POLITECNICO DI MILANO

Facoltà di Ingegneria dei Processi Industriali

Corso di Laurea in
Ingegneria Nucleare

MAINTENANCE MODELING BASED ON EFFECTIVE AGE, FUZZY LOGIC AND MONTE CARLO SIMULATION

Relatore: Prof. Enrico ZIO

Co-relatori: Ing. Piero BARALDI
Ing. Ludovic BENETRIX
Ing. Michele COMPARE
Ing. Antoine DESPUJOLS

Tesi di Laurea di:

Andrea BALESTRERO Matr. 709221

Anno Accademico 2009 - 2010

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In this thesis we address the problem of building a model in support of maintenance optimization of a generic component of an industrial power plant, when the only available information is that elicited from an expert. In particular, we introduce the concept of effective age for modeling the degradation process affecting the component. To be practical, we propose a systematic way to elicit the information from the expert within the theoretical framework of Fuzzy Logic (FL) to deal with his/her qualitative statements. Then we apply a hybrid approach based on the Monte Carlo (MC) simulation and FL to evaluate the performance of a given maintenance policy. Finally, we show how the proposed methodology can be applied in practice, by way of a real case study dealing with a medium voltage test network.

Keywords: Condition-Based Maintenance (CBM), Maintenance optimization, Effective Age, Fuzzy Logic, Expert elicitation, Monte Carlo simulation.

RIASSUNTO DELLA TESI

In questa tesi si propone una metodologia che permette di costruire un modello per l'ottimizzazione della manutenzione di un generico componente di un impianto industriale nel caso particolare in cui le uniche informazioni disponibili siano quelle ricavate a partire dalla conoscenza di un esperto. Nel **Capitolo 1** si sottolinea come questa situazione, comune in molti contesti industriali, necessiti di sviluppare nuove soluzioni modellistiche. Infatti, le informazioni elicitate dall'esperto sono soggettive, qualitative e molto spesso espresse in forma implicita; per esempio, l'esperto generalmente si affida a variabili linguistiche come "alto", "spesso", "lento" e fornisce valutazioni qualitative come "le condizioni ambientali sono buone". Questo tipo di informazioni deve essere appropriatamente interpretato, rappresentato e propagato tramite un modello adeguato. Per fare ciò, grazie alla sua capacità di gestire dati imprecisi ed espressi in modo vago, in questo lavoro ci muoveremo all'interno dell'inquadramento teorico offerto dalla logica fuzzy.

Nel **Capitolo 2** si descrive in dettaglio la modellizzazione del processo di invecchiamento del componente e dell'attività di manutenzione. In tutta generalità, l'efficacia di un modello di degrado – fondamentale per lo sviluppo di un adeguato programma di manutenzione – è tanto maggiore quanto più esso è in grado di cogliere le specificità del componente che derivano dalle particolari condizioni operative nelle quali il componente lavora. Infatti, in analogia con quanto avviene agli esseri umani, due componenti simili (cioè dello stesso lotto di produzione) aventi la stessa "età anagrafica" saranno probabilmente in stati di degrado differenti se hanno operato in condizioni diverse o sotto differenti stati di stress. Per modellizzare questa situazione introduciamo il concetto di "età efficace" di un componente proprio per descrivere il fatto che lo stato di degrado può evolvere più o meno rapidamente a seconda delle condizioni operative di funzionamento, caratterizzate dall'esperto tramite la definizione di un adeguato numero di “fattori influenzanti” (umidità, temperatura, etc...). Dunque, l’età efficace può essere considerata a tutti gli effetti come un indicatore dello stato di degrado del componente, lasciando al modello il compito di trovare la relazione fra l’età efficace e le condizioni di funzionamento. Un’evoluzione continua dello stato di degrado, tuttavia, mal si adatta alla percezione tipicamente discreta che l’esperto ha nel giudicare l’invecchiamento del componente; in mancanza di un monitoraggio continuo, infatti, l’esperto tenderà a ragionare in termini di stati discreti (e finiti) di degrado, probabilmente associati a misure ed osservazioni saltuarie. Per modellizzare questa situazione poniamo un numero adeguato di soglie sull’età efficace per definire altrettanti stati di degrado; in questo modo, si stabilisce una relazione fra il valore dell’età efficace del componente e il suo stato di degrado corrente. Il passo successivo è quello di trovare un legame fra la velocità con la quale il degrado avanza e le condizioni di funzionamento. L’elicitazione diretta della velocità di degrado da parte dell’esperto non è fattibile in pratica, dal momento che tale concetto è piuttosto intangibile. Piuttosto, ciò che si può agevolmente fare è chiedere all’esperto di valutare, per una data combinazione dei fattori influenzanti (cioè per una data condizione di funzionamento) la lunghezza dell’intervallo di tempo che il componente impiega per cambiare il proprio stato di degrado, e ricavare da questa informazione il valore della velocità di degrado. Dal momento che l’informazione fornita dall’esperto sarà in generale del tipo “Se le condizioni operative sono buone, allora il tempo di transizione è piccolo”, è necessario adottare un approccio fuzzy per dare un senso quantitativo a termini come “buono” e “piccolo”.

