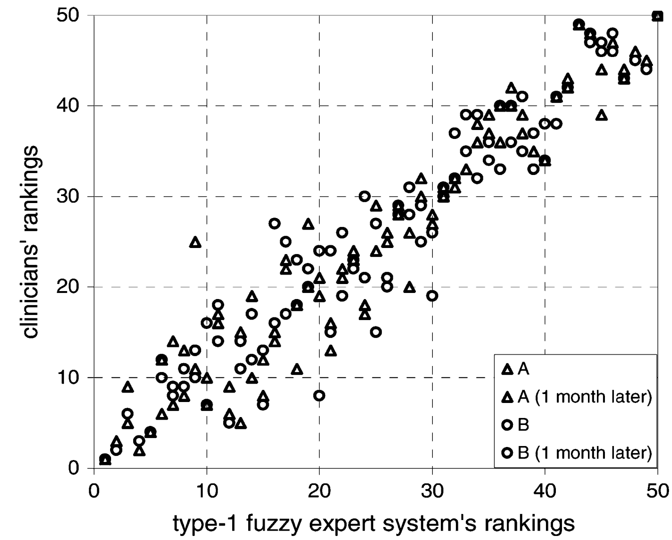


Fig. 8. Intraexpert variation: two expert's rankings plotted against the ranking given by the original type-1 FES.

agreement was then calculated from these rankings to ensure that there was an acceptable body of opinion. If the experts agree on the ranking of the cases, then comparison of the rankings of the FES and the human experts' rankings can be taken as a valid measure of the expertise captured by the FES. The Spearman rank order correlation coefficient of interexpert agreement was found to be 0.91.

Following the membership function tuning process, a Spearman rank order correlation coefficient of 0.93 was obtained between the FES and the clinicians. This was a very high level of agreement—indeed the system-clinician agreement of 0.93 was actually *higher* than the interexpert agreement of 0.91. This indicated that the FES had been tuned to represent a good "consensus view" of the clinicians. In order to visualize the agreement, a graph was plotted of each of the six expert's rankings plotted against the ranking given by the FES, as shown in Fig. 7. If perfect agreement had been found, a straight line through from (1,1) to (50,50) would have been obtained. However, as can be seen from Fig. 7, there is neither perfect agreement between the FES and any of the experts nor among the experts themselves (if two experts agreed with each other, then all their points would be superimposed). It can also be observed that there appears to be higher agreement at the extreme cases, whereas there is less agreement in the cases that fall in the middle of the range. For example, for the case that the type-1 FES ranked the worst (ranking 1), all the clinicians also ranked the worst and all the points are superimposed. On the other hand, for the case that the type-1 FES ranked 25th worst, the clinicians ranked 23rd, 24th, 26th, 27th, 29th and 32nd. That is, the distribution appears to present the characteristic of an elliptic envelope around the diagonal from (1,1) to (50,50).

E. Determination of Intraexpert Variability

It had become apparent from other studies (e.g., [22]) that individual human experts will show variation in their own decision making over time (intraexpert variation) as well the variation found between experts (interexpert variation). In order to assess the intraexpert variation in this domain a new study was carried out in which two of the original six experts were asked to reassess the same cases at least four weeks apart. The nature of the task was explained carefully to the experts and they were asked not to make any recording of the cases or to put any particular efforts into remembering them. The experts were then given the same 50 cases as before in randomised order at four week intervals and asked to rank them from "worst" to "best" once more (termed a "repeat" experiment). Fig. 8 shows the rankings given to each case by the experts in each repeat, plotted by the ordering of the original type-1 FES. For example, the case labeled the "worst" by the FES (rank 1) was also labeled "worst" by both experts at both repeats, so all four points are coincidental at location (1,1). For the case assigned rank 3 by the FES, expert A assigned ranks of 5 and 9, hence a triangle symbol appears at (3,5) and (3,9), while expert B assigned a rank of 6 in both repeats, hence the coincidental circle symbol at (3,6).

F. Experiments in Uncertain Fuzzy Reasoning

A number of experiments were carried out to investigate the distribution of outputs (decision making) obtained through the nonstationary FES and the boundaries of outputs obtained through the interval type-2 FES. Three sets of experiments were carried out to investigate centre variation, width variation and the addition of white noise. In these experiments, the Δ parameter was varied between 1% and 5%.

For the non-stationary FES, uniformly distributed random numbers in $[-1, 1]$ were used for ϵ in all cases. As an aside, it should be noted that it is trivial to use nonuniformly distributed random numbers rather than uniformly distributed: the equivalent type-2 systems would then have to be noninterval. In each case the non-stationary FES was run 100 times, each time on the same 50 cases as given to the clinicians. The outputs of the non-stationary FES were recorded in each trial and used to rank order the cases from best to worst. Histograms could then be plotted of the rankings obtained from the nonstationary FES plotted against the original rankings obtained from the type-1 FES.

For the interval type-2 FES, the footprint of uncertainty of the interval type-2 membership functions was determined from (7)–(12). The interval type-2 FES was then run and Mendel's method used to obtain the lower and upper bound of the output centroids. These were combined using (16) to obtain the centroid for the health index. A type-2 FES must reduce to a type-1 FES when the uncertainty around the membership functions disappears. This was verified by running the interval type-2 FES with 0% deviation in the membership function parameters and checking that the rankings obtained were the same as those obtained from the original type-1 FES. A method was then required to obtain the lower and upper bounds of rankings which may be obtained from a set of 50 health index intervals. This is a far from straight-forward task in itself. Suppose, for example, three cases are considered, labelled A, B, and C. For these three cases, suppose the following intervals are obtained, [3,6] for case A, [4,7] for case B and [5,8] for case C. To obtain the full set of possible rank orders, we need to consider the rankings obtained using the lower and upper bounds of each case. For three intervals there as thus 2^3 possible rank orders. If we consider the lower bound of A (3), it can be seen that, regardless of B and C, A will be ranked lowest (rank 1). If we consider the upper bound of A (6) together with the lower bounds of B and C (4 and 5, respectively) it can be seen that A will be ranked highest (rank 3). Thus case A has rank bounds of $[1, 3]$. It can be verified by inspection that, in this specific example, this is also true of both B and C. Hence, all three cases have rank bounds of [1,3]. In general, therefore, this process would need to be repeated for n cases, and so 2^n rankings need to be considered—for 50 cases this would have been computationally prohibitive and a more efficient ranking method was necessary.

