A fuzzy expert system for the human reliability analysis of crews in simulated nuclear emergency procedures

E. Zio*, P. Baraldi and M. Librizzi

Dipartimento di Energia
Politecnico di Milano
Via Ponzio 34/3
20133 Milan, Italy
E-mail: enrico.zio@polimi.it
E-mail: piero.baraldi@polimi.it
E-mail: massimo.librizzi@polimi.it
*Corresponding author

Abstract: The use of Human Reliability Analysis (HRA) for a particular safety assessment is still a difficult problem. In order to perform a comparison of the available methods, the best approach is simulation. In this regard, the pilot study proposed by the Organisation for Economic Co-operation and Development (OECD) Halden Reactor Project (HRP) is intended to provide a first guidance in HRA methods evaluation through experimental data on crew performance in simulated scenarios. The quantitative evaluation of the results of these simulations in terms of crew performance and Human Error Probabilities (HEPs) is quite a difficult task. In this paper, a fuzzy expert system for systematically assessing crew performance is presented. The feasibility of the method is proved on a case study concerning a scenario of an incomplete scram in a Boiling Water Reactor (BWR).

Keywords: human reliability analysis; HRA; fuzzy expert system; FES; crew performance assessment; human error probability; HEP.

Reference to this paper should be made as follows: Zio, E., Baraldi, P. and Librizzi, M. (2010) 'A fuzzy expert system for the human reliability analysis of crews in simulated nuclear emergency procedures', *Int. J. Nuclear Knowledge Management*, Vol. 4, No. 1, pp.42–58.

Biographical notes: Enrico Zio completed his BS in Nuclear Engineering from Politecnico di Milano, Italy in 1991, MSc in Mechanical Engineering from UCLA in 1995, PhD in Nuclear Engineering from Politecnico di Milano in 1995 and PhD in Nuclear Engineering from MIT in 1998. He is the Director of the Graduate School of Politecnico di Milano, a Full Professor of Computational Methods for Safety and Risk Analysis, and an Invited Lecturer at various Master's and PhD programmes in Italy and abroad. His research topics include the analysis of the reliability, safety and security of complex systems under stationary and dynamic operation, particularly by Monte Carlo simulation methods and cellular automata, and also the development of soft computing techniques (neural networks, fuzzy logic and genetic algorithms) for safety and reliability applications, system monitoring, fault diagnosis, and optimal design. He is a coauthor of two international books and more than 200 papers in international journals. He has served as a referee of ten international journals.

Piero Baraldi completed his BS and PhD in Nuclear Engineering from the Politecnico di Milano, Italy in 2002 and 2006, respectively. He is a Postdoctoral Researcher at the Department of Nuclear Engineering at the Politecnico di Milano. His research efforts are devoted to the development of soft computing techniques (neural networks, fuzzy logic and genetic algorithms) for system health monitoring, fault diagnosis, prognosis and human reliability qualification and quantification.

Massimo Librizzi holds a PhD in Radiation Science and Technology and a degree in Nuclear Engineering from the Politecnico di Milano, Italy. His research interests are the application of computational methods (fuzzy logic, Monte Carlo simulation, cellular automata, *etc*.), evaluation of the reliability of complex systems (phase-mission systems and networks systems) and human reliability analysis. He works at the Nuclear Technical Area of Enel in Italy.

1 Introduction

Various Human Reliability Analysis (HRA) methods (*e.g*., Swain and Guttman, 1983; Moieni *et al.*, 1994; Mosleh and Chang, 2004; Reer *et al.*, 2004; Cooper *et al.*, 1994; Hollnagel, 1998) are applied in the Probabilistic Safety Assessments (PSAs) of complex systems such as nuclear power, chemical and process plants. These methods differ in their approaches, underlying models and aims, due to the context in which they were developed; all present advantages and drawbacks when applied to different situations. This calls for procedures of comparison and validation in order to guide the choice of the appropriate approach for a given situation (Zio, 2009).

As a step forward in this direction, a test has been designed at the Halden Man-Machine Laboratory's (HAMMLAB) facility of the Organisation for Economic Co-operation and Development (OECD) Halden Reactor Project (HRP) with the aim of providing the technical basis for the comparison of the performance and findings of different HRA methods. The study is intended to be a pilot test aimed at establishing a first guidance in the assessment and improvement of HRA methods through the information gained by simulator data (Dang *et al.*, 2007; Broberg *et al.*, 2008; Lois *et al.*, 2008; Bye, 2006a–b).

The initial testing study will focus on the performance of a number of crews in the HAMMLAB experimental facility which reproduces the digital instrumentation and control systems and equipment in actual nuclear Pressurised Water Reactors (PWRs) and Boiling Water Reactors (BWRs). The comparison between the results of the HRA methods and the actual experimental performance is to be made in terms of the 'driving factors' that most influence human performance and of the estimated values of Human Error Probability (HEP).

Although the bulk of the comparative analyses focuses on the qualitative modelling of crew behaviour, the quantitative results play a fundamental role as they eventually need to be input in the PSA. In this respect, the experiments are expected not to yield statistically significant experimental values of HEPs due to the very high level of performance of the trained crews and the small number of simulated sessions; henceforth, a surrogate model for retrieving experimental values of HEPs is needed.

In this work, a Fuzzy Expert System (FES) (Klir and Yuan, 1995) is purported to address the problem of HEP assessment and crew performance evaluation from the experimental simulations. Fuzzy Logic (FL) provides a systematic framework for the representation and treatment of the linguistic, subjective expert judgements used in the characterisation of the crew performance in the experiment. On these premises, the FL model is designed to reproduce an expert evaluation of crew performance, providing a degree of action success/unsuccess in the simulated scenario.

To demonstrate the proposed method, Scenario 4.2 'incomplete scram/start of the boron system' of Laumann *et al.* (2005) is taken as case study. Although it is not the scenario chosen for the pilot study under development, it is believed to include most of the relevant aspects worth of analysis.

