Performance analysis of a fuzzy decision-support system for culling of dairy cows

R. LACROIX[1], M. STRASSER[1,2], R. KOK[2] and K.M. WADE[1]

[1]*Department of Animal Science; and* [2]*Department of Agricultural and Biosystems Engineering, McGill University, 21111 Lakeshore Road, Ste-Anne de Bellevue, QC, Canada H9X 3V9.* *Received 5 October 1997; accepted 7 April 1998.*

Lacroix, R., Strasser, M., Kok, R. and Wade, K.M. 1998. **Performance analysis of a fuzzy decision-support system for culling of dairy cows.** Can. Agric. Eng. **40**: 139-152. To investigate the use of fuzzy logic in decision-support systems for dairy cattle breeding, a prototype software system was developed. The objectives were to determine advantages and limitations of fuzzy logic for this type of application and to establish a methodological basis for the development of more complete decision-support systems in the future. The goal of the prototype decision-support system was to make culling decisions on the basis of monthly production data. During the development phase, three experts in the area of animal breeding were interviewed. The final version comprised three rule sets which considered a total of five input variables. The membership functions for most of the input variables were made herd-specific. Results showed that the use of fuzzy sets could increase the flexibility and adaptivity of rule-based expert systems. The same rule sets were appropriate under various scenarios (e.g., herds, regions, and breeds) with inferences being made specific to each context by adjusting the membership functions associated with the fuzzy sets. Results also showed that the inferences from fuzzy sets could be used as an alternative to methods currently used for within-herd cow rankings. The development of expert systems based on fuzzy logic seems relatively easy and such expert systems may require a smaller number of rules than traditional approaches to achieve similar output variations. **Keywords:** dairy cattle, culling, decision-support systems, fuzzy logic.

Un logiciel prototype a été développé afin d'explorer l'utilisation de la logique floue dans les systèmes informatiques d'aide à la décision (SIAD). Les objectifs consistaient à déterminer les avantages et limites de la logique floue pour ce type d'application, et à établir une base méthodologique pour le développement de SIAD plus complets dans l'avenir. Le SIAD devait prendre des décisions de réforme de vaches laitières, basées sur les données de contrôle. La version finale du prototype, développée à partir de connaissances acquises auprès de trois experts, considérait cinq variables input. Pour la majorité de ces variables, les fonctions d'appartenance étaient spécifiques à chaque toupeau. Pour cette raison, les systèmes experts basés sur les ensembles flous sont flexibles et adaptatifs. Les mêmes bases de règles peuvent être utilisées dans divers contextes (e.g., troupeaux, régions, races), les inférences étant rendues spécifiques en ajustant les fonctions d'appartenance. Les résultats obtenus ont aussi démontré que les ensembles flous constituaient une alternative aux méthodes de classement des vaches actuellement utilisées par les agences de contrôle laitier. Les systèmes experts basés sur la logique floue pourraient requérir un moins grand nombre de règles que les approches conventionnelles pour fournir des variations aussi grandes au niveau des output.

INTRODUCTION

The animal breeding program is an integral part of any efficient dairy enterprise, the major components of which involve decisions concerning culling, replacement, and mating. Culling decisions are concerned with achieving a balance between genetic progress for specific traits of economic importance and an animal's length of productive life, while replacement and mating decisions cover issues such as age structure of the herd and the identification of traits which need to be improved in the overall genetic profile of animals in the herd. On an individual cow basis, breeding decisions can include, for example, optimal culling time, if and how she should be replaced, or which sire should be mated with her to better improve the genetic makeup of the next generation. In general, breeding decisions are complex and must take into account a large number of interrelated factors. Therefore, they require the consideration of many information sources, part of which is collected on the farm (e.g., milk production, fertility, and conformation), the other portion being made available through external agencies such as Artificial Insemination units, Dairy Herd Improvement Agencies (DHIA), and Breed Associations (e.g., genetic proofs for conformation and pedigree information). Potentially, the larger the number of factors considered, the better the conclusions can be.

While optimal breeding decisions require the processing of large volumes of information, the human ability to carry out this task is limited. Humans are able to make complex decisions through information analysis and reasoning, but, within a certain period of time, they can analyse only a finite number of data contained in tables and figures. Also, at some point during analysis, humans become saturated and can no longer absorb further information. In contrast, computers are particularly good at rapidly and endlessly carrying out well-structured, procedural tasks; in short, they are good information processors. Therefore, humans and computers both possess different strengths and, when trying to make an optimal decision, the best approach would seem to consist of combining computerized treatment and human thinking. With this approach, the computer software components form decision-support systems (DSS), which act as information pre-digesters and which establish the basis for final decisions by human managers. This approach can be applied to dairy cattle breeding, whereby the role of DSS would be, for example, to identify potentially problematic areas, to make preliminary diagnostics, and to formulate recommendations. The DSS would then carry out lower-level or repetitive analysis tasks and would leave higher level or final decisions to the farm manager.

To help the final decision-making process of humans, it is important to develop as much as possible the analytical ability of DSS. A good approach for this consists of capturing the expertise of specialists in well-defined, narrow areas, and then embodying it in software modules for addition to DSS. These modules, traditionally called "expert systems", can become important components of DSS for dairy cattle breeding and can be particularly helpful in diagnosing problems and suggesting recommendations for solving them. However, in order for expert systems to be able to mimic the reasoning and decision-making processes of specialists, they must be able to deal with vagueness, ambiguity, and uncertainty (Grinspan et al. 1994; Zadeh 1989). Many mathematical tools exist to help develop such software, e.g., tools based on confirmation, Bayesian probability, and fuzzy set theories (Graham and Jones 1988; Heatwole and Zhang 1990; Leung and Lam 1988; Zimmermann 1991). In the last decade, much attention has been given to fuzzy logic in various economic sectors and many commercial applications have been developed, based on this approach (Williams 1992). An advantage of fuzzy logic is that it allows for approximate reasoning and decision-making based on vaguely defined, linguistic variables, which generally characterize the decision-making processes of experts. Since fuzzy logic has been applied successfully in many agricultural areas (Ambuel et al. 1994; Edwards and Canning 1995; Grinspan et al. 1994; Thangavadivelu and Clovin 1991), it is reasonable to consider its use in the area of dairy cattle management.

The goal of this research was to investigate the use of a fuzzy-logic approach in the development of DSS for the area of dairy-cattle breeding. This investigation was done through the development of a prototype decision-support system that would make culling recommendations for individual cows, based on test day records. The specific objectives were to: 1) develop a prototype DSS, 2) analyse its performance as well as factors affecting it, and 3) establish a basis for the development of more complete DSS for breeding decisions.