G. An Efficient Ranking Method for Intervals

The following two observations were used to simplify the ranking process.

1) Observation 3.1: Let $[l_i,\ u_i]$, for $i = 1, \ldots, n$, be the n health index intervals sorted in ascending order according to l_i, where l_i is the lower bound and u_i is the upper bound. Now, consider a case $j, j \in [1, 50]$. If, $\forall k$ in $[1, j-1]$, $\max_k(u_k) < l_j$, then the rank of the cases $1 \ldots j-1$ will never be higher than the ranks of the cases $j \ldots n$.

So the n cases can be divided into two separate groups for ranking; the same process can be repeated for the cases $j \ldots n$ again which may result in yet smaller groups to rank.

2) Observation 3.2: Let $[l_i,\ u_i]$, for $i = 1, \ldots, n$ be the n health index intervals, where l_i is the lower bound and u_i is the upper bound. The minimum rank of case k is the rank of l_k among all u_i except u_k. The maximum rank of case k is the rank of u_k among all l_i except l_k.

Observation 3.1 can be used to find all the possible ranks of a set of intervals, whereas observation 3.2 can only be used to find the minimum and maximum rank for each interval. Using observation 3.1 the computation time required for ranking 20 cases was decreased to a few minutes and for 50 cases to less than 30 mins. Using observation 3.2, minimum and maximum rankings of 50 cases were determined in a few minutes.

TABLE V
AN EXAMPLE OF SPLITTING THE PROBLEM OF RANKING A GROUP OF INTERVALS INTO SUB-GROUPS USING OBSERVATION 3.1

		Health Index		
	Case ID	Lower Bound	Upper Bound	Rank Range
1	P00865	11.08	11.08	1
2	P00862	16.08	25.73	
3	P09216	17.45	18.80	2-3
4	P03977	26.23	39.50	
5	P07320	28.68	47.93	
6	P09905	29.98	44.53	
7	P02370	31.55	32.43	
8	P03444	36.33	47.25	
9	P10297	39.23	51.53	
10	P00486	39.93	53.10	
11	P00136	47.85	57.80	
12	P06463	48.53	58.23	
13	P09308	49.45	59.85	
14	P10828	52.38	60.83	
15	P08722	55.05	63.00	
16	P06617	59.30	68.83	
17	P02844	63.83	70.45	4-17
18	P08070	74.50	83.63	
19	P06086	77.68	87.85	
20	P09151	81.20	88.08	18-20

Example: In Table V, 20 cases are presented with the lower and upper bounds of their health indices. Using observation 1 the ranking problem can be divided into three ranking problems, as follows. It can be seen that the lower bound of case 2 (16.08) is higher than the maximum upper bound of case 1 (11.08); thus case 1 forms the first sub-group with 1 member. Then, the lower bound of case 4 (26.63) is higher than the maximum upper bound of cases 2-3 (25.73); thus cases 2 and 3 form the second sub-group with 2 members. Next, the lower bound of case 18 (74.5) is higher than the maximum upper bound of cases 4-17 (70.45); thus cases 4-17 form the third sub-group with 14 members. Cases 18–20 form the fourth sub-group with 3 members. Thus a total of $2^1 + 2^2 + 2^{14} + 2^3 = 16396$ combinations need to be considered, rather than the $2^{20} = 1048576$ combinations to be considered in the original ranking problem. Note that, any degenerate intervals such as case 1 in Table V of course reduce the total number of combinations to be considered yet further.

H. Results of Uncertain Fuzzy Reasoning

1) Variation in Membership Function Centers: Figs. 9(a), (c), (e), show the distributions in rankings obtained as the variation in the center of the membership functions is increased from 1%, to 3%, and to 5%. The x-axis corresponds to the original type-1 FES rankings, the y-axis corresponds to the rankings of the nonstationary FES, and the z-axis corresponds to the number of times the appropriate ranking was obtained. Figs. 9(b), (d) and (f) show the lower and upper bounds of the rankings obtained by the interval type-2 FES with centre variation plotted against the rankings obtained by the original type-1 FES. It can be seen that the lower and upper bounds of rankings expand very rapidly, so that at only 5% deviation in the centre, almost all cases have ranking bounds close to [1,50], i.e., the "correct" rank ordering of the cases is almost completely lost.