Particolare attenzione è posta alla modellizzazione degli effetti delle attività di manutenzione. In letteratura è possibile trovare numerosi approcci in cui le azioni di manutenzione hanno un impatto sull’età efficace; tuttavia, tali modelli non soddisfano il requisito – fondamentale in questo lavoro – di contare unicamente sulla conoscenza dell’esperto, che generalmente non pensa in termini di riduzione di età efficace per giudicare l’efficacia di una riparazione. Per questo motivo, introduciamo l’insieme dei possibili esiti di una azione di riparazione; ogni membro di questo insieme rappresenta lo stato di degrado nel quale il componente è lasciato al termine dell'attività di manutenzione. In questo contesto, l'esperto deve valutare la probabilità che una riparazione porti il componente in uno degli stati finali possibili.

Il **Capitolo 3** descrive il processo di elicitazione dell'esperto che, in questo lavoro, gioca un ruolo fondamentale dal momento che il nostro modello di manutenzione si basa interamente sulle sue valutazioni. Negli anni passati sono state sviluppate molte tecniche per ottenere informazioni obiettive da persone esperte. Per rendere concreto il processo di raccolta delle informazioni, definiamo una serie di interviste, ognuna incentrata su un aspetto chiave del modello: stati di degrado, fattori influenzanti, effetto dei fattori influenzanti sulla velocità di degrado, manutenzione e valutazione dei parametri per la simulazione.

Un aspetto molto importante di questo lavoro, affrontato nel **Capitolo 4**, è la valutazione quantitativa del modello tramite l'integrazione del modulo fuzzy, che fornisce la probabilità di guasto del componente (funzione univoca del suo stato di degrado), con un modulo Monte Carlo che simula i guasti casuali del componente e il cambio delle condizioni di funzionamento tramite un opportuno modello dell'ambiente. La simulazione fornisce, fra le altre cose, una stima dell'indisponibilità media del componente e i costi totali di gestione, quantità che possono essere prese in pratica come indicatori dell'efficacia di una data strategia di manutenzione.

I **Capitoli 5** e **6** descrivono un ampio e dettagliato caso studio, basato sull'ottimizzazione della manutenzione di un interruttore di media tensione che protegge un impianto per prove di corto circuito, nel quale il modello teorico è applicato in pratica al fine di verificarne le potenzialità e definire gli aspetti da sviluppare in futuro, riassunti e commentati nel **Capitolo 7**.

1 INTRODUCTION

CBM can be thought of as a dynamic Preventive Maintenance (PM) practice, in which the decision to do maintenance and the maintenance action itself are performed on the basis of the observed conditions or upon failure (i.e., corrective maintenance actions). In principle, this allows to avoid the main drawbacks of PM, which are linked to the danger of imposing actions when unnecessary, interrupting operation and possibly introducing malfunctions due to errors of maintenance operators. Obviously, CBM lies at the heart of an optimization process aimed at identifying the combination of decisions that typically maximizes the availability of the system and/or minimizes the cost of its performance. This entails that the availability and/or cost indicators are evaluated and compared in correspondence of different decisions, for finding the optimal maintenance policy.