MATERIALS AND METHODS

Fuzzy sets

In decision-making processes, experts often use qualitative terms to describe something about which they are reasoning. For example, an expert may qualify the milk production level of a specific cow as being 'medium' and her fertility as being 'low'. An important aspect is that no sharp boundary exists between the qualifiers used. For example, if an expert assesses a cow's average calving interval of 392 days as 'medium', he will most likely not describe a calving interval of 393 days as 'long'. At the same time, many experts might agree that an average calving interval of 420 is definitively 'long' and a calving interval of 390 days is definitively 'medium'. Fuzzy mathematics can be used to deal with such situations in a quantitative manner.

Fuzzy mathematics is based on fuzzy sets, which correspond to the qualifiers employed by specialists. Each possible qualifier used to describe a situation or an entity corresponds to one fuzzy set, with a series of fuzzy sets used to cover all possible qualification levels (e.g., from 'very short' to 'very long' in the case of 'average calving interval'). In contrast to classical set theory, where an element belongs either completely or not at all to a specific set (e.g., the set of cows with an average calving interval larger than 400 days), fuzzy set theory allows elements to belong partially to different sets. For this reason, once fuzzified, a numerical value is characterized by one or more fuzzy sets (i.e., the sets to which they belong), and a degree of membership in each of these sets.

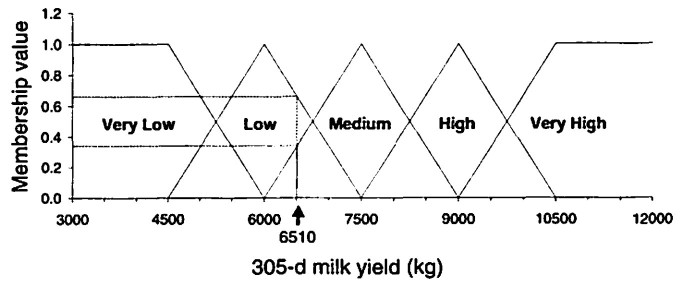

Fig. 1. Illustration of fuzzy sets characterizing 305-day milk yield.

An example of fuzzy sets which describe 305-d milk yield is shown in Fig. 1. In that figure, all cows with a milk production less than 4,500 kg belong entirely to the set 'VeryLow'. Between 4,500 and 6,000 kg, the cows belong to two sets: 'VeryLow' and 'Low'. As production increases from 4,500 kg, the degree of membership in the set 'VeryLow' decreases, while the membership in the set 'Low' increases. At 6,000 kg, the production is no longer 'VeryLow' and is definitively 'Low'. During a specific inference, the mapping of a numerical, crisp value to its fuzzy equivalent (i.e., the determination of the fuzzy sets to which it belongs and the associated degrees of membership) is called the "fuzzification" process. For example, in Fig. 1, a milk production of 6510 kg was classified simultaneously as 'Low' with a membership value (MV) of 0.66, and as 'Average' with a degree of membership of 0.34. This is denoted as:

MilkProd = (Low, 0.66), (Average, 0.34)

where "MilkProd" is the variable for 305-d milk yield of a specific dairy cow.

Fuzzy Logic

Fuzzy logic is concerned with drawing inferences using rules constructed with fuzzy variables. A fuzzy inference process consists of determining the MV of the fuzzy variables contained in a rule's conclusion. This value is a function of the degree of truth of the premise. The truth of the premise is a function of 1) the degree of membership associated with the values of the fuzzy variables contained in the premise and 2) the logical operator(s) that link(s) these variables. The use of the AND operator in fuzzy inferencing is illustrated in Fig. 2 where there is a set of hypothetical rules, applied to the production profile of a cow (as previously discussed in Fig. 1) and with a reproductive efficiency classified as VeryLow at 0.42 and Low at 0.58:

MilkProd = (Low, 0.66), (Average, 0.34)

ReprodEff = (VeryLow, 0.42), (Low, 0.58)

RULE 1:	IF MilkProd = Low AND ReprodEff = Low THEN Cull = Yes
RULE 2:	IF MilkProd = Average AND ReprodEff = VeryLow THEN Cull = Yes
RULE 3:	IF MilkProd = Average AND ReprodEff = Low THEN Cull = No

Fig. 2. Example of a rule set involving fuzzy variables.

where "ReprodEff" is the reproductive efficiency. The premise of the first rule in Fig. 2 is composed of two conditions. Using extension principles, the degree of truth of such a statement can be theoretically established using various methods (Zimmermann 1991). However, the most common method consists of taking the minimum MV of the fuzzy sets involved. In this case, since one of the milk production values is (Low, 0.66) and one of the reproductive efficiency values is (Low, 0.58), the premise of RULE 1 is true to a degree of 0.58. Consequently, the degree of membership of the variable "Cull" to the fuzzy set 'yes' is:

Cull = (yes, 0.58).

When the same operation is applied to RULE 2, it is now established that:

Cull = (yes, 0.34).

By applying this operation in RULE 3, it is concluded that

Cull = (no, 0.34).

During an inference process, many rules in a rule set (RS) can be fired since fuzzy variables can possess more than one value; in this example, three rules were fired. For this reason, different rules may assign a variable to the same fuzzy set but with different MV. For example, RULE 1 and RULE 2 led to the conclusion that the variable "Cull" belonged to the fuzzy set 'yes' with MV of 0.58 and 0.34 respectively. In this case, it is necessary to assign a final degree of membership to the set, and, again, this can be done using different methods. The simplest method consists of assigning the maximum degree of membership (i.e., the common fuzzy union operation). Other methods, derived from extension principles, or from confirmation theory, may also be used. For example, it may sometimes be convenient to use methods that are based on the supposition that two conclusions in agreement confirm and reinforce each other. One such method, called the 'probability sum method' has been proposed for confirmative certainty algebra in expert systems (Holsapple and Whinston 1986). Using this method in the previous example, the new degree of membership to the set 'yes' for the variable "Cull" would be:

$$\mu_{cull,\,yes} = 0.58 + 0.34 - (0.58 \cdot 0.34) = 0.72$$

where $\mu_{cull,\,yes}$ is the degree of membership to the set 'yes'. Thus, with this method, the firing of Rules 1 through 3 would lead to the following values for "Cull":

Cull = (yes, 0.72), (no, 0.34).