2) Variation in Membership Function Widths: Figs. 10(a), (c), and (e) show the distributions in rankings obtained as the variation in the width of the membership functions is increased

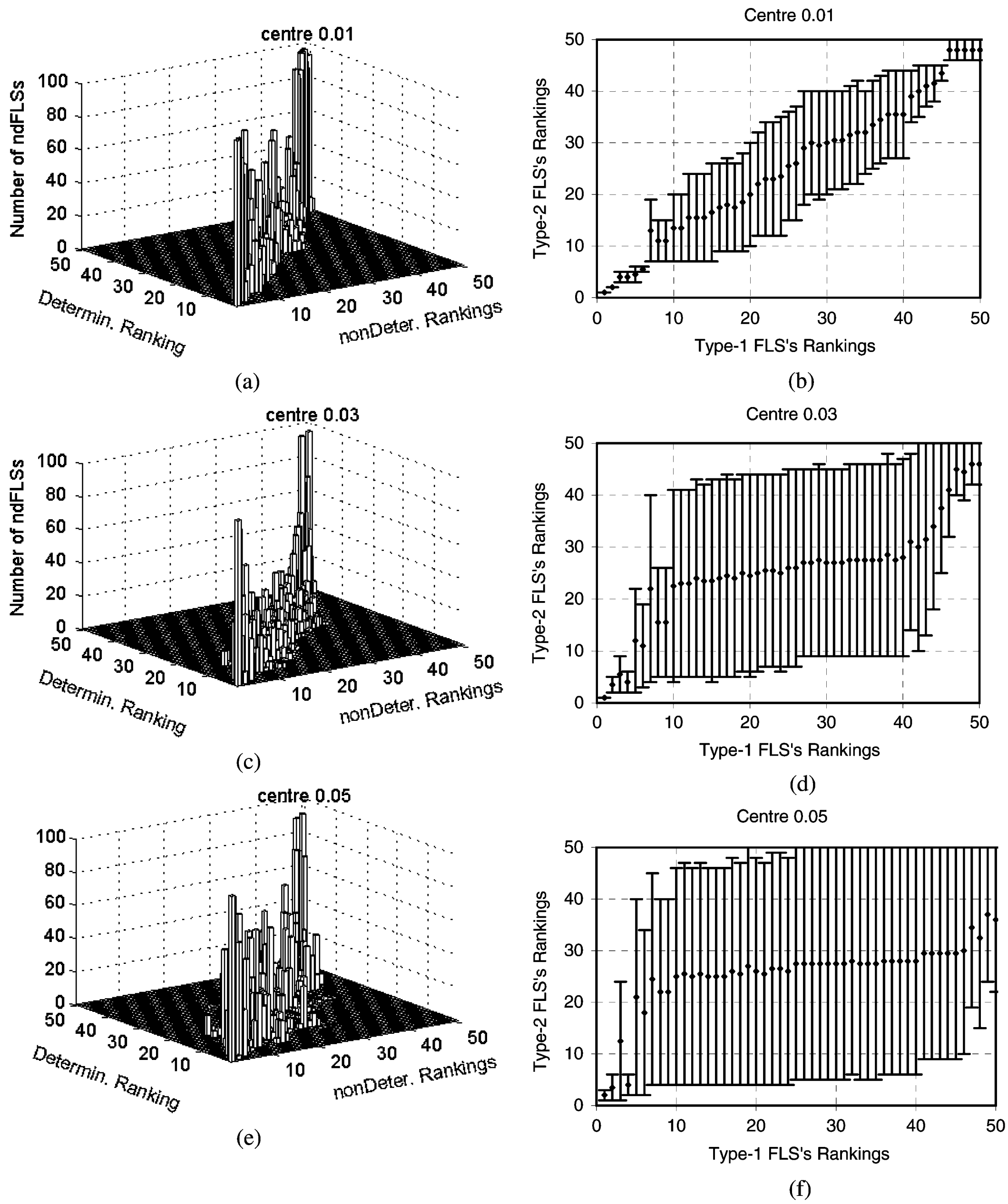


Fig. 9. Histograms of the distributions and boundaries of decision making obtained when the centre points of membership functions are varied. (a), (c), and (e) Distributions obtained from the non-stationary type-1 FES. (b), (d), and (f) Theoretical boundaries derived from the type-2 FES. In each case, the original fuzzy expert system's rankings are used as the reference and are plotted on the x-axis. (a) Distribution of non-stationary FES rankings obtained with 1% centre variation. (b) Lower and upper bounds of interval type-2 FES rankings obtained with 1% centre variation. (c) Distribution of nonstationary FES rankings obtained with 3% centre variation. (d) Lower and upper bounds of interval type-2 FES rankings obtained with 3% centre variation. (e) Distribution of nonstationary FES rankings obtained with 5% center variation. (f) Lower and upper bounds of interval type-2 FES rankings obtained with 5% center variation.

from 1%, to 3%, and to 5%. Figs. 10(b), (d), and (f) show the lower and upper bounds of the rankings obtained by the interval type-2 FES with width variation plotted against the rankings obtained by the original type-1 FES. It can be seen that, in contrast to those obtained for centre variation, the lower and upper bounds of rankings increase less rapidly with increasing variation in widths of the membership functions. It is also apparent that the shape of the lower and upper boundaries of rankings, particularly in Figs. 10(d) and (f), is highly reminiscent of the elliptic envelope seen in Fig. 7.

3) Addition of White Noise to Membership Functions: Figs. 11(a), (c), and (e) show the distributions in rankings obtained as the variation in the white noise added to the membership functions is increased from 1%, to 3%, and to 5%. Figs. 11(b), (d), and (f) show the lower and upper bounds of the rankings obtained by the interval type-2 FES with width variation plotted against the rankings obtained by the original type-1 FES. It can be seen that the behaviour is very similar to that obtained for width variation. (a) Distribution of nonstationary FES rankings obtained with the addition of 1% white noise. (b) Lower and upper bounds of interval type-2 FES rankings obtained with the addition of 1% white noise (c) Distribution of nonstationary FES rankings obtained with the addition of 3% white noise. (d) Lower and upper bounds of interval type-2 FES rankings ob-