A number of works have been propounded in the literature, which address the issue of optimizing a CBM policy in different applications (e.g., [1]-[9]). These CBM models can be regarded as particular instances of the general scheme in Figure 1, which is made up of four modules:

1) the Failure/degradation behavior model, which describes the behavior of the component with respect to the degradation mechanisms and failure modes to which it is exposed.

2) The Monitoring model, which timely provides not just the alarm at failure occurrence (as in the PM), but also the degradation state of the monitored component.

3) The Maintenance model, which describes the time distribution of the actions taken and their effects on the future degradation and failure behaviors.

4) The Decision model, in which decisions are implemented by setting the variables that define the maintenance policy (e.g., the monitoring intervals, trade-off between performing a corrective maintenance and replacement or repair, etc.). In particular, in CBM applications, maintenance actions are dynamically performed when the degradation level of the component falls beyond a given limit threshold, which is a fundamental variable defined in the Decision module.

The optimal policy is found by evaluating and comparing the availability and/or the cost corresponding to different combinations of values of the decision variables.

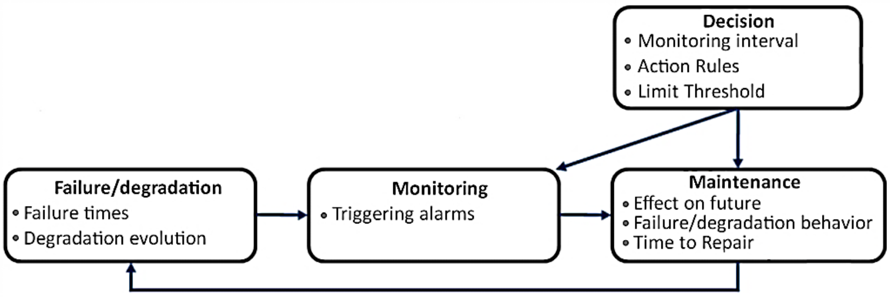

Figure 1: CBM optimization scheme

In the present work, we propose a methodology that allows to address the issue of building a CBM model in support of maintenance decision-making for a generic electrical component, in the case in which the only available information is that elicited from an expert. This situation, very common in industrial contexts, calls for the development of novel modeling solutions. In fact, the information elicited from experts is subjective, qualitative and very often in implicit form; for example, experts usually resort to linguistic variables like "high", "often" and "slow" and provide qualitative statements like 'the environmental conditions are good'. This kind of information needs to be properly interpreted, represented and propagated in an adequate model. To do this, the present work resorts to the theoretical framework of FL ([10]-[12]), due to its capability of dealing with imprecise variables and linguistic statements.

The reminder of the thesis is organized as follows. Section 2 describes in details the modeling solutions adopted to address each of the modules of the CBM optimization scheme in Figure 1. Section 3 is dedicated to the description of the method applied for eliciting information from the expert. In Section 4, an overview of the MC simulation algorithm is provided. A case study dealing with the optimization of a CBM policy of a medium-voltage circuit breaker is presented in Section 5. Results are given in Section 6 and conclusions are drawn in Section 7.

2 CBM Model

2.1 Failure/Degradation Module

Modeling the degradation process of a component is a complex problem, and can be tackled in several ways. In all generality, the effectiveness and precision of degradation models increase when these are able to capture the specificity of the component, which derives from the particular environment and operating conditions in which it works. In fact, in analogy to what happens with human beings, two similar components (i.e., of the same production lot) with the same calendar age will probably be in a different state if they have been operated differently, in different environmental conditions and/or under different stress levels. To model such situation, we resort to the concept of ‘effective age’ of a component to describe the fact that age may evolve faster or slower than chronological time in adverse or favorable working conditions, respectively. Then, the effective age can be taken as indicator of the degradation state of the component; in other words, it can be considered alike a physical variable that is representative of the health state of the component (e.g., in the same way as the crack length may be used to indicate the degradation state of a mechanical component, [8], [9]). Under this concept, the objective of degradation modeling becomes the identification of the relations between the environment and operating conditions of the component and its ‘effective age’.