The paper is organised as follows: Section 2 illustrates the simulated scenario and the experimental data. The FES is described in Section 3 and a case study concerning the evaluation of a crew is presented in Section 4. The results are discussed in Section 5 and conclusions on the work are drawn in the last section.

2 Simulated scenario

The scenario under analysis is the 'incomplete scram/start of the boron system' in a BWR which is initiated by a failure in the main feedwater pumps and a leakage in a feedwater pipe, leading to a low level in the reactor vessel and the subsequent scram command by the feedwater isolation signal (Laumann *et al.*, 2005). However, in the scenario, 30 control rods are not inserted because of failure to open two scram valves. A more detailed description of the scenario can be found in the appendix.

The scenario has been simulated with the HAMMLAB BWR simulator (Laumann *et al.*, 2005). Seven crews (labelled from A to G), each one composed by a turbine operator, a reactor operator and a shift supervisor, participated in the exercise. The data from the experiment were analysed in terms of the completion time of the important actions in the scenario and by in-depth qualitative analyses of the crews' communications. Additional recorded data include the debriefing of crews after the simulation in terms of important Performance-Shaping Factors (PSFs), the notes of some observers of the simulation on the difficulties encountered by the crews, the decision processes undertaken, the communications and other characteristics that seem to have contributed to the crew performance.

The most important crew activity in the scenario described above is to bring the reactor to a subcritical state by starting the boron system manually. Other intermediate and mandatory target actions in the emergency procedure were monitored and their occurrence recorded together with their timing. Table 1 summarises the occurrence times, expressed in minutes, of nine target actions (a '0' value indicates that the crew did not perform the action).

Table 1 Crews' occurrence times of target actions in minutes

Crew	*Start boron system manually*	*Open scram valve 354VB1*	*Insert SIRM*[a]	*AUX pump A*	*AUX pump B*	*AUX pump D*	*Open valve 322VD2*	*Open relief valve 314VA17*	*Open relief valve 314VA23*
A	3.18	0	5.59	14.14	5.02	31.48	1.49	2.29	2.4
B	1.04	5.51	0	8.15	8.04	7.45	1.36	1.51	2.07
C	4.27	7.22	3.56	19.43	9.15	11.45	23.44	2.53	2.54
D	1.19	1.14	9.53	0	4.49	4.16	5.16	2.42	3.02
E	3.04	9.19	3.26	0	1.47	11.37	13.16	2.48	3.04
F	2.01	0	6.49	17.27	2.58	2.35	0	0	0
G	11.43	0	18.56	10.53	1.3	2.21	0	0	19.04

Note: [a] Source and Intermediate Range Monitor (SIRM).

On the basis of the available qualitative and quantitative data, an expert has been required to evaluate crew conduct and judge its performance, *e.g.*, in terms of 'Well' (W) or 'Not Well' (NW) done. A further detailed judgement addressed the timing of action accomplishment in terms of Early (E) or Late (L). A typical expert assessment is as follows:

> The crew solved the scenario *well*. They started the boron system, inserted SIRM detectors and closed the pressure relief valve (314 VA2) *fast*. They also *fast* send a field operator to open the scram valves in group B1 (354 VB1) and group C1 (354 VC1). The crew was *late* in starting the auxiliary feedwater system and the level was down to 2.1 meters. …. The crew did not trigger an evacuation alarm or clearly informed the field operators. The reactor operator did not clearly report first checks.

It is important to note that the choice of the target actions is done at the design phase of the simulation. On the contrary, the judgement scale regarding E or L action is not clearly defined. It may refer, in absolute terms, to agreed time windows or be founded on relative terms, *i.e.*, by comparison between crews or even between actions of the same crew.

3 The fuzzy expert system for crew performance assessment

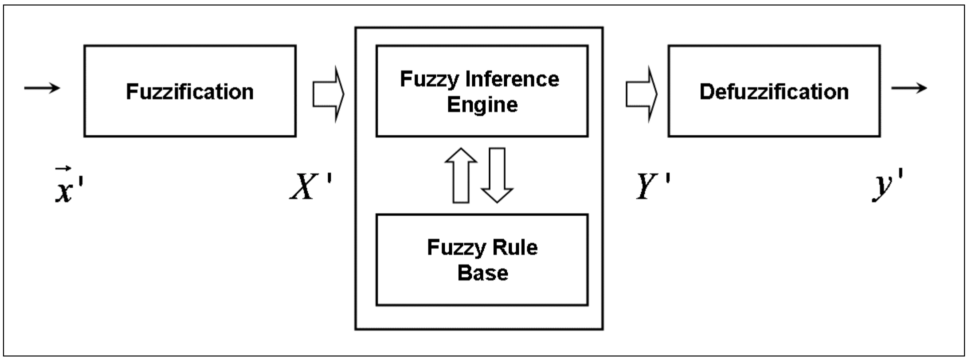
As seen before, the judgement on the crew performance is expressed in qualitative, linguistic terms, generally smoothing out differences between crews which turn out to merit the same final assessment, albeit holding different conducts during the simulation. For a more definite qualification, it seems important to assess the crew performance as a 'degree' of success, a higher degree being associated to a crew that accomplishes all the target actions in due time and a lower degree to a crew that does so with bad timing. This can be done by building an FES based on the available experimental data and the corresponding judgement on the crew performance.

3.1 An overview on fuzzy expert systems

In general terms, an expert system is an Artificial Intelligence (AI) system designed to mimic how experts solve problems. The decision-making process must be explicitly modelled and the relevant ambiguities and uncertainties must be properly taken into consideration.

To account for the ambiguities and uncertainties affecting the subjective assessment procedure, overlapping Fuzzy Sets (FSs) can be used to quantitatively represent the input and output values. Such expert system framework, called FES, allows quantifying of the linguistic judgements by the expert, *e.g.*, 'cold', 'warm', 'hot', with regard to the state of a variable x, *e.g.*, 'temperature'. To this aim, a Membership Function (MF) $\mu_{X^v}(x)$ is associated to each linguistic label X^v defining the variable. Such MF quantifies the degree to which x belongs to X^v for all values of x in its range of variability called Universe of Discourse (UOD).