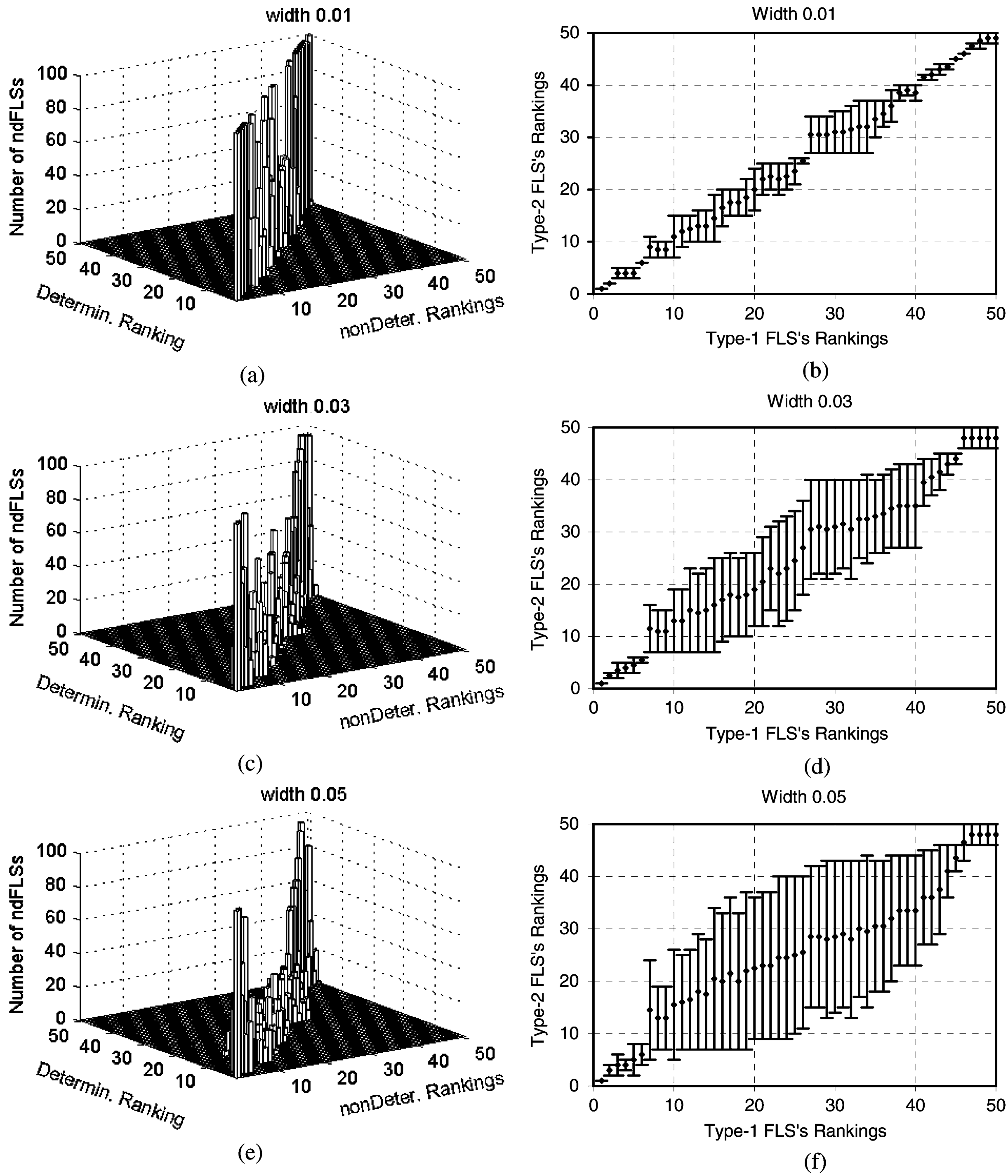


Fig. 10. As Fig. 9, but showing the boundaries of decision making obtained when the widths of membership functions are varied. (a) Distribution of nonstationary FES rankings obtained with 1% width variation. (b) Lower and upper bounds of interval type-2 FES rankings obtained with 1% width variation. (c) Distribution of nonstationary FES rankings obtained with 3% width variation. (d) Lower and upper bounds of interval type-2 FES rankings obtained with 3% width variation. (e) Distribution of nonstationary FES rankings obtained with 5% width variation. (f) Lower and upper bounds of interval type-2 FES rankings obtained with 5% width variation.

tained with the addition of 3% white noise. (e) Distribution of nonstationary FES rankings obtained with the addition of 5% white noise. (f) Lower and upper bounds of interval type-2 FES rankings obtained with the addition of 5% white noise.

I. Best Models of Inter- and Intraexpert Variation

Figs. 12 (h) shows the graph of inter-expert variation (Fig. 7) overlaid with the boundaries of rankings obtained for the interval type-2 FES for a selection of different variation mechanisms and values of Δ. In the ideal situation the lower and upper boundaries of interval type-2 FES rankings would exactly overlay the interexpert variation observed. However, it is highly improbable that such a perfect match would ever be obtained in practice. If the interval type-2 FES boundaries collapsed to zero width, then all expert variations would lie outside the boundaries; if, on the other hand, the boundaries were expanded to the