The concept of effective age, also called virtual age, is not new; in fact, a number of works in the literature resort to this idea to model the aging of repairable components (e.g., [14]-[16]). In particular, the Accelerated Life Models (ALM) and Proportional Hazard Models (PHM) [14] have proven capable of effectively accounting for the environment and operating conditions of the component [14]. Both approaches start from a baseline probability distribution describing the evolution of the degradation process in normal conditions, which is modified depending on the magnitude of the influence of the component working conditions on the degradation process. These latter are characterized by means of a set of Influencing Factors (IFs), which are conditioning aspects of the component life (e.g., environment, quality, etc.). The difference between ALM and PHM lies in the way in which the dependence of the aging process on the component working conditions is modeled. However, as proposed in the literature, both PHM and ALM models require the determination of the baseline functions; this is not the case in this paper.

On the other side, the evolution of the degradation process is generally described as a stochastic transport process within a set of degradation states, which can be discrete or continuous. In this respect, both the PHM and ALM models proposed in the literature consider continuous states degradation processes. On the contrary, the practical view undertaken in this work of building the degradation model based only on the expert’s information compels to consider discrete (and finite) states reachable by the degradation process, in recognition of the fact that experts are more familiar with this way of thinking of the degradation mechanisms ([8], [9]). These two features (i.e., the lack of a stochastic model of the degradation in normal operating conditions and the necessity to consider a discrete state process), call for the development of a new degradation modeling paradigm; what we propose is discussed in the following.

Firstly, the concept of effective age (indicated by $w(t)$) is introduced as defined in the following Equation and sketched in Figure 2:

$$
\begin{aligned}
& w(t) = \Psi\big(IF(t')\big)\cdot(t - t') + w(t') \qquad for \quad t' \le t < t'' \\
& w(0) = 0
\end{aligned}
\tag{1}
$$

where $\Psi\big(IF_1(t'),...,IF_k(t')\big)$ is the 'age speed', which depends on the values of the IFs at $t = t'$; its value is 1 in nominal working conditions. The variable t' is the last time instant at which the occurrence of an event has changed the age speed of the component; finally, t'' is the next time instant at which the age speed will experience a further change (i.e., between t' and t'' no event occurs that changes the age speed).

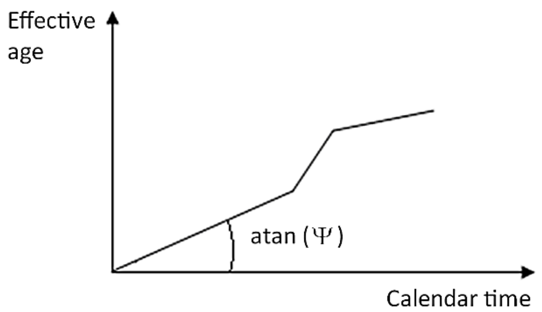

Figure 2: Example of time evolution of the effective age depending on the component working conditions through the age speed Ψ.

Once the effective age has been defined, the possible degradation states of the system need to be identified. To this aim, $n-1$ thresholds, $w_1,...,w_{n-1}$, are set on the effective age, which define the n degradation states $\mathrm{D} = 1,2,..,n$ (Figure 2). In this way, a relation between the effective age of the component and its discrete degradation states is established.

The next step is to model the dependence of the age speed Ψ on the component working conditions; these latter are characterized by means of a set of K IFs identified by the expert which define the function $\Psi\big(IF_1,...,IF_K\big)$. In this regard, direct elicitation from the expert of the value of the age speed is not feasible, since this concept remains rather intangible. Rather, what one can do is to ask the expert to assess, for a given combination of the IFs, the length of the time interval that the component takes to change its degradation state, and connect the age speed to this value. For example, if the expert knows that in certain conditions (characterized by a particular set of values $\{IF_1^*,...,IF_K^*\}$ of the IFs) the transition time between degradation states D=2 and D=3 is $t_{2\to3}^*$, then the corresponding age speed Ψ^* can be defined as :

$$\Psi* = \frac{w_2 - w_1}{t_{2\to3}^*} \quad (2)$$

where w_1 and w_2 are the thresholds on the effective age which separate *D=1* from *D=2* and *D=2* from *D=3*, respectively (Figure 3 and
Figure 4).