At the end of a fuzzy inference process, the output variables generally possess more than one value, as is the case in this example. Each value is composed of a fuzzy set and an MV representing the degree of membership to that set. However, a single and crisp value is usually needed in order to generate an overall conclusion. The process by which this single, crisp value is obtained is called "defuzzification". In the case of a discrete variable, the defuzzification approach can simply consist of choosing the fuzzy set with the largest MV. Using this approach, the final conclusion of our example would be a culling decision of 'yes' (since 0.72 > 0.34).

Development of the Software Prototype

In this project, the fuzzy logic-based DSS prototype for culling recommendations was constructed through readings and interviews with three local specialists. The knowledge acquisition was done in three stages: 1) informal discussion with the specialists, 2) variable definition and rule formulation, and 3) determination of membership functions. In the initial stages, interviews were conducted with the experts to discuss the factors (variables) on which the DSS should focus when making culling recommendations for individual cows, given available data. Subsequent interviews were conducted to determine links among variables and these links were used to design the architecture of the prototype DSS. The experts were also asked which descriptors (i.e., fuzzy sets) they would use for each variable, and which conclusions they would draw from various sets of conditions (i.e., from various combinations of fuzzy sets representing different variables). This led to the development of the rules. When the base system was developed, numerical values for all input variables were presented to the experts. These values represented individual test day records retrieved from a data set supplied by the local DHIA (Québec Dairy Herd Analysis Service). The experts were asked to qualify the numerical values, and this information was used to determine the membership functions.

The software system was developed with GURU 3.0 (Micro Data Base Systems 1991), running under the operating system OS/2. GURU is an integrated product which includes a rule-based expert system shell, a data base management system, a procedural language and spreadsheet capacity. GURU also allows the manipulation of fuzzy (multi-valued) variables and contains many tools for the manipulation of fuzzy sets and their associated MV. The expert system shell is conceived for making inferences using fuzzy variables and certainty factors, and various inference methods are available (backward, forward and mixed chaining, rule firing based on cost or priority criteria, etc.).

Performance Analysis

Once the software prototype was developed, numerical experiments were carried out to evaluate its performance within various contexts. The main objective was to gain some insight regarding the impact of input values, rules and membership functions on the outputs of the DSS. The experiments were performed with individual Holstein test day records representing 30 herds and 804 cows after edits for incoherent data. The first step of the performance analysis consisted of studying the inference processes produced by the DSS. This was done in detail for three arbitrarily chosen herds (Herds 9, 21 and 30) during the system development and the validation period. The second step consisted of analysing the overall results of the inference process for all 30 herds.

RESULTING PROTOTYPE

Overall decision-making process

For the purpose of this prototype, it was decided that the decisions of the DSS would be based on the production level of the cow, her reproductive efficiency and her lactation number. Also, the DSS was to be applicable only to multiparous cows. The variables involved in the overall decision-making and their interrelationships are shown in Fig. 3. Most variables were defined as fuzzy and the fuzzy descriptors, listed in Table 1, were used. Using a modular approach, the architecture of the DSS was organized so that the overall decision-making process would be carried out by three rule sets (RS), one to evaluate reproductive efficiency; a second to make a culling decision (in fuzzy terms); and a third to defuzzify the culling decision after analysis.

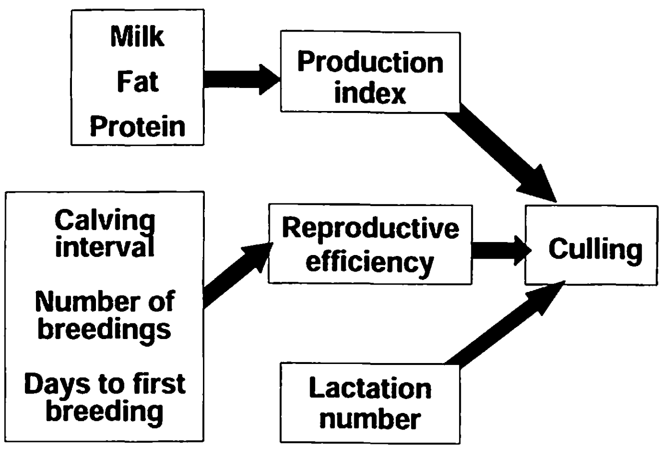

Fig. 3. Variables involved in the overall decision-making process.

Table I. Fuzzy sets used in the prototype decision-support system.

Variable name*	Fuzzy descriptors
Calving interval (CalvInt)	Short, Average, Long
Days to first breeding (FirstSrv)	Early, Average, Late
Number of breedings (NumServ)	Few, Average, Many
Reproductive efficiency (ReprEff)	Unsatisfactory, Poor, Good, Excellent
Production index (ProdIdx)	Very low, Low, Medium, High, Very high
Culling (Cull)	Yes, No

*The name indicated between brackets is used in the appendix.

The reproductive efficiency of the cow was evaluated by RS 1 using calving interval, days to first breeding and number of breedings. For the three input variables, the membership functions used to determine the sets in which a certain cow belonged and the degree of membership to each set were specific to each herd; they were constructed relative to herd average values for each variable as explained below. Twenty-seven rules were developed to cover all possible combinations of fuzzy sets associated with the three input variables. Using these rules, which are partially listed in the Appendix, RS 1 characterized reproductive efficiency with one of its four descriptors (Unsatisfactory, Poor, Good, or Excellent). During inferences, the minimum method was used for AND operations, and confirmative calculations across the rules were done using the probability sum method.

The second rule set (RS 2) required the values for reproductive efficiency, lactation number and production index to make a culling decision. The lactation number had two possible values: 'less than three' or 'three and greater'. The production index was a linear combination of milk, fat and protein production. This was done since the relative economic importance of each dairy component varies from region to region and from time to time. The production index was calculated using:

$$\text{ProdIndex} = \alpha * \text{Milk} + \beta * \text{Fat} + \gamma * \text{Protein} \qquad (1)$$

where:
ProdIndex = Production Index (kg),
Milk = average 305-Day milk production (kg),
Fat = average 305-Day fat production (kg),
Protein = average 305-Day protein production (kg), and
α, β, γ = weighting factors for milk, fat and protein, respectively.

For this prototype, the values for α, β and γ were 1, 10 and 20, respectively, which approximate current economic values for each of the three production traits. Once fuzzified, the production index was described using five qualifiers. The membership functions associated with the five sets were different for each herd and were determined using the herd's average production values. RS 2 was composed of 10 rules (listed in the Appendix) with two possible outputs: 'yes' and 'no'. Reproductive efficiency (in conjunction with lactation number) was *only* considered in decisions concerning cows with a production index of 'medium'. Thus, reproductive efficiency was used only to discriminate between 'medium' producing cows. During inferences, the 'Cull' variable's degree of membership in each fuzzy set (i.e., 'yes' and 'no') was determined using the minimum operation and the probability sum method.