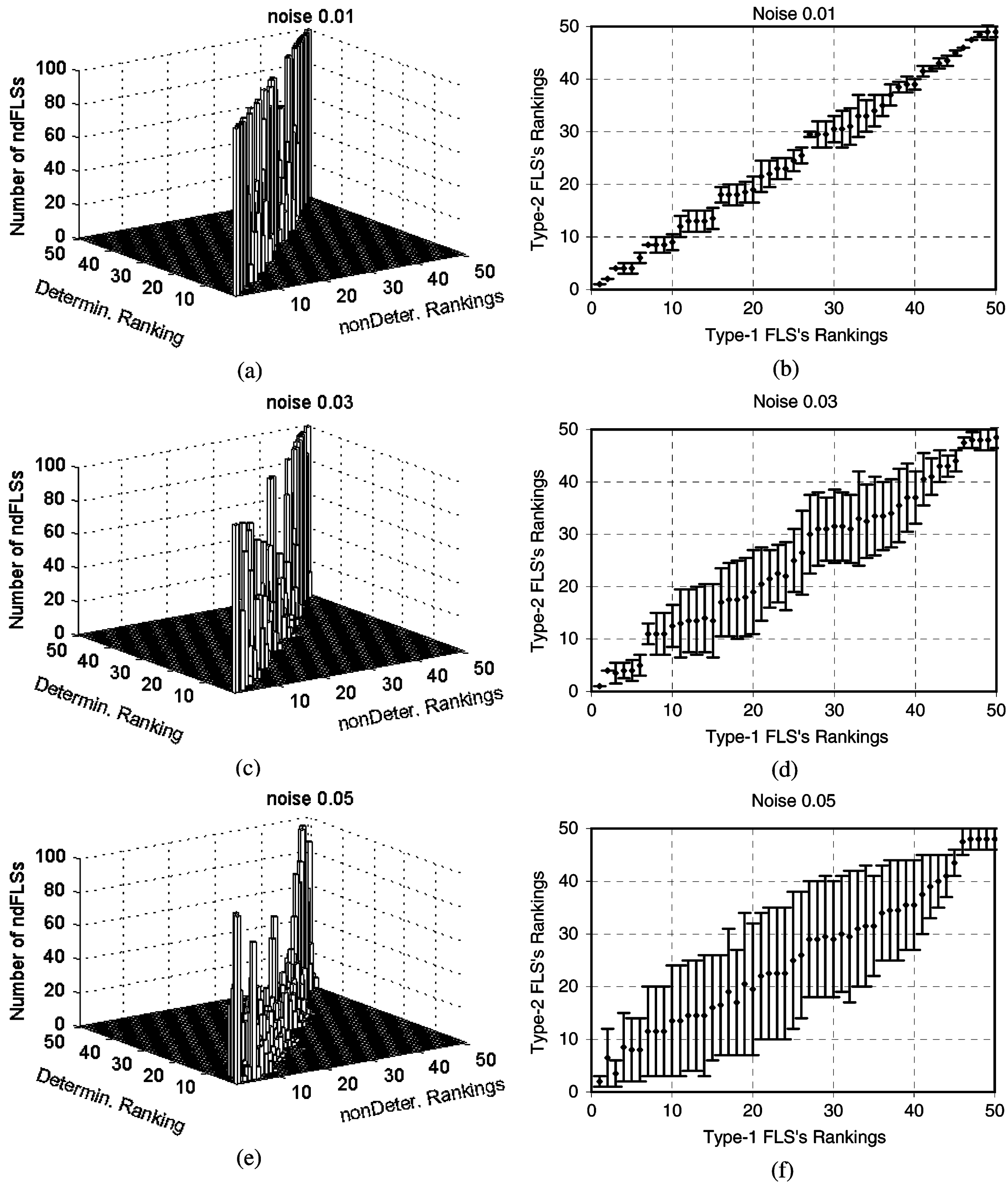


Fig. 11. As Fig. 10, but showing the boundaries of decision making obtained when white noise is added to the membership functions. (a) Distribution of nonstationary FES ranking obtained with the addition of 1% white noise. (b) Lower and upper bounds of interval type-2 FES rankings obtained with the addition of 1% white noise. (c) Distribution of nonstationary FES rankings obtained with the addtion of 3% white noise. (d) Lower and upper bounds of interval type-2 FES rankings obtained with the addtion of 3% white noise. (e) Distribution of nonstationary FES rankings obtained with the addition of 5% white noise. (f) Lower and upper bounds of interval type-2 FES rankings obtained with the addition of 5% white.

maximum, then all expert variations would lie inside. Neither of these two extremes are appropriate for capturing a notion of the "best" match of interexpert variation. We therefore propose a measure of root-mean-square error (RMSE) between the boundary limits and all the observed expert points. Table VI shows the RMSE obtained for Fig. 12(a)–(h) together with the number of experts' rankings that lay outside the span of interval type-2 FES's rankings. The two lowest measures of RMSE have been highlighted in bold. These are the boundaries for 3% noise addition (the lowest RMSE with 49 outliers) and 3% width variation (the second lowest RMSE with only 41 outliers).

Finally, Figs. 13(a) and (b) show the scatter of intraexpert variation overlaid with boundaries in variability of the "best" models of interexpert variability (3% width variation and 3% noise addition).