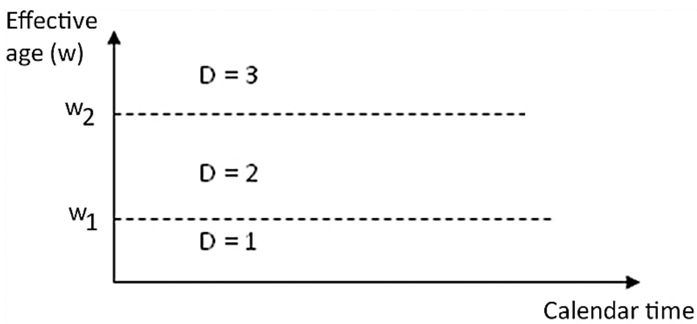

Figure 3: The degradation states are defined by setting a number of thresholds on the effective age for passing from one degradation state to another.

Obviously, one cannot expect that the expert provides precise statements such as "*If* the amplitude of the vibration is 1 mm, *then* the transition time between *D=2* and *D=3* is 1240 h". Rather, the information provided is expected to be of the form: *"If* the environment is *Mild*, *then* the transition time is *Small".* The fuzzy approach is applied in this work to deal with this type of qualitative information, since it is particularly apt to represent and propagate the imprecision associated to linguistic variables ([10]-[12]).

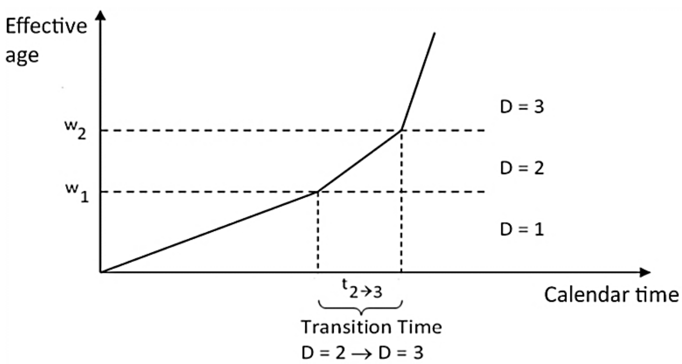

Figure 4: Depending on the IFs, the expert specifies the age speed by giving the transition time between one degradation state and the following one.

Obviously, one cannot expect that the expert provides precise statements such as "*If* the amplitude of the vibration is 1 mm, *then* the transition time between *D=2* and *D=3* is 1240 h". Rather, the information provided is expected to be of the form: *"If* the environment is *Mild*, *then* the transition time is *Small".* The fuzzy approach is applied in this work to deal with this type of qualitative information, since it is particularly apt to represent and propagate the imprecision associated to linguistic variables ([10]-[12]).

In practice, the Universe of Discourse (UoD, i.e., the set containing all the possible values) of each IF is firstly partitioned in a suitable number of fuzzy sets, e.g., "Good", "Medium" and "Heavy". The same is done for the transition time, which may be reasonably partitioned in "Small", "Medium" and "High", according to the expert's view. Then, a set of fuzzy rules such as *"If* IF_1 is Good *and ....and* IF_K is Heavy *then* the transition time is *Small"* are elicited from the expert to relate each combination of the IFs to the transition time. A degree of truth is associated to each rule, inferred on the basis of the degrees of truth of its constituents, which describe to which extent the actual values of the variables describing the current situation under analysis match the expert view. All the rules can then be logically aggregated to provide a fuzzy set that describes the implication on transition time; this is finally defuzzified ([10]-[13]) into a crisp value of the transition time, which is provided in input to Eq. (2) to find the age speed. The uncertainty represented by the fuzzy set may also be propagated to the transition time to characterize its uncertainty. Sensitivity analysis may also be performed to verify the robustness of the expert's assignment [31].

At a first glance, it may appear that the thresholds on the effective age can be quite arbitrarily fixed, since the only information to set them is the transition time between two consecutive degradation states. However, this degree of freedom is lost when the constraint that the age speed in normal operating conditions must be 1 is introduced; that is, the calendar time must coincide with the effective age if the component works in the conditions considered in the design phase.