Figure 1 The fuzzy reasoning process



Four main elements constitute the fuzzy reasoning process underpinning an FES. These are illustrated schematically in Figure 1 with reference to the mapping of an m-dimensional input vector $\vec{x}' = \{x_k', k = 1,2,...,m\}$ into a one-dimensional output value y' (Klir and Yuan, 1995).

1 *The fuzzification module*. It converts numeric (crisp) values into FSs in order to take into account the uncertainties and ambiguities inherent in the input data. In the application of the present work, the data are in the form of point values $x_k', k = 1,2,...,m,$ with associated intervals $[a_k', b_k']$ reflecting the uncertainty. The conversion of x_k' into the FS X_k' is obtained by constructing the MF $\mu_{X_k'}(\cdot)$, supported on the interval $[a_k', b_k']$ and equal to unity in correspondence to x_k'.

2 *The Fuzzy Rule Base* (FRB): In an FES, the expert knowledge is modelled into a set of N_r heuristic rules capturing the relationships between the different values of the input and output variables. The generic j-th fuzzy rule ($j = 1,2,\ldots,N_r$) is made up of a number of antecedents and a consequent linguistic statement related by appropriate fuzzy connections:

$$\text{Rule } j\text{: If } (x_1 \text{ is } X_1^j) \text{ and } (\ldots) \text{ and } (x_m \text{ is } X_m^j) \text{ then } (y \text{ is } Y^j).$$

The linguistic variables x_k, $k = 1,2,\ldots,m$ are the antecedents, represented in terms of the FSs X_k^j with MFs $\mu_{X_k^j}(x_k)$ on the UOD X_k. The linguistic variable y is the consequent represented by the FS Y^j with MF $\mu_{Y^j}(y)$ on the UOD Y.

3 *The fuzzy inference engine.* The fuzzy inference engine receives the fact constituted by the variables sent by the fuzzification module, *viz.*:

Fact: x_1 is $X_1^{'}$ and … and x_m is $X_m^{'}$

where $X_k^{'} \subseteq U_{x_k}$ = an FS on the UOD of the k-th variable x_k. The fuzzy inference engine compares these data with those in the antecedents of the FRB and arrives at the conclusion $y = Y^{'}$, where $Y^{'} \subseteq U_y$ = an FS on the UOD of the output. The FL procedure used to associate the conclusion to a given fact may vary depending on the modelling approach adopted. In this work, the Mamdani procedure is employed (Huang *et al.*, 2001).

4 *Defuzzification module.* The output of the fuzzy inference engine consists of an FS $Y^{'}$ with compact support (η_1, η_2), whose MF is $\mu_{Y^{'}}(y)$. In many instances, one is interested in a crisp number $y^{'}$ representing the information encoded in the output FS $Y^{'}$. This conversion, called defuzzification, may be done in several ways (Klir and Yuan, 1995), the most common being the Center of Area (COA) method (Dubois, 1997):

$$y^{'} = y_{COA} = \frac{\int_{\eta_1}^{\eta_2} y \cdot \mu_{Y^{'}}(y) \cdot dy}{\int_{\eta_1}^{\eta_2} \mu_{Y^{'}}(y) \cdot dy}.$$

The crisp number $y^{'}$ thereby obtained can be taken as the numerical output resulting from the given input $\vec{x}^{'}$.

3.2 Design of an FES for crew performance assessment

The first step in the FES design process consists in the selection of the relevant actions that the crews are required to perform. The guidelines for this selection are as follows:

- the knowledge base of the scenario definition that includes the list of target actions that the expert has chosen as intermediate goals
- a thorough analysis of the expert judgements of crew evaluation in order to infer what actions the expert believes to be the keys in obtaining the scenario goal.

With reference to the scenario illustrated in Section 1, the set of nine target actions considered relevant for the scenario goal is given in Table 2. In the following, these actions are treated as equally important, although this may not be the case and a more accurate description of the scenario and crew performance could be achieved on the basis of a definition of primary, secondary and auxiliary tasks.

In the most simplistic modelling of the expert assessment of each action, it is conceivable to introduce a pair of linguistic terms E and L for characterising the action timing.

Table 2 Target actions for the FES

Action number	*Action description*
1	Start boron system
2	Open valve 354VB1
3	Insert SIRM detector
4	Activate pump A
5	Activate pump B
6	Activate pump D
7	Open valve 322VD2
8	Open relief valve 314VA23
9	Open relief valve 314VA17

Table 3 reports the sorting of the crews with respect to the occurrence time of the nine actions and their subdivision in terms of the judged E and L conditions. Crew D is not considered in the design phase of the model in order to use its action timing for testing.

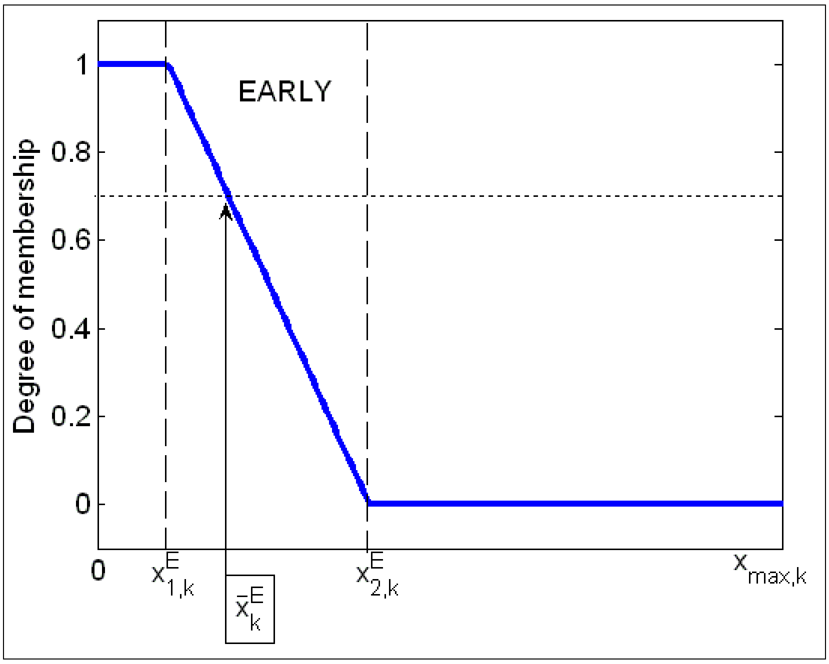
Table 3 Crew sorting in terms of occurrence time to each action and subdivision in E and L linguistic categories