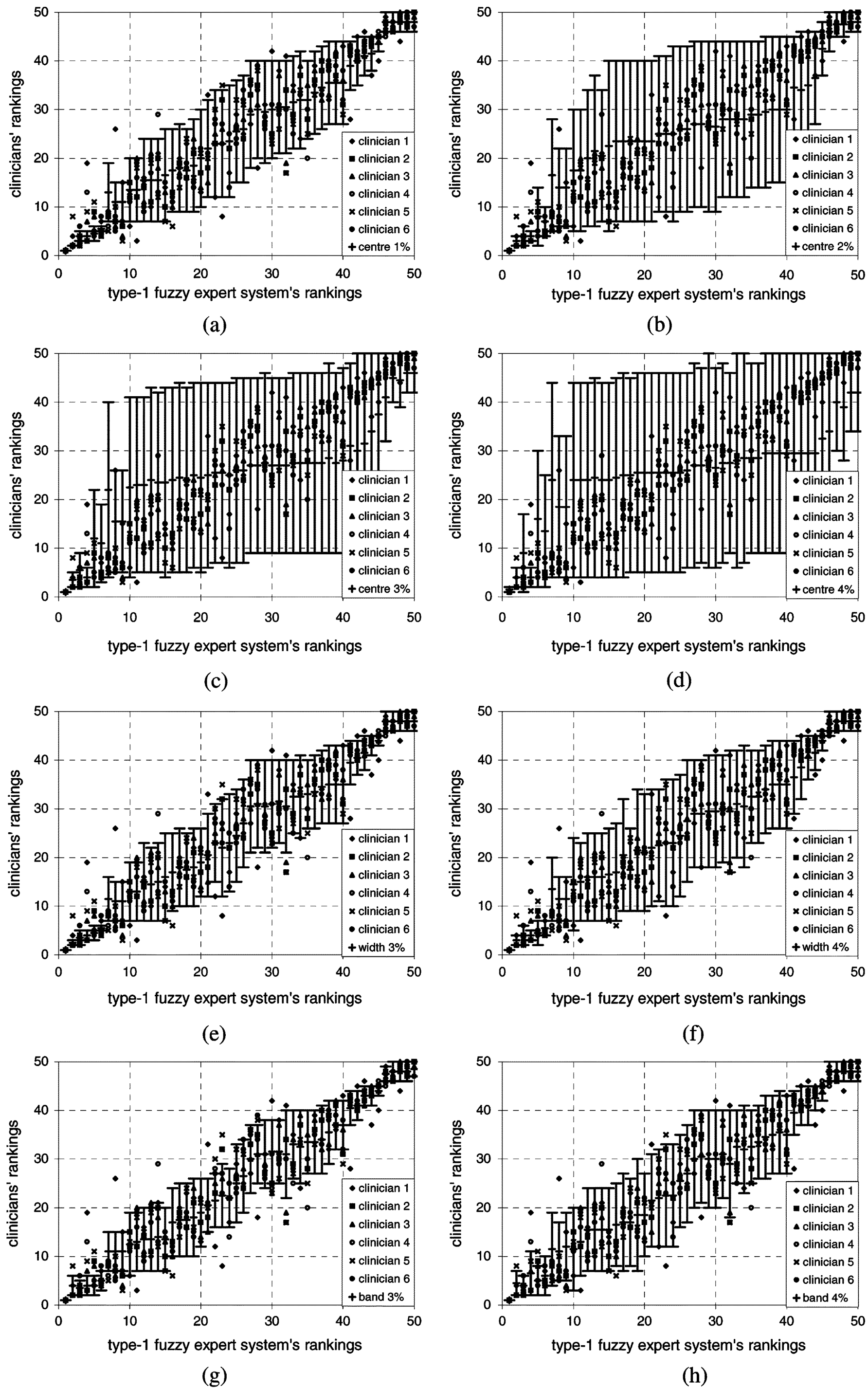
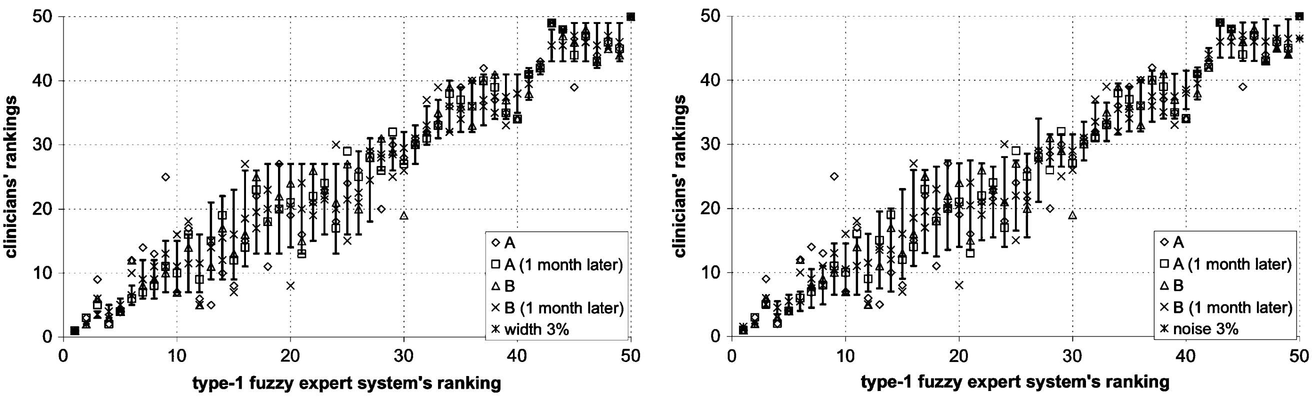


Fig. 12. Selected plots showing the original scatter of clinicians' rankings against the original fuzzy expert system, overlaid with boundaries in variability of different types and sizes derived from the type-2 FES. These charts provide an informal "visual" indication of which variation best matches interexpert variability. (a) Plot of interexpert variation overlaid with boundaries from 1% center variation. (b) Plot of interexpert variation overlaid with boundaries from 2% center variation. (c) Plot of interexpert variation overlaid with boundaries from 3% center variation. (d) Plot of interexpert variation overlaid with boundaries from 4% center variation. (e) Plot of interexpert variation overlaid with boundaries from 3% width variation. (f) Plot of interexpert variation overlaid with boundaries from 4% width variation. (g) Plot of interexpert variation overlaid with boundaries from 3% noise addition. (h) Plot of interexpert variation overlaid with boundaries from 4% noise addition.