Finally, with reference to the failure behavior of the component, in this work we assume that each degradation state has an associated shock failure time distribution, supposed exponential with a mean time between shocks dependent on the degradation state. This choice is supported by the fact that a number of works have been proposed in the literature to investigate this kind of failure behavior (e.g., [8], [9], [17], and [18] and the references therein).

Figure 5 summarizes the information flow of the proposed method, from the working conditions to the evaluation of the failure rate.

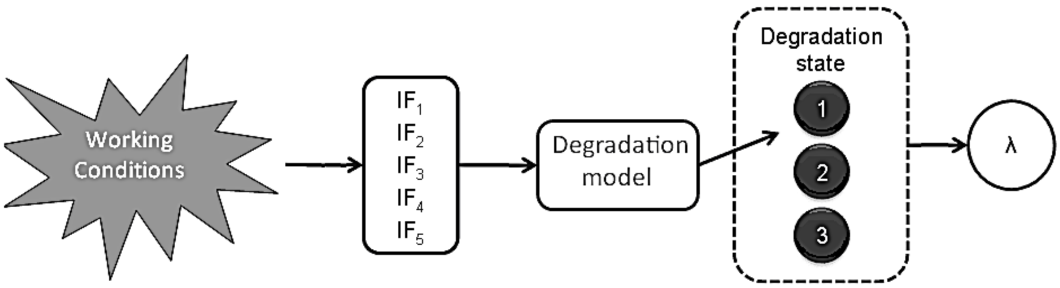

Figure 5: The degradation model receives in input the influencing factors, which represent the working conditions, and returns the current degradation state.

Notice also that in the present work we limit ourselves to the case of a single component affected by a single degradation process; however, the impact of the stochastic behavior of other components or external events on the degradation process is accounted for in the modeling, by considering the influence that the degradation states of the external components have on the IFs of the component of interest.

2.2 Maintenance module

Modeling of the effect of a maintenance action on the failure behavior of the component is one of the issues that an optimization model has to give due account to. The basic assumptions of repair efficiency are commonly referred to as minimal repair or 'As Bad As Old' (ABAO) and perfect repair or 'As Good As New' (AGAN) [13]. In the former case, the repair action leaves the equipment in the same state as it was before failure, whereas in the latter case repair is perfect and leaves the equipment as if it were new. Reality is reasonably between these two extreme cases and repairs better-than-minimal-but-not-perfect are commonly called 'Imperfect Repairs'. These have been widely investigated in the literature (e.g., [19]-[23]). In particular, models in which the maintenance actions impact on the effective age have been proposed (e.g., [14], Arithmetic Reduction of Age (ARA) models in [22], [24]). However, these models do not fulfill the requisite of relying only on the knowledge of the expert, who usually does not think in terms of age reduction to judge the efficiency of a repair. In fact, the entity of this reduction is not easy to be elicited from the expert, being the concept of effective age somewhat abstract. More likely, he/she will assess a maintenance action in terms of reduction of the degradation state. Thus, a different model needs to be developed to take into account this point of view, in which maintenance actions impact directly on the discrete degradation states $D=1,2,..,n$ of the component.

To this aim, a set $\{O_0, O_1, ..., O_n\}$ containing all the possible outcomes of a repair action is introduced; each member of this set represents the degradation state in which the component is left after the maintenance action. Namely, O_0 refers to the ABAO maintenance actions, whereas the generic outcome O_i describes the event "after the repair, the degradation state of the component is left to the beginning of *D=i*". This maintenance scheme allows to take into account also bad maintenance actions or operators errors; it suffices to consider the events $O_j . j > i$ when the component is in degradation state *D=i*.

Similarly to what is done in [25], the expert is asked to assess the likelihood that the maintenance action will actually bring the component in *D=i*, for *i=1,2,…n*. In the spirit of this work, the likelihood of a future event is partitioned in 4 fuzzy sets, namely "Very low", "Low", "High" and "Very High", as shown in Figure 6.