Action number	*Crew with an E judgement*	*Crew with an L judgement*
1	B F E A C	G
2	B C	E A F G
3	E C A F	G B
4	B G	A F C E
5	G E F A B C	
6	G F B	E C A
7	B A E C	F G
8	B A E C	F G
9	B A E C	F G

This way of proceeding naturally leads to a fuzzy formulation in which each action is an antecedent linguistic variable, x_k, $k = 1,2,..,9$, described in terms of the linguistic descriptors E and L, mathematically expressed as FSs $X_k^v, v = E, L$ on a time interval called UOD in fuzzy terminology.

The partitioning of the UOD into the supports of the MFs $\mu_{X_k^v}(x_k)$ associated to the FSs X_k^v is defined based on the available experimental data (action occurrence time) and the associated expert judgement (E or L action). In doing so, one must care to reduce as much as possible the arbitrary parameters while using all available information contained in the data. The method here propounded considers for each action k, a sorting of the timing data from the fastest to the slowest crew and extracts the largest crew action time $\overline{x}_k^E$ judged positively, *i.e.*, E, by the expert and the shortest time $\overline{x}_k^L$ judged negatively, *i.e.*, L.

Figure 2 Trapezoidal MF of FS 'EARLY' (see online version for colours)



The MFs have been arbitrarily chosen of trapezoidal shape with parameters $(0, 0, x_{1,k}^E, x_{2,k}^E)$, $k = 1,2,\ldots,9$ for the FSs of the linguistic variable E and $(x_{1,k}^L, x_{2,k}^L, x_{\max,k}$, $x_{\max,k})$, $k = 1,2,...,9$ for the FSs of the linguistic variable L. For the generic action $k = 1,2,...,9$, the parameters $x_{1,k}^E$ and $x_{2,k}^E$ are calibrated so as to satisfy the following conditions (Figure 2):

- The MF value of the FS X_k^E evaluated at time $\bar{x}_k^E$ is equal to 0.75, *i.e.*, $\mu_{X_k^E}(\bar{x}_k^E) = 0.75$; this choice amounts to giving a significant weight (≥0.75) to all time data judged E by the expert, the largest time value (*i.e.*, the slowest crew) still holding a 0.75-degree membership to E.
- At $x_{1,k}^E, \mu_{X_k^E}(x_{1,k}^E) = 1$ and the distance between the lower base value $x_{1,k}^E$ and $\bar{x}_k^E$ is set equal to the standard deviation σ_k of the available crew occurrence times, *i.e.*, $\bar{x}_k^E = x_{1,k}^E + \sigma_k$; this allows definition of the slope of the trapezoidal MF.
- $x_{2,k}^E$ is set by linear interpolation at $\mu_{X_k^E}(x_{2,k}^E) = 0$.

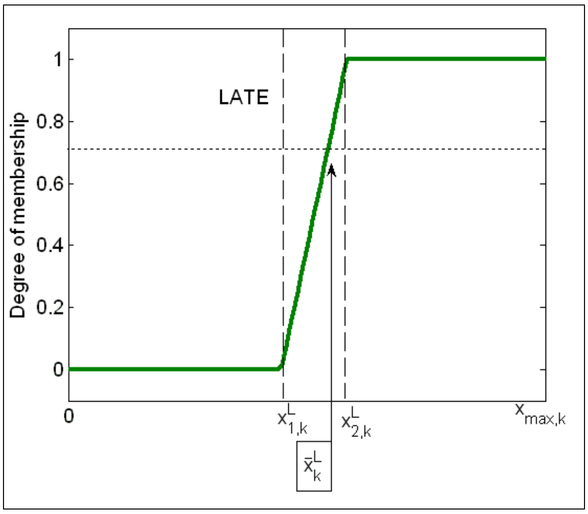
The parameters $x_{1,k}^L$ and $x_{2,k}^L$ are computed similarly (Figure 3):

- The MF value of the FS X_k^L evaluated at time $\bar{x}_k^L$ is equal to 0.75, *i.e.*, $\mu_{X_k^L}(\bar{x}_k^L) = 0.75$; this choice amounts to giving a significant weight (≥0.75) to all time data judged L by the expert, the shortest time value (*i.e.*, the faster crew) still holding a 0.75-degree membership to L.

- At $x_{2,k}^L, \mu_{X_k^E}(x_{2,k}^L) = 1$ and the distance between the lower base value $x_{2,k}^L$ and $\bar{x}_k^L$ is set equal to the standard deviation σ_k of the available crew occurrence times, *i.e.*, $x_{2,k}^L = \bar{x}_k^L + \sigma_k$; this allows definition of the slope of the trapezoidal MF.

- $x_{1,k}^L$ is set by linear interpolation at $\mu_{X_k^E}(x_{1,k}^L) = 0$.

Figure 3 Trapezoidal MF of FS 'LATE' (see online version for colours)



The parameter $x_{\max,k}$ is set as double the largest time value of the action occurrence time among the crew data available. For computational reasons, this value is also assumed as the numerical value of action time associated to those crews who actually did not perform the action k. Finally, the range $(0, x_{\max,k})$ defines the UOD of the variable.