IV. Discussion

All human beings, including "experts," exhibit variation in decision making. Up to now it has been an implicit assumption that expert systems, including fuzzy expert systems, should not exhibit such variation. Indeed, the property of being 100% consistent—always giving the same answer (outputs) when presented with the same data (inputs)—has been advocated as a key advantage of computerised expert systems. In this paper, a new type of fuzzy expert system, termed a nonstationary FES, has



(a) (b)

Fig. 13. Plots showing the scatter of intraexpert variation, overlaid with boundaries in variability of the "best" models of interexpert variability. (a) Plot of intraexpert variation overlaid with boundaries from 3% width variation. (b) Plot of intraexpert variation overlaid with boundaries from 3% noise addition.

TABLE VI
MEASURES OF THE MATCH BETWEEN DIFFERENT MECHANISMS OF FES VARIATION AND INTER-EXPERT VARIABILITY

Varied Parameter	RMSE	Number of Outliers
Centre 1%	4.42	6
Centre 2%	7.99	6
Centre 3%	9.76	11
Centre 4%	11.09	40
Width 3%	**3.98**	**41**
Width 4%	5.00	24
Noise 3%	**3.77**	**49**
Noise 4%	4.00	26

been presented in a case study in order to examine whether the inter- and intraexpert variability found in a particular decision making domain can be successfully modelled using an FES in which small variations in the membership functions are found.

Three different mechanisms for introducing variation into an FES have been proposed, and the effect of these mechanisms have been investigated by creating nonstationary FESs to allow the distributions of outputs to be studied and by creating interval type-2 FESs to allow the lower and upper boundaries of outputs to be studied. In the case study presented here, it has been found that 3% variation in the width of membership functions and the addition of 3% "white noise" are both good candidates for the best match of inter- and intraexpert variability. It can be observed from these trials that there is a direct relationship between the uncertainty in the membership functions used in the nonstationary FES and interval type-2 FES. Not surprisingly, as the variation in the membership functions is increased, the variation in decision making is observed to increase. However, it is interesting to note that the nature of the variation in the output can match the nature of the interexpert and intraexpert variability quite closely. The elliptic envelope observed in the interexpert variability (there is apparently more variability in the middle range of cases than at the extremes) can be mimicked quite closely by the nonstationary FES.

It is not clear at the present time why variation in the centre of membership functions apparently causes a much higher degree of variability than similar levels of variation in the width of membership functions or the addition of white noise to the membership functions. It may be tentatively suggested that the centre point of the membership functions in an FES has more effect on the system than the widths of the membership functions. However, it may be that this observed effect is purely an artefact of the rules used in this particular system and is *not* a general finding, but this would need to be explored further before general conclusions may be reached.

The primary membership functions used in this system are sigmoids or products of sigmoids, rather than the more common Gaussian membership functions. There is no specific reason for this, other than they had been found in empirical studies to better match the experts' opinions of what the membership functions "should" look like (when the original type-1 FES had been designed). Of course, whether this is in any way relevant is a matter, among other things, as to whether the experts had (or have) a real understanding of what fuzzy membership functions are. Theoretical results for inferencing with type-2 interval sets based on sigmoidal primary membership functions have not been established here. It is possible that, in general, the type-2 inferencing used may not hold in all cases. However, the fact that the type-2 system produced the same result as the type-1 FES when the intervals were reduced to zero is a hopeful sign that the inference may be valid in general.

The question of whether such nonstationary FESs may be helpful in real decision making remains unanswered at present. We make two suggestions of scenarios in which we believe they *may* be useful.

1) In expert system validation. Expert systems might be validated using a form of "Turing Test," in which a panel of experts must differentiate between the expert system under validation and a group of their peers. If an intelligent system cannot be differentiated from its human counterparts, this may provide a convincing argument for its usage. However, a conventional expert system would immediately be identified due to its lack of variability. A nonstationary FES, which exhibits the same variability as the human participants, *might* be better placed to pass such a test.
2) In situations where a range of opinions is desirable. In some situations a variety of alternative decisions might be useful whereas an "average" decision is less useful or even undesirable. Consider, for example, a situation where a driver

of a vehicle may be about to collide with an obstacle immediately ahead. Two decisions might be recommended to avoid the imminent collision: "turn left" or "turn right": either is acceptable, but the average decision is not. Nonstationary systems may be able to produce a range of alternative decisions from which the human decision-maker may choose the best.

Pardoxically, almost, many practical successes of fuzzy methods have been in fuzzy control, but this success has not yet been matched in modelling human reasoning—Zadeh's "Computing with Words" paradigm [23]. We believe that nonstationary FESs may be a useful step in increasing the uptake of fuzzy methods in modelling human reasoning.

V. Future Work

The research on understanding and modelling the dynamics of variation in human decision making is ongoing. There are many avenues of investigation that may be explored. The relationship between nonstationary FESs and interval type-2 FESs needs to be further explored and clarified, in order to be able to answer questions such as whether the boundaries of interval type-2 FES decisions are theoretically identical to the limits of the nonstationary decisions. It may be that there is a role for non-stationary FESs in being approximators to general type-2 inferencing. An non-stationary FES is implemented as a type-1 FES; as mentioned in Section III.F, it is trivial to use nonuniformly distributed random numbers in the variation of the membership functions in an non-stationary FES. If this were to be done and the nonstationary FES were to be run (say) 30 times, then statistics on the mean and standard deviation of, for example, the output centroid could easily be obtained. This could be used, perhaps, to approximate the inferencing of a full type-2 FES with non-interval secondary membership functions. Further studies on inter- and intra-expert variability need to be carried out, particularly with different fuzzy expert systems in different decision-making domains so that the results presented in this paper may begin to be generalised.

Acknowledgment

The authors would like to thank the referees for their helpful comments which helped to improve the paper.