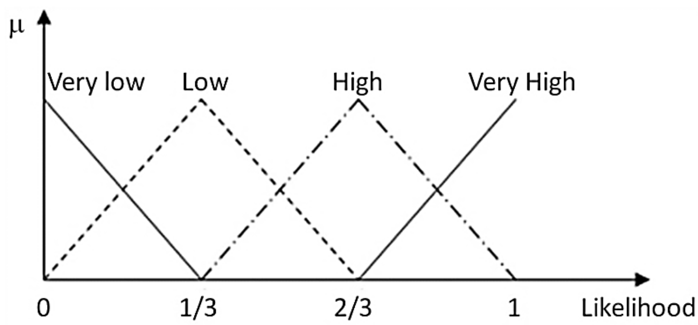

Figure 6: Fuzzification of the likelihood of the events O_i.

For example, let us consider the particular situation in which only three degradation states are defined (i.e., *D=1, 2, 3*), and the CBM policy foresees that the component overtakes repair actions if it is found in degradation state *D=2* at inspection; no maintenance action is performed in *D=1,* whereas the component is replaced if found in *D=3*. Then, the possible outcomes of a repair action are:

- O_0: the degradation is left unchanged (ABAO)

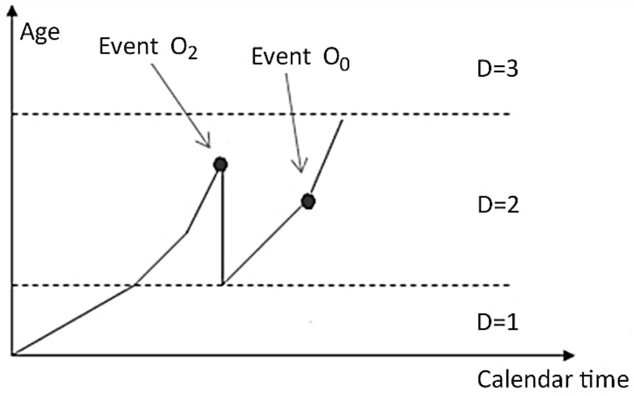

Figure 7: Possible outcomes of a repair. O_0 leaves the degradation as it is, while O_2 lowers the degradation back to the beginning of *D=2*.

- O_1: the degradation is lowered to the beginning of *D=1* (AGAN)
- O_2: the degradation is lowered to the beginning of *D=2*
- O_3: the degradation is increased to the beginning of *D=3* (bad maintenance)

For the sake of simplicity, let us assume that the expert believes that only the outcomes O_0 and O_2 are possible (Figure 7). In this case, denoting by L_0 and L_2 the likelihood that the outcome of a repair will be O_0 and O_2, respectively, the expert's assessment may be expressed by the following sentences:

- L_0 is Low
- L_2 is High

On the other side, several factors may influence the likelihood of the outcome of a repair, such as the number of repairs already done, the starting degradation state, the skill of the operator, etc. In the present illustrative example, only the number of repairs overtaken in the past is considered. In particular, two fuzzy sets "Low" and "High" for the linguistic variable "number of repairs already done" are defined (Figure 8). In this case, denoting by N the number of repairs performed on the component, the expert's assessments may be expressed by the following rules:

1) *If* N is Low *then* L_2 is High.
2) *If* N is Low *then* L_0 is Low.
3) *If* N is High *then* L_2 is Low.
4) *If* N is High *then* L_0 is High.

These rules can be implemented for any specific value of N (Figure 9); then, rules 1) and 3) are aggregated to give L_2, and rules 2) and 4) to give L_0. Two crisp values are finally obtained from the defuzzification of L_2 and L_0, which are then normalized to represent the probabilities associated to the two possible results of the maintenance action.

For example, let us consider N=5; then, the sets "Low" and "High" are activated with degrees 0.17 and 0.83, respectively (Figure 9). The implementation and the aggregation of the four rules are represented in Figure 9 and Figure 10. The operation of defuzzification by the center of area method [10]-[13] gives back L_2 = 0.33 and L_0 = 0.89. These values are transformed in probability masses by simply dividing them by 1.22 (=0.33+0.89, i.e., sum-to-one normalization [30]).

The results are reported in Table 1. The outcome of the maintenance can be sampled from the distribution derived from these probability masses.