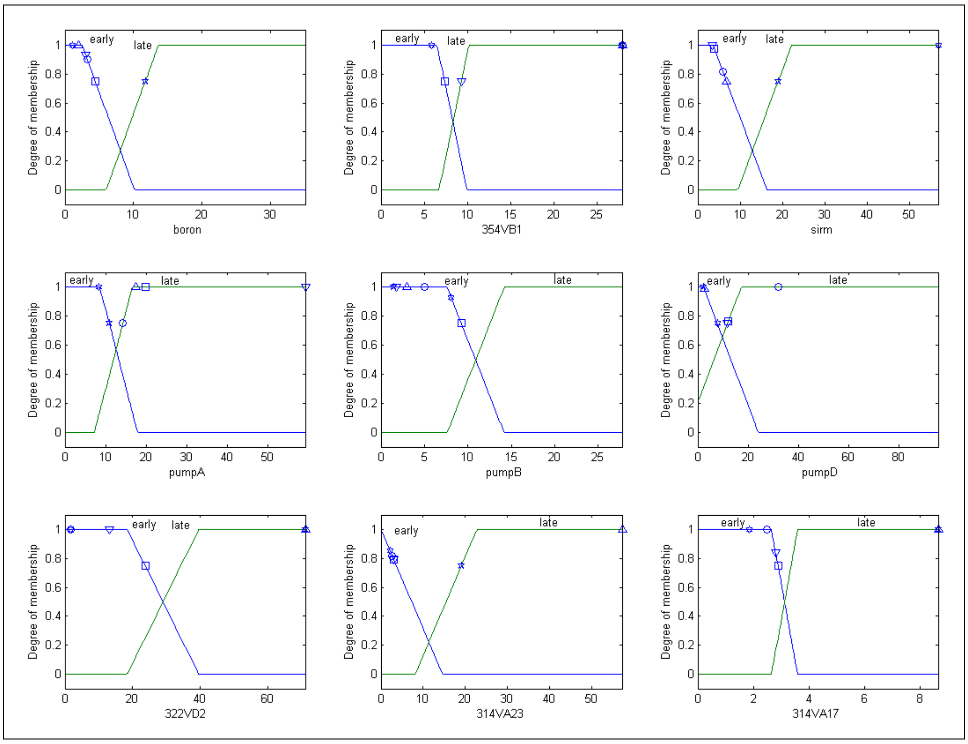
In case there is no information on $\bar{x}_k^L$, *i.e.*, if for action k there is no indication regarding the fastest among the L crews, the MF of the FS 'LATE' is derived to give values complementary to the MF of the FS 'EARLY', *i.e.*:

$$\mu_{X_k^L}(x_k) = 1 - \mu_{X_k^E}(x_k), \forall x_k.$$

Figure 3 reports the partitioning of the time UODs of the nine actions. It seems worthwhile stressing that the design of the MFs relevant to the modelling is derived completely from the information contained in the experimental data, albeit under some arbitrary assumptions. Thanks to this procedure, the membership degree assignments turn out to be physically reasonable. Clearly, by construction, the slowest of the E crews is given a membership 0.75 to E; the fastest of the L crews is given a membership of 0.75 to L.

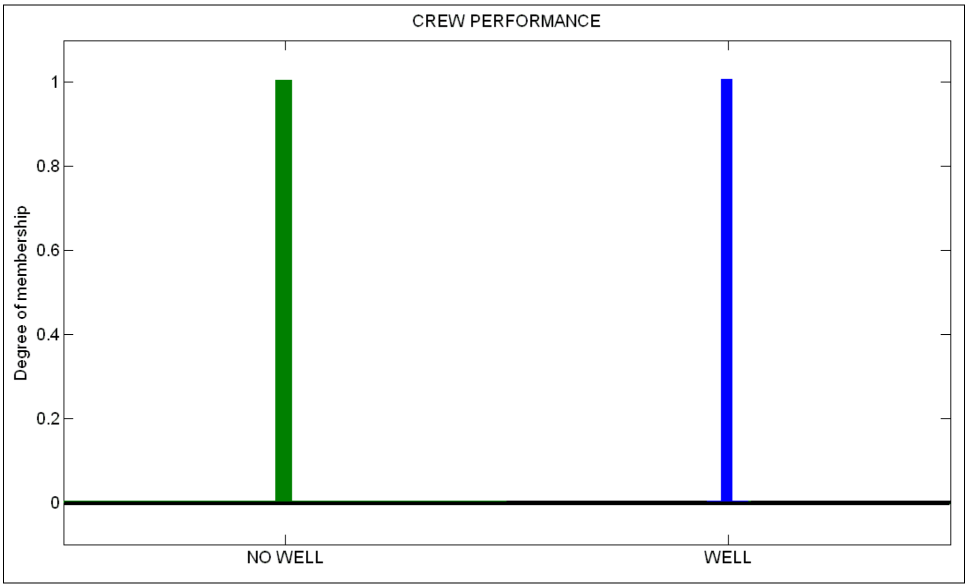
The second step in the design process of the FES consists in the definition of the output of the model, *i.e.*, the consequent of the fuzzy rules. In the present case, the consequent of interest is the linguistic variable 'crew performance', a discrete variable characterised by NW and W linguistic judgements representing the two possible evaluations by the expert on the crew performance. In fuzzy terminology, the consequent linguistic variable z is described by the discrete FSs Z^v, $v = NW, W$ in Figure 5.

Figure 4 Trapezoidal MFs of the FSs E and L for the nine target actions (see online version for colours)



Note: The points indicated on the curves correspond to the values of the seven crews.