References

[1] J. M. Garibaldi, "Intelligent techniques for handling uncertainty in the assessment of neonatal outcome," Ph.D. dissertation, Univ. Plymouth, Plymouth, U.K., 1997.

[2] T. Ozen and J. M. Garibaldi, "Investigating adaptation in type-2 fuzzy logic systems applied to umbilical acid-base assessment," in *Proc. Eur. Symp. Intelligent Technologies, Hybrid Systems and Their Implementation on Smart Adaptive Systems*, Oulu, Finland, Jun. 2003, pp. 289–294.

[3] T. Ozen, J. M. Garibaldi, and S. Musikasuwan, "Preliminary investigations into modelling the variation in human decision making," in *Proc. Information Processing and Management of Uncertainty in Knowledge Based Systems*, Perugia, Italy, Jul. 2004, pp. 641–648.

[4] T. Ozen and J. M. Garibaldi, "Effect of type-2 fuzzy membership function shape on modelling variation in human decision making," in *Proc. IEEE Int. Conf. Fuzzy Systems*, Budapest, Hungary, Jul. 2004, pp. 971–976.

[5] ——, "Modelling the variation in human decision making," in *Proc. Fuzzy Sets in the Heart of Canadian Rockies, NAFIPS-04*, Banff, Canada, Jun. 2004, pp. 617–622.

[6] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Inform. Sci.*, vol. 8, pp. 199–249, 1975.

[7] M. Mizumoto and K. Tanaka, "Some properties of fuzzy sets of type-2," *Inform. Control*, vol. 31, pp. 312–340, 1976.

[8] ——, "Fuzzy sets of type-2 under algebraic product and algebraic sum," *Fuzzy Sets Syst.*, vol. 5, pp. 277–290, 1981.

[9] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New York: Academic, 1980.

[10] E. Hisdal, "The IF THEN ELSE statement and interval-values fuzzy sets of higher type," *Int. J. Man-Mach. Stud.*, vol. 15, pp. 385–455, 1981.

[11] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[12] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 6, pp. 643–658, Dec. 1999.

[13] Q. Liang and J. M. Mendel, "Interval type-2 fuzzy logic systems: Theory and design," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 535–550, Oct. 2000.

[14] J. M. Mendel and R. I. John, "Type-2 fuzzy sets made simple," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 117–127, Apr. 2002.

[15] N. N. Karnik, Q. Liang, and J. M. Mendel, Type-2 fuzzy logic software 2001 [Online]. Available: http://sipi.usc.edu/mendel/software

[16] J. M. Garibaldi and E. C. Ifeachor, "Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 1, pp. 72–84, Feb. 1999.

[17] J. M. Garibaldi, J. A. Westgate, and E. C. Ifeachor, "The evaluation of an expert system for the analysis of umbilical cord blood," *Artif. Intell. Med.*, vol. 17, pp. 109–130, 1999.

[18] O. Siggaard-Andersen, "An acid-base chart for arterial blood with normal and pathophysiological reference areas," *Scand. J. Clinical Labo. Invest.*, vol. 27, pp. 239–245, 1971.

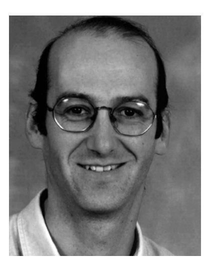
[19] J. A. Thorp and R. S. Rushing, "Umbilical cord blood gas analysis," *Obstet. Gynecol. Clin. North Amer.*, vol. 26, no. 4, pp. 695–709, 1999.

[20] J. M. Garibaldi and E. C. Ifeachor, , P. Szczepaniak, P. J. G. Lisboa, and J. Kacprzyk, Eds., "The development of a fuzzy expert system for the analysis of umbilical cord blood," in *Fuzzy Systems in Medicine*. New York: Springer-Verlag, 2000, pp. 652–668.

[21] E. L. Lehmann and H. J. M. D'Abrera, *Nonparametrics: Statistical Methods Based on Ranks*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[22] R. D. F. Keith, S. L. Beckley, J. M. Garibaldi, J. A. Westgate, E. C. Ifeachor, and K. R. Greene, "Multicentre validation of an intelligent system for managing labour," *Expert Syst. Appl.*, vol. 11, no. 4, pp. 537–541, 1996.

[23] L. A. Zadeh, "Fuzzy logic = computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 1, pp. 103–111, Feb. 1996.



J. M. Garibaldi received the B.Sc. (hons.) degree in physics from Bristol University, U.K., and the M.Sc. and Ph.D. degrees from the University of Plymouth, U.K., in 1984, 1991, and 1997, respectively.

He is an Associate Professor within the Automated Sheduling, Optimization, and Planning (ASAP) Research Group in the School of Computer Science and Information Technology at the University of Nottingham, U.K.. The ASAP Group conducts research into models, heuristics, and algorithms for automatically producing high-quality solutions to a variety of real world decision support and optimization problems. In particular, the ASAP Group is undertaking research into a broad range of optimization approaches that can operate at a higher level of generality than current search technology. His main research interest is in modelling uncertainty in human reasoning, with particular emphasis on medical decision support. He has published over fifty papers on the subject.

T. Ozen received the B.Sc. degree from the Middle East Technical University, Ankara, Turkey, the M.Sc. degree from Imperial College, London, U.K., and the Ph.D. degree from Leicester University, Leicester, U.K., in 1997, 1998, and 2002, respectively.

He was a Research Associate within the Automated Sheduling, Optimization, and Planning (ASAP) Research Group in the School of Computer Science and Information Technology, the University of Nottingham, Nottingham, U.K., at the time this work was carried out. He now works in the software industry.