Since inferences with the previous RS led, in most cases, to two values ('yes' and 'no'), with various degrees of membership for the two descriptors, a defuzzification of this conclusion was required for the DSS to furnish a specific, crisp, recommendation. This was achieved by a third rule set (RS 3), which assessed the value of the crisp variable 'Culling' (see Appendix). The three possible crisp values for the 'Culling' variable were: 'yes', 'no' or 'unknown'. As an example of defuzzification, RS 3 considered the difference between the degree of membership for the two possible values of the variable 'Cull' produced by RS 2; if the difference was too small, the decision was 'unknown'.

Membership functions

During the knowledge acquisition phase, the specialists consistently requested herd-average values when trying to make a culling decision for any specific cow. Early on in that phase, it was realized that the exact meaning of the descriptors used by the specialists, when presented with numerical values (e.g., calving interval values), varied from one herd to another. At the same time, each combination of descriptors (in the rules' premises) consistently led to the same conclusion, independent of the herd average values. The use of fuzzy logic then appeared to be ideal in mimicking the reasoning of the specialists in this context. Indeed, it was possible to use the same RS with different herds, making the inference herd-specific by adjusting the membership functions associated with the various fuzzy sets.

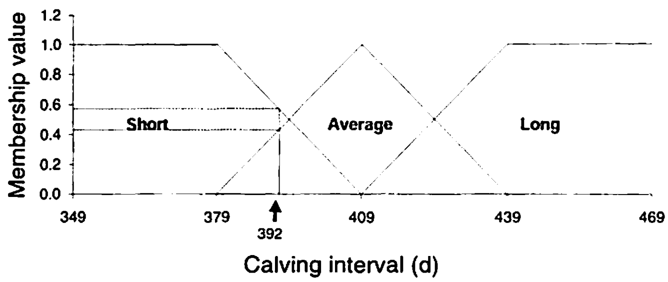

Fig. 4. Fuzzy sets for the variable 'calving interval'.

In the present DSS, it was assumed that fuzzy conditions 'normal', 'medium' or 'average' would correspond to the average values for a specific herd. The fuzzy sets and their respective membership functions were then defined relative to these average conditions, and applied to the variables calving interval, days to first breeding and production index. All fuzzy sets were defined using triangles and trapezoids. The MV for any set varied between 0 and 1, and the sum of the MV associated with the fuzzy sets at any point always equaled 1 as well. Except in the case of 'number of breedings', the middle fuzzy sets were centered on herd-average values. The position of the other sets was determined by what is referred to as a 'critical point', which corresponds to the summit for triangles or the point of discontinuity in the upper portion of trapezoids. For 'calving interval' and 'days to first breeding', the critical points were set by default at plus and minus 30 days, and at plus and minus 10 days, respectively. This is illustrated for 'calving interval' in Fig. 4, where the central value (409 days) corresponds to the average calving interval for a specific herd. In this example, a calving interval of 392 days (for a hypothetical cow belonging to this herd) would be simultaneously considered as (short, 0.57) and (average, 0.43).

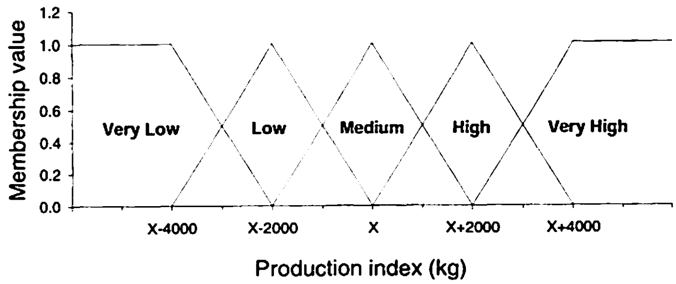

Fig. 5. Fuzzy sets for the variable 'production index'.

For the production index, the critical points were set to plus and minus one and two increments of 2000 kg (Fig. 5). For number of breedings, the position and the shape of the sets were the same for all herds; critical points were at 1, 1.5 and 2 breedings.

PERFORMANCE ANALYSIS

Individual inference process

An example of the inference process carried out by the DSS for an individual cow is shown in Fig. 6. The inputs to each of the three RS are also listed before each partial inference, and the rules are displayed in the order in which they were fired. It should be noted that many of the rules can be fired during an inference process. Also, some output variables may contain more than two values after an inference (e.g., reproductive efficiency). The effect of the confirmative calculation with the probability sum method can be seen in Fig. 6. For example, the firing of RULE 17 from the first RS should result in a reproductive efficiency value of (poor, 0.66). However, since reproductive efficiency has already been determined as being (poor, 0.18) after firing RULE 15, the degrees of membership are combined to produce a new value of 0.72.

Within herd analysis

The inference results obtained for the individual cows of three specific herds (9, 21 and 30) are listed in Tables II through IV. The results are presented in a decreasing order of the MV associated with the 'no' value. To a certain extent, the MV for 'yes' and 'no' are complementary: low MV for 'yes' generally correspond to high MV for 'no', and vice-versa. This leads to the interesting observation that, although the DSS was developed to generate culling decisions, the MV that it produces can be used to rank the cows. When considering the final culling decisions of the DSS (last column of Tables II through IV), it can be observed that the proportion of 'yes' cases is similar for Herds 9 and 21 (26% and 24%, respectively). This is true even though average production indices for these two herds differ considerably, as indicated in Table V. However, for Herd 30, the proportion of 'yes' cases is close to 50%, even though the critical values that were used were specific to this herd. Specifically, a 'no' decision was produced for all cows (except one) in this herd where the production index was larger than the average production index for the herd - 19223 kg - (see Table V), and a 'yes' decision was produced for all cows (except one) with a production index lower than the average production index for the herd. This is different to what occurs in Herds 9 and 21, where eight cows and five cows, respectively, with a production index lower than the herd average were not

FROM 1st RULE SET

Input variables are:

1) Calving interval =
(average, .98), (short, .02)

2) Days to first breeding =
(late, .85), (average, .15)

3) Total number of breedings =
(average, .66), (many, .34)

RULE 5 (fired)
Calving interval is (short, .02), days to first breeding is (average, .15), and number of breedings is (average, .66):
Therefore reproductive efficiency is (good, .02).