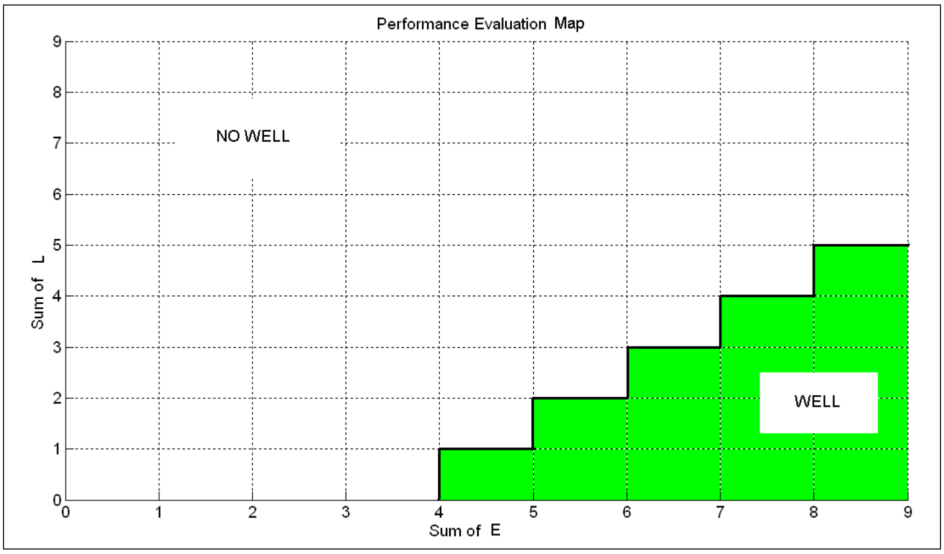
Figure 5 Discrete FS of the consequent 'crew performance' (see online version for colours)



The last design step of the FES is the construction of the FRB, setting the relationships between the nine action antecedents x_k and the crew performance consequent z which model the expert judgement process. To build the 29 rules representative of all possible relationships between the nine antecedents, conditions (E, L) and crew performance (NW, W) by expert interview is unfeasible not only in terms of time but also because the experts would most likely be capable of distinguishing few relationships possibly mainly those associated to extreme conditions of the antecedents or related to known experimental tests.

To overcome this difficulty, in this work, a mapping procedure inspired by the Cognitive Reliability Error Analysis Method (CREAM) for HRA is proposed. For each rule, the numbers of antecedents with FSs E and L are separately computed and then input as abscissa and ordinate of the map in Figure 6, respectively. If the representative point of the antecedents of the rule falls in the white area, the rule is assigned an NW crew performance consequent; if the point falls in the shaded area, the crew performance consequent is W.

Figure 6 Crew performance consequent map (see online version for colours)



Notes: The white area represents the NW performance zone whereas the shaded area represents the W zone. The thick line represents the boundary.

By so doing, the elicitation process is limited to the definitions of the boundary of the two consequent zones; this can be, for example, done by assigning for each number of E actions in abscissa the lowest number of L actions that would lead to judging the crew performance as NW.

Again, it seems worthwhile to remark that the procedure does not make any distinction in the importance of the different actions as mentioned at the beginning of this section. The relevance of the different actions in the emergency procedure can be accounted for by associating weights to the actions which are summed in the E and L counts as proper, so that a highly weighted action executed E (L) will push the judgement on crew performance in the W (NW) zone of the map.

Obviously, it is expected that the zone of NW crew performance lies on the north-west corner of the map characterised by a large number of L-executed actions and a small number of E-executed actions, whereas the zone of W crew performance is expected to be positioned in the south-east region with a large number of actions executed timely and, at most, only a few L ones.

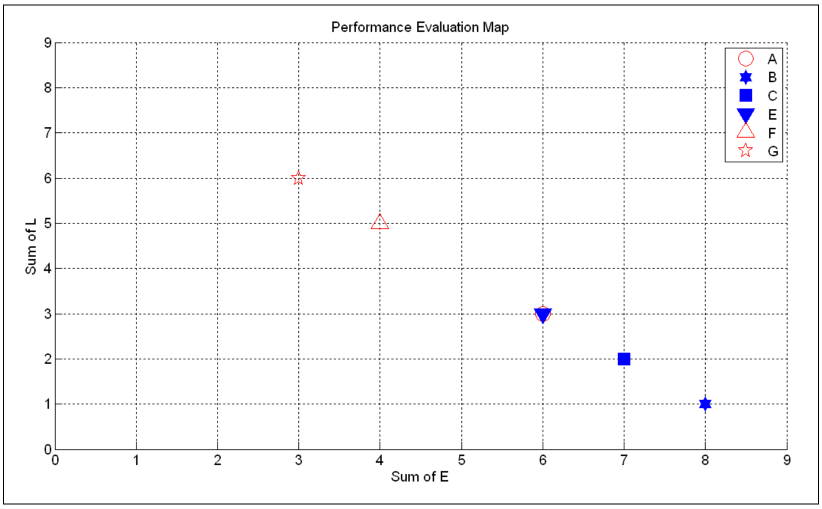
In the absence of the possibility of eliciting the boundary of crew performance from an expert, an alternative way to proceeding, here followed, amounts to building the boundary directly from the available experimental data. For each crew, the number of actions executed early and late is counted and the associated crew performance judgement is recorded (Table 4). Again, the data concerning crew D are not used in this design phase either.

Table 4 Total number of E and L actions and associated crew performance judgements

Crew	*(E,L)*	*Judgement*
A	(6,3)	NW
B	(8,1)	W
C	(7,2)	W
E	(6,3)	W
F	(4,5)	NW
G	(3,6)	NW

Each pair of numbers of E and L actions in Table 4 is a point on the map, labelled as W/NW.