RULE 6 (fired)
Calving interval is (short, .02), days to first breeding is (average, .15), and number of breedings are (many, .33):
Therefore reproductive efficiency is (good, .04).

RULE 8 (fired)
Calving interval is (short, .02), days to first breeding is (late, .85), and number of breedings are (average, .66):
Therefore reproductive efficiency is (poor, .02).

RULE 9 (fired)
Calving interval is (short, .02), days to first breeding is (late, .85), and number of breedings are (many, .34):
Therefore reproductive efficiency is (poor, .04).

RULE 14 (fired)
Calving interval is (average, .98), days to first breeding is (average, .15), and number of breedings are (average, .66):
Therefore reproductive efficiency is (good, .18).

RULE 15 (fired)
Calving interval is (average, .98), days to first breeding is (average, .15), and number of breedings are (many, .33):
Therefore reproductive efficiency is (poor, .18).

RULE 17 (fired)
Calving interval is (average, .98), days to first breeding is (late, .85), and number of breedings are (average, .66):
Therefore reproductive efficiency is (poor, .72).

RULE 18 (fired)
Calving interval is (average, .98), days to first breeding is (late, .85), and number of breedings are (many, .34):
Therefore reproductive efficiency is (unsatisfactory, .34).

FROM 2ND RULE SET

Input variables are:

1) reproductive efficiency = (poor, .72), (unsatisfactory, .34), (good, .18)

2) production index = (high, .76), (medium, .23)

3) lactation number = 9

RULE 2 (fired)
Production index is (high, .76):
Therefore culling is (no, .76).

RULE 4 (fired)
Production index is (medium, .23), and reproductive efficiency is (good, .18):
Therefore culling is (no, .80).

RULE 6 (fired)
Production index is (medium, .23), reproductive efficiency is (poor, .72), and lactation number is larger than 3:
Therefore culling is (yes, .23).

RULE 7 (fired)
Production index is (medium, .23) and reproductive efficiency is (unsatisfactory, .34):
Therefore culling is (yes, .41).

FROM 3RD RULE SET

The input variable is:

1) Culling =
(no, .80), (yes, .41)

RULE 3 (fired)
Since the difference between degrees of membership to NO and YES is more than .20, and the culling suggestion of the 2nd expert system is NO with a membership value larger than .60, the cow should not be culled.

Fig. 6. Example of an inference process.

Table II. Inference results for Herd 9.

Cow identification	Calving interval (d)	Days to first breeding	Number of breedings	Lactation number	Production index (kg)	MV for 'yes'	MV for 'no'	Final culling decision
448	448	108	3.0	2	15838	0	98	no
453	412	95	1.7	3	15616	9	95	no
464	410	99	1.7	3	16357	0	92	no
473	395	93	1.5	2	17983	0	92	no
494	373	90	1.0	2	15914	0	83	no
420	466	87	2.5	3	17364	0	82	no
424	380	77	3.0	4	18278	0	82	no
418	381	80	1.7	4	17323	24	81	no
486	435	74	2.0	2	15432	18	81	no
488	393	84	1.5	2	15448	18	81	no
429	397	96	1.3	4	16214	55	77	no
430	393	71	1.5	4	15358	22	77	no
454	403	110	2.0	2	15356	22	77	no
474	362	81	1.0	2	15593	10	76	no
395	436	85	2.3	5	18629	0	75	no
465	352	84	1.0	3	18848	0	75	no
467	366	71	2.5	2	16809	0	75	no
490	405	84	1.5	2	15012	39	72	no
487	362	89	1.0	2	15007	40	59	unknown
431	416	103	1.3	4	16194	86	55	yes
384	414	68	2.3	5	15870	47	53	unknown
461	438	78	1.7	3	14234	78	34	yes
422	410	93	1.8	5	14197	87	19	yes
492	390	86	2.0	2	13981	91	15	yes
458	363	66	1.7	3	13669	93	0	yes
489	423	99	1.0	2	13379	83	0	yes
485	373	89	3.0	2	12981	75	0	yes

categorized as 'yes'. The disparity in the proportions of 'yes' cases is due to the difference in the distribution of the individual production indices within the herds, the membership functions associated with production index, and the rules which compose RS 2 in the DSS. In the case of Herd 30, the production index of the cows was very variable and the deviations were generally quite large about the herd average. Due to this variability on both sides of the herd average value (which is indicated by a high standard deviation in Table V), the degree of membership to the set 'medium' was generally not very high (i.e., from 0.05 to 0.61). Consequently, even when reproductive efficiency and lactation were considered in the inference process of 14 out of 21 cows (or 67%) for Herd 30, their influence on the final conclusions was small. This is due to the use of the 'minimum' operation in the premise of the

Table III. Inference results for Herd 21.

Cow identification	Calving interval (d)	Days to first breeding	Number of breedings	Lactation number	Production index (kg)	MV for 'yes'	MV for 'no'	Final culling decision
412	411	76	2.3	4	20171	0	94	no
430	388	86	2.3	3	18274	0	90	no
440	393	70	2.0	3	17991	3	89	no
450	359	77	3.0	2	20305	0	89	no
439	364	79	1.0	3	19159	0	85	no
453	447	70	2.5	2	20446	0	84	no
442	368	86	2.0	3	19167	0	83	no
355	411	90	1.8	6	20521	0	82	no
432	393	75	2.3	3	18704	0	81	no
434	430	75	2.3	3	20591	0	80	no
461	367	88	2.0	2	17706	17	77	no
427	431	98	2.3	3	18846	0	76	no
428	540	85	3.0	2	20795	0	76	no
426	369	72	1.7	3	17256	39	60	no
459	381	131	1.5	2	17193	42	57	unknown
454	357	83	2.5	2	17187	43	56	unknown
451	388	94	1.0	2	16503	77	22	yes
455	349	58	2.0	2	12974	100	0	yes
417	368	87	1.0	4	15297	76	0	yes
452	388	96	1.0	2	14847	76	0	yes
437	416	74	2.7	3	15181	75	0	yes

rules; when the degree of membership to 'medium' for production index is very small, then the probability that it be the element limiting the degree of truth of the whole premise is high. Therefore, in the case of Herd 30, the discrimination among cows was based mostly on production indices, which were evenly distributed about the average for the herd. In the case of Herds 9 and 21, decisions relating to cows with a lower production index than the herd average and which were not categorized as 'yes', were favourably influenced by reproductive efficiency, which was good to excellent, or by the lactation number when reproductive efficiency was poor or unsatisfactory (i.e., lactation number was less than or equal to three in this case). In these cases, the degree of membership to the set 'medium' for production index was generally large (from about 0.57 to 0.97) and, consequently, the weight of reproductive efficiency on the final decision was large.