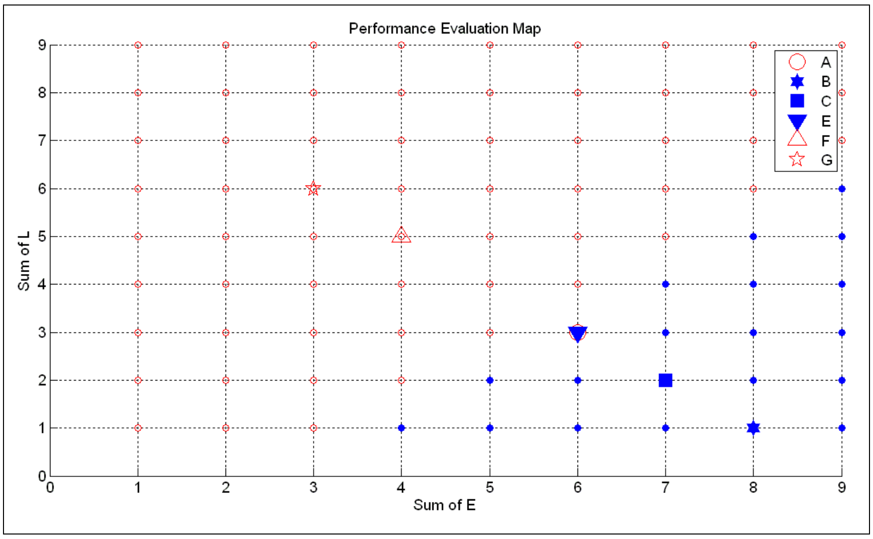
Figure 7 Performance evaluation map (see online version for colours)



Note: Void symbols indicate NW judgements and filled symbols indicate W judgements.

These representative points can serve to guide the splitting of the map into the two W/NW zones (Figure 8).

Figure 8 Performance evaluation map split into the two W/NW zones (see online version for colours)



4 Crew performance assessment

Once the FES is built, it can be fed as input with the occurrence times of the nine target actions as performed by a new crew, namely D in Table 1. These times are represented as nine antecedent singletons (Figure 9) and then elaborated by a Mamdani fuzzy inference procedure on the FRB to infer the final FS Z' of Figure 10, representing the degree of possibility that the crew performance can be judged W or NW.

Figure 9 Antecedent singletons fact FSs (see online version for colours)

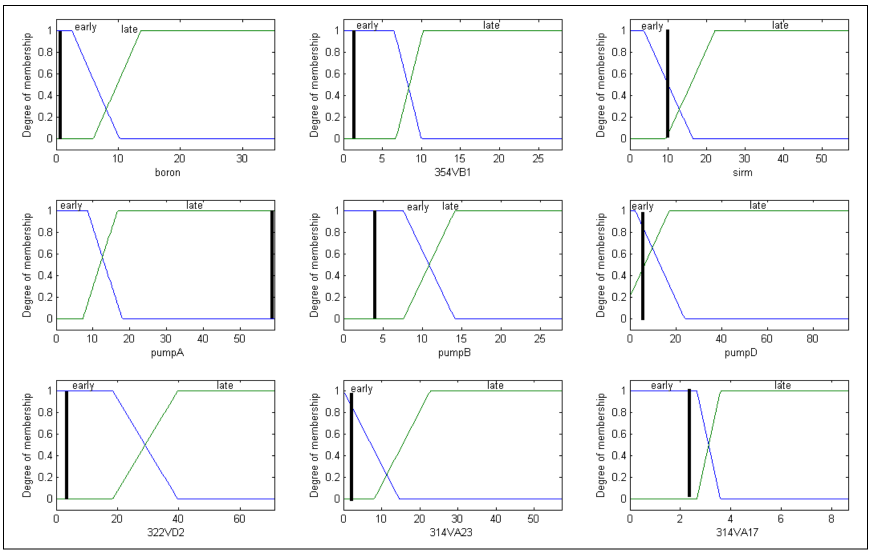


Figure 10 Inferred consequent FSs

Crew D turns out to have a degree of possibility of 0.40 being judged W and of 0.00 of being judged NW.

5 Discussion of the results

The results of the FES evaluation of the new crew D are consistent with the experimental expert judgement (W). Furthermore, notice that the use of the structured FES guarantees the repeatability and traceability of the assessment. This represents an advancement with respect to the current crew evaluation practice which relies to a great extent to arbitrary expert judgement. Repeatability of the assessment comes from the fact that the proposed method is based on an explicitly structured computable model whereas traceability is assured by a systematic elicitation procedure based on empirical data, *i.e.*, once the model is built, it is possible to generate an explanation of its conclusions, tracing the steps of its reasoning within the explicit and direct fuzzy rule formalism. In other words, the if-then linguistic rules that contribute to the FES output assessment can be retrieved and constitute the basis for the understanding of the reasons behind the model conclusions and for identifying the causes of possible disagreements with the expert intuition and expectation.

Finally, notice that the Mamdani-style inference leads to an output FS with membership less than one for all values in its support; this may not be desirable and could be avoided by considering different types of fuzzy rules, for example, the gradual rules (Dubois and Prade, 1992).

6 Conclusions

The FES presented in this paper is a first step in the investigation of the use of FL modelling for the evaluation of human performance from simulated data. Simulation is very useful for scrutinising human behaviours that can result in human performance-related problems since it allows researchers to systematically observe human behaviours in coping with hypothetical accidents. Possibly, simulators are the only way of observing human performance under emergencies and their insights are invaluable in spite of the several discrepancies from the real situation (*e.g.*, the level of stress and/or fidelity, *etc.*).

How to retrieve information useful for HRA from simulation studies is still open. In this sense, an FES such as the one proposed here provides a quantitative assessment of the qualitative judgements produced by the experts evaluating the simulation. This is done through an explicit representation of the crew performance assessment and a traceable and transparent inference process. It can be applied to evaluate the performance of a crew and it can be updated by experts as new experimental data for crew performance evaluation.

The FES construction is deeply founded on the simulation data which can replace the expert input when they are unavailable and allow adjusting potential inconsistencies in the evaluations. Obviously, the data-based model provides an approximation which does not reproduce exactly the expert assessment of a crew performance due to the inevitable introduction of arbitrary parameters in the model definition.

Also, the FES design process must be supported by an HRA expert analyst who defines the human behaviour model to be associated to the experimental evidence, *e.g.*, in terms of how the actions must be treated depending on the human way to tackle them, for example, a long action occurrence time due to the thinking phase of decision making might be considered positively because it provides a better understanding of the problem whereas a short reaction time might be interpreted as hazardous; on the contrary, in other safety recovery actions, rapidity might be required.

The numerical output of the expert assessment on the different crews can be further manipulated in order to produce aggregate results on the average performance in the simulated scenario. Future research should address the possibility of defining a relationship between this aggregate result and the quantitative attributes of interest in PSA, like the HEP.

Acknowledgements

The authors wish to thank the Institute for Energy Technology, OECD Halden Reactor Project for providing the data and Drs. Luca Podofillini and Vinh N. Dang of the Paul Scherrer Institute for their constructive comments which have allowed improving the paper.

This work is funded by the project Virthualis (FIS5-1999-00250, funded by the European Union). The views expressed in this article are solely those of the authors.

References

Broberg, H., Hildebrandt, M., Massaiu, S., Braarud, P.O. and Johansson, B. (2008) 'The international HRA empirical study: experimental results and insights into performance-shaping factors', *Proceeding of PSAM9, Probabilistic Safety Assessment and Management*, Hong Kong, China, 19–22 May.

Bye, A. (2006a) 'Minutes', *Workshop Meeting on HRA Method Investigation*, *PSAM8*, MTO-IND-294, OECD Halden Reactor Project, New Orleans, 13–14 May.

Bye, A. (2006b) *Workshop on Empirical Testing of HRA Methods*, MTO-IND-312, OECD Halden Reactor Project, 4–7 December.

Cooper, S.E., Ramey-Smith, A.M., Wreathall, J., Parry, G.W., Bley, D.C., Luckas, W.J., Taylor, J.H. and Barriere, M.T. (1994) *A Technique for Human Error Analysis (ATHEANA)*, NUREG/CR-6093, Washington, DC: US Nuclear Regulatory Commission.

Dang, V.N., Bye, A., Lois, E., Forester, J., Kolaczkowski, A.M. and Braarud, P.O. (2007) 'An empirical study of HRA methods – overall design and issues', *Joint 8th IEEE HFPP Conference on Human Factor and Power Plants and 13th HPRCT Annual Workshop on Human Performance, Root Cause, Trending, Operating Experience, Self-Assessment*, Monterey, California, 26–31 August, pp.368–372.

Dubois, D. (1997) *Fuzzy Sets and Systems: Theory and Applications*, Orlando, FL: Academic Press, Inc.

Dubois, D. and Prade, H. (1992) 'Gradual inference rules in approximate reasoning', *Inform. Sci*, Vol. 61, pp.103–122.

Hollnagel, E. (1998) *Cognitive Reliability and Error Analysis Method*, Oxford: Elsevier.

Huang, D., Chen, T. and Wang, M.J.J. (2001) 'A fuzzy set approach for event tree analysis', *Fuzzy Sets and Systems*, 16 February, Vol. 118, No. 1, pp.153–165.

Klir, G.J. and Yuan, B. (1995) *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall.

Laumann, K., Braarud, P.Ø. and Svengren, H. (2005) 'The task complexity experiment 2003/2004', HWR-758, August.

Lois, E., Dang, V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P.Ø., *et al.* (2008) 'International HRA empirical study – pilot phase report. Description of overall approach and first pilot results from comparing HRA methods to simulator data', HWR-844, OECD Halden Reactor Project, Halden, Norway.

Moieni, P., Spurgin, J. and Singh, A. (1994) 'Advances in human reliability analysis methodology – Part I: frameworks, models, and data', *Reliab Eng Syst Saf*, Vol. 44, pp.27–55.

Mosleh, A. and Chang, Y.H. (2004) 'Model-based human reliability analysis: prospects and requirements', *Reliab Eng Syst Saf*, Vol. 83, pp.241–253.

Reer, B., Dang, V.N. and Hirschberg, S. (2004) 'The CESA method and its application in a plant-specific pilot study on errors of commission', *Reliab Eng Syst Saf*, Vol. 83, pp.187–205.

Swain, A.D. and Guttman, H.E. (1983) *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*, NUREG/CR-1278.

Zio, E. (2009) 'Reliability engineering: old problems and new challenges', *Reliab Eng Syst Saf*, Vol. 94, pp.125–141.

Appendix

The 'incomplete scram/start of the boron system' scenario

In this scenario, the operators are supposed to start the boron system in order to reduce the reactor power. Before the initiating event, the plant is at full power in normal conditions, with one of the four emergency pumps of the safety system under planned maintenance. The safety system is designed with four electrically and physically separated auxiliary trains in two out of four logic.

The initiating sequence is made up of three main events:

1 the failure of the main feedwater pump

2 a leakage in the pipe of the standby main feedwater pump

3 the unavailability of 30 control rods.

Following the failure of the main feedwater pump, the reactor power is reduced to 55% and the standby main feedwater pump is called in action. However, the leakage in the feedwater pipe leads to a feedwater isolation signal that on one hand gives order to scram the reactor and on the other hand stops all feedwater pumps and closes the steam and feedwater lines. The reactor power is greater than 2% since not enough control rods are inserted in the reactor. To cope with this situation, an emergency procedure is followed. The operators must manually start the boron system, control the effectiveness of the reaction reduction (monitoring the power of the reactor) and secure an appropriate cooling of the reactor core (monitoring the level of water in the vessel). To supervise the neutron flux and thus check that the reactor is reaching a subcritical condition, the operators must insert an SIRM into the core. The crew must monitor the effectiveness of the auxiliary pumps system that automatically starts after the failure in the main feedwater pumps. However, the first pump is under an electrical failure that requires the help of the maintenance personnel *in situ*; the second pump must be activated manually because the reactor level is too high for the automatic trigger signal; the third pump is under maintenance and, finally, the last pump must be manually activated from the control room due to a failure.