For those cows for which the discrimination was based only on production, it would be expected that the degrees of membership to 'yes' or 'no' be directly proportional to the production level. For example, for the cows whose production index is higher by 2000 kg or more than the herd average, the degree of membership to 'no' should increase with an increase in production index. However, this is not necessarily what happened here. For example, the degree of membership to 'no' for Cows 412, 450, 453 and 355 of Herd 21 was inversely proportional to their production level (Table III). Similar patterns occurred for low-producing cows (e.g., Cows 417 and 437 of Herd 21) for which the degree of membership to 'yes' was lower for lower production levels. These patterns are due to a combination of factors: 1) the use of more than one set to describe the production index for deviations larger than 2000 kg and smaller than 4000 kg (i.e., 'high' and 'very high' for positive deviations, and 'low' and 'very low' for negative deviations), 2) the use of rules testing different conditions, but leading to the same conclusion (e.g., Rules 1 and 2 of RS 2), and 3) the method used for confirmative calculation. The situation can be best understood when looking at Fig. 7, which shows the variation of the final degree of membership to 'no' for

Table IV. Inference results for Herd 30.

Cow identification	Calving interval (d)	Days to first breeding	Number of services	Lactation number	Production index (kg)	MV for 'yes'	MV for 'no'	Final culling decision
506	689	87	2.5	2	25849	0	100	no
465	516	173	2.0	2	23171	0	97	no
501	361	93	2.0	2	19448	0	89	no
428	511	183	1.8	7	21000	11	88	no
493	388	98	1.3	3	19798	0	84	no
508	400	115	1.0	2	20779	0	82	no
481	397	97	1.3	5	19120	5	77	no
451	529	124	3.5	6	21984	0	76	no
513	448	87	1.5	2	20002	0	76	no
421	418	102	1.6	5	20178	0	75	no
469	455	84	2.0	3	20134	0	75	no
520	402	74	2.5	2	17819	70	29	yes
453	407	83	2.2	6	17704	75	24	yes
489	503	126	1.5	2	17650	78	21	yes
527	394	77	3.0	2	17648	78	21	yes
434	493	148	2.2	5	19612	80	19	yes
529	436	125	2.0	2	13016	100	0	yes
526	455	32	2.5	2	17120	94	0	yes
532	416	131	1.0	2	17036	91	0	yes
475	656	165	3.8	6	18841	84	0	yes
517	476	71	3.5	2	15782	80	0	yes

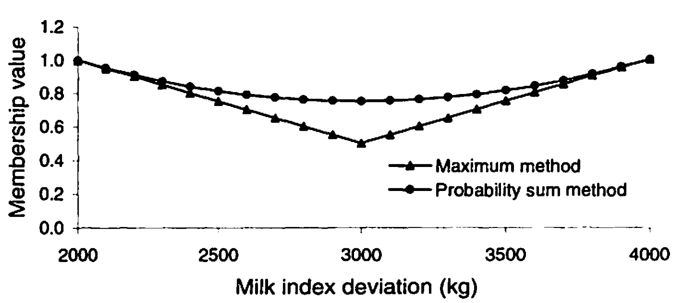

Fig. 7. Variation of confirmative membership values calculated with maximum and probability sum methods.

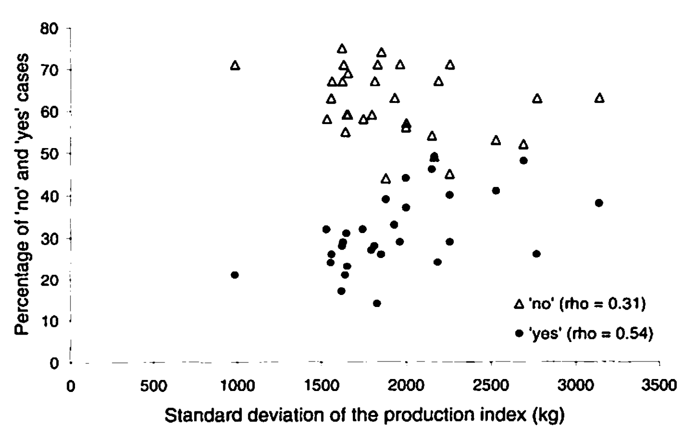

Fig. 8. Variation of the percentage of 'yes' and 'no' resulting from DSS's inferences, as a function of the herd standard deviation of the production index.

Table V. Herds' average and standard deviation for calving interval, days to first breeding, number of breedings and production index.

Herd number	Number of cows	Calving interval (d)		Days to first breeding (d)		Number of breedings (d)		Production Index (kg)	
		Avg.	(Std.)	Avg.	(Std.)	Avg.	(Std.)	Avg.	(Std.)
1	21	393	31	78	13	1.8	0.6	16288	1961
2	20	379	90	73	11	1.5	0.6	14812	2251
3	19	422	86	92	19	1.1	0.2	15485	1851
4	18	405	53	75	22	2.4	0.9	14580	1812
5	29	384	27	71	10	1.9	0.6	14189	1645
6	62	389	38	83	19	1.5	0.5	13524	1555
7	22	431	62	90	28	1.8	0.6	13148	1652
8	19	407	45	78	12	1.9	0.6	15252	1745
9	27	400	29	87	12	1.8	0.6	15810	1560
10	19	404	45	87	18	1.7	0.7	12856	1529
11	25	386	32	84	17	1.9	0.8	17420	1995
12	13	369	22	78	17	1.1	0.2	14859	1652
13	7	427	35	115	24	1.8	0.6	20341	1829
14	24	417	57	81	15	2.5	1.0	14892	2147
15	18	381	19	80	7	1.8	0.6	13499	1877
16	22	404	44	91	17	2.1	0.9	14324	1793
17	18	396	23	82	14	1.9	0.6	15027	1622
18	16	393	47	80	12	2.6	1.5	17017	3134
19	14	422	92	101	50	1.2	0.3	13911	2254
20	19	416	59	109	28	1.8	0.8	16947	2768
21	21	396	43	83	15	2.0	0.6	18053	2185
22	24	401	36	90	15	1.9	0.4	13859	1928
23	17	390	32	83	12	2.2	0.8	15693	2526
24	29	414	39	94	19	1.8	0.6	16590	1638
25	14	401	57	65	12	2.6	1.0	13984	985
26	12	418	62	93	24	1.7	0.4	13947	1620
27	33	398	46	82	11	2.0	1.1	14407	2161
28	49	394	53	83	22	1.4	0.5	14040	1996
29	21	400	36	77	9	2.3	1.0	16251	1626
30	21	464	84	108	37	2.1	0.8	19223	2688

deviations, about the herd average production index, varying between 2000 kg and 4000 kg. Within this interval, both Rules 1 and 2 of RS 2 are fired; therefore, since the conclusion for the rules is the same, there is a confirmative calculation of the final degree of membership to 'no'. Figure 7 contains the results of confirmative calculations based on two methods: 1) the probability sum of method, used in this project, and 2) the maximum method, more traditionally used in fuzzy logic. For both methods, an increase in the production level leads to a decrease in the degree of membership in the first portion of the interval, to reach a minimum when the degree of membership to both 'high' and 'very high' equals 0.50. Similar (but inverted)

Table VI. Inference results for 30 herds.

Herd number	Number of cows	No		Yes		Unknown	
		Count	%	Count	%	Count	%
1	21	15	71	6	29	0	0
2	20	9	45	8	40	3	15
3	19	14	74	5	26	0	0
4	18	12	67	5	28	1	6
5	29	17	59	9	31	3	10
6	62	39	63	15	24	8	13
7	22	13	59	5	23	4	18
8	19	11	58	6	32	2	11
9	27	18	67	7	26	2	7
10	19	11	58	6	32	2	11
11	25	14	56	11	44	0	0
12	13	9	69	3	23	1	8
13	7	5	71	1	14	1	14
14	24	13	54	11	46	0	0
15	18	8	44	7	39	3	17
16	22	13	59	6	27	3	14
17	18	12	67	5	28	1	6
18	16	10	63	6	38	0	0
19	14	10	71	4	29	0	0
20	19	12	63	5	26	2	11
21	21	14	67	5	24	2	10
22	24	15	63	8	33	1	4
23	17	9	53	7	41	1	6
24	29	16	55	6	21	7	24
25	14	10	71	3	21	1	7
26	12	9	75	2	17	1	8
27	33	16	49	16	49	1	3
28	49	28	57	18	37	3	6
29	21	15	71	6	29	0	0
30	21	11	52	10	48	0	0

patterns are found with negative deviations. This had no impact on the final culling decisions of the DSS. However, it led to unsatisfactory rankings for some of the cows in the given production level interval. It should be pointed that with the maximum method, the ranking of the cows would have been worse; the final results would even have changed from 'no' to 'unknown' for the cows whose degree of membership to 'no' would have been smaller than 0.60.

Between herd analysis

The inference results for all 30 herds are presented in Table VI. The culling rate (i.e., instances where 'yes' is recommended) varies from 14% to 49%, with an average of 31%. It should be noted that this average value is close to what is often suggested for dairy farms. The proportion of 'no' fluctuates between 44 and 75%, with an average value of 62%, while the percentage of 'unknown' cases varies between 0 and 24, the average being 8%. Since it had previously been observed that the distribution of final decisions (e.g., the overall culling rate) was influenced by the distribution of input values about herd averages, it was hypothesized that a relationship may exist between the distribution of the outputs and the standard deviation of the input variables. Thus, the proportion of 'no' and 'yes' cases was plotted against the standard deviation of the production index for the 30 herds. This plot is shown in Fig. 8 and a certain trend can be observed (higher standard deviations often lead to larger proportions of 'yes' cases and to smaller proportions of 'no' cases), even if the relationships are not very strong (the linear correlation coefficients were 0.54 and 0.31 for 'yes' and 'no' cases respectively). This can be explained by the fact that, the higher the standard deviation of the production index, the higher the probable number of cows in a herd for which reproductive efficiency was not considered (the correlation coefficient between standard deviation of the production index and percentage of cows for which the deviation of the production index about herd average was larger than 2000 kg was 0.76). Also, production index alone is used to discriminate cows for which the deviation of the production index about herd average is larger than 2000 kg. Since, on average, the cows were fairly well distributed about herd averages, herds with high standard deviations of the production index tended to generate final results that were more evenly distributed (i.e., the number of 'yes' cases were closer to the number of 'no' cases). For other variables such as calving interval, similar trends were not detected. This absence of relationship is most likely due to the fact that a variable such as calving interval affects the final inference results indirectly through reproductive efficiency; also, reproductive efficiency was only considered in the case of medium producing cows. However, it should be mentioned that, on average, 71% of the cows had a production index deviation about herd average of smaller than 2000 kg. Consequently, on average, reproductive efficiency was considered in the inference process of 71% of the cows, and production index alone did not explain the inference results for those cows.

DISCUSSION

An interesting observation from this study was that DSS, based on fuzzy sets, are flexible and can easily be adapted to various contexts without changing the rules contained in the knowledge base. The parameters that characterize specific inference processes (i.e. membership functions) are external to the RS and can be defined at each inference. For example, the same sets of rules could be used for farms with different levels of productivity or objectives, by adapting the DSS to each farm through a simple modification of the membership functions associated with the fuzzy sets. This means that a fuzzy logic based DSS could easily be adapted to different regions or to various breeds of cattle, thereby prolonging its period of usefulness compared to a situation where crisp variables are incorporated into rules. Such a DSS might arguably remain valid despite changes due to genetic improvement or even improved management.

The analyses, performed with the prototype DSS under various conditions, permitted insight into the intrinsic characteristics of the DSS itself, and the general uses of fuzzy systems. Analyses displayed the fact that results were influenced by many factors. Indeed, the inference processes relied heavily on membership functions, values of the input variables and rules. The performance analyses also demonstrated that, consequently, the response of the DSS under various conditions was not always easy to explain and the establishment of input-output relationships was not a straight-forward process. This was true despite the fact that the structure of the DSS was relatively simple and leads to the assessment that fuzzy systems may exhibit some complex behaviour, even if they possess a simple structure. Therefore, response analysis constitutes an important step when developing fuzzy systems; it permits a better understanding of the interactions among rules, membership functions, input values and outputs.

Another, related observation is that, fuzzy systems (even simple ones) can lead theoretically to an infinite number of output values (before defuzzification), due to the combined use of rules and continuous membership functions. For example, in this project, despite its simple structure, the DSS could lead to any degree of membership to the two sets associated with the output variable. In order to obtain a similar behaviour with crisp sets, a larger number of rules would have been required so as to cover many possible cases within the variation ranges of the input variables. If the number of rules, required to embed an area of expertise, is considerably lower with fuzzy sets than with crisp sets, then the implementation and debugging of fuzzy expert systems could potentially be easier, while leading to lower development time and costs as well. However, when developing fuzzy systems, one must be aware that the potential number of rules increases rapidly when combining variables with a large number of sets. For example, three variables that are each represented by 5 fuzzy sets will lead to 125 possible rules. In this case, redundant rules (or rules with impossible outcomes) may need to be eliminated from a fuzzy system to ease its management and to accelerate the inference process. This would also simplify the tuning of the system, which needs to be done through adjustments of the membership functions.

Although the system which has been developed to date does not take account of all factors that a producer might consider when making culling decisions, it has shown promising results. As it is, it can already be used for cow rating, with minor adjustments. Indeed, one of the interesting aspects observed during this project is that the degrees of membership associated with the descriptors of the fuzzy variable 'culling' could be used for ranking of cows within a herd. Such an approach might constitute an interesting alternative to rating methods currently used by some DHIA. This may even constitute an approach which would be preferable for many producers, who would themselves make the final decisions about culling the cows or not; they would use the ranks produced by the DSS in conjunction with other factors to make their final decision. However, in order to obtain a ranking that would reflect better the production level, the prototype DSS requires some adjustments. One possible adjustment consists of eliminating two fuzzy descriptors for production level ('very high' and 'very low'). Also, in order to increase the discriminating ability of reproductive efficiency, the membership functions associated with production level may be adjusted. For example, the shape for the set 'medium' could be changed from a triangle to a trapezoid, or to some type of logistic functions (which would permit smoother variations in ranking). Another possible modification is that the membership functions for each herd be established not only about the herd averages, but also as a function of the distribution of the input variables about herd averages. For example, membership functions for production level could be a function of the standard deviation, as it was shown to be correlated to the number of cows for which reproductive efficiency was considered in the inference processes. This approach would help homogenizing the consideration of reproduction efficiency among herds. Additional modifications to the current DSS are necessary so as to make it applicable to primiparous cows. For these cows, reproductive efficiency can not be evaluated; only fertility can be assessed, based on the number of breedings. Also, 305-day values are not known for those cows, and decisions must be based on projection factors. Therefore, for primiparous cows, both fertility and production will add uncertainty in the inference processes, and this uncertainty will need to be dealt with.

CONCLUSION

Fuzzy logic permits the encoding of knowledge using a terminology which is close to that used by experts, i.e., based on linguistic descriptions. This eases the encoding of knowledge, which may accelerate the development and the implementation of DSS that accurately reproduce the reasoning of specialists. This would make them more rapidly available to producers. Since fuzzy logic seems promising in the development of knowledge-based systems, it will be used to develop a more complete DSS aimed at helping dairy producers to establish their herd breeding policy, their herd breeding program and their breeding program for individual cows. Results obtained in this project constitute a basis for the elaboration of a methodology and of a complete framework that will help in that area.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada and the Cattle Breeding Research Council of Canada. Data were provided by the Québec Dairy Herd Analysis Service (Programme d'analyse des troupeaux laitiers du Québec).

REFERENCES

Ambuel, J.R., T.S. Colvin and D.L. Karlen. 1994. A fuzzy logic yield simulator for prescription farming. *Transactions of the ASAE* 37(6): 1999-2009.

Edwards, D.B. and J.R. Canning. 1995. Developing a radio-controlled log skidder with fuzzy logic autonomous control. *Transactions of the ASAE* 38(1):243-248.

Graham, I. and P.L. Jones. 1988. *Expert Systems: Knowledge, Uncertainty and Decision.* New York, NY: Chapman and Hall.

Grinspan, P., Y. Edan, H.E. Kahn and E. Maltz. 1994. A fuzzy logic expert system for dairy cow transfer between feeding groups. *Transactions of the ASAE* 37(5):1647-1654.

Heatwole, C.D. and T.L. Zhang. 1990. Representing uncertainty in knowledge-based systems: confirmation, probability, and fuzzy set theories. *Transactions of the ASAE* 33(1):314-323.

Holsapple, C.W. and A.B. Whinston. 1986. *Manager's Guide to Expert Systems Using GURU.* Homewood, IL: Dow Jones-Irwin.

Leung, K.S. and W. Lam. 1988. Fuzzy concepts in expert systems. *Computer* 21(9):43-56.

Micro Data Base Systems. 1991. GURU - Integrated components manual. Lafayette, IN: Micro Data Base Systems, Inc.

Thangavadivelu, S. and T. S. Clovin. 1991. Decision analysis using linguistic approximation in an agricultural environment. In *Proceedings of the Simulation Multiconference on Artificial Intelligence and Simulation. Simulation Series 23*(4):98-103.

Williams, T. 1992. Fuzzy logic is anything but fuzzy. *Computer Design* 31(4):113-127.

Zadeh, L.A. 1989. Knowledge representation in fuzzy logic. *IEEE Transactions on Knowledge and Data Engineering* 1(1):89-100.

Zimmermann, H.-J. 1991. *Fuzzy Set Theory and its Applications*, 2nd edition. Norwell, MA: Kluwer Academic Publishers.

APPENDIX

Rule Set 1

RULE 1:
IF: CalvInt = "Short" AND FirstSrv = "Early" AND NumServ = "Few"
THEN: ReprEff = "Excellent"

RULE 2:
IF: CalvInt = "Short" AND FirstSrv = "Early" AND NumServ = "Average"
THEN: ReprEff = "Good"

.....

RULE 7:
IF: CalvInt = "Short" AND FirstSrv = "Late" AND NumServ = "Few"
THEN: ReprEff = "Good"

RULE 8:
IF: CalvInt = "Short" AND FirstSrv = "Late" AND NumServ = "Average"
THEN: ReprEff = "Poor"

.....

RULE 13:
IF: CalvInt = "Average" AND FirstSrv = "Average" AND NumServ = "Few"
THEN: ReprEff = "Good"

RULE 14:
IF: CalvInt = "Average" AND FirstSrv = "Average" AND NumServ = "Average"
THEN: ReprEff = "Good"

.....

RULE 19:
IF: CalvInt = "Long" AND FirstSrv = "Early" AND NumServ = "Few"
THEN: ReprEff = "Poor"

RULE 20:
IF: CalvInt = "Long" AND FirstSrv = "Early" AND NumServ = "Average"
THEN: ReprEff = "Poor"

.....