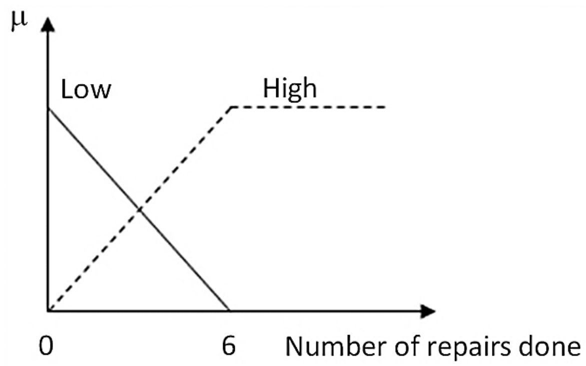

Figure 8: Fuzzification of the number of repairs already done.

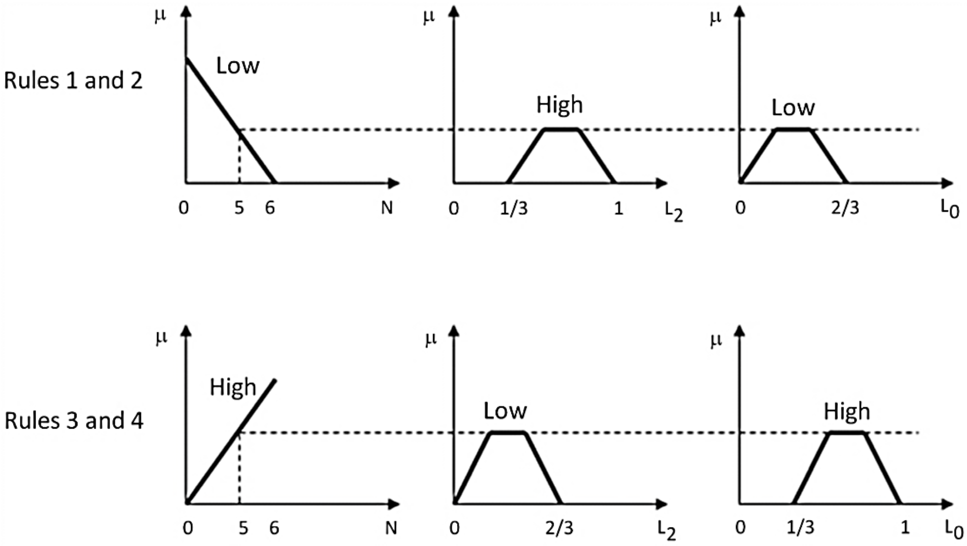

Figure 9: Rule implementation.

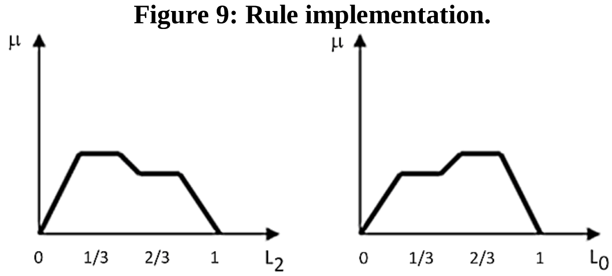

Figure 10: Rule aggregation.

Maintenance after 5 repairs	
P_2= probability to bring the effective age at the beginning of the D=2	0.27
P_0=probability of ABAO maintenance	0.73

Table 1: Probability of the outcomes of a repair action, for N=5

2.3 Monitoring

Generally speaking, monitoring can be continuous or periodic (e.g., at inspection). In the latter case, the length of the monitoring interval is an important factor, defined within the Decision module. In general, the smaller the period between two successive monitoring acquisitions, the earlier the stage at which problems are revealed. On the other hand, the monitoring interval influences the operating cost and the performance of the plant: a short interval may lead to a large cost of monitoring, whereas a long one increases the risk of failure. Further, continuous monitoring may require a costly technological solution, which allows to detect immediately the occurrence of a failure or the achievement of a limit threshold, whereas visual inspections may be cheaper but allow evaluating the system conditions only at inspections.

In this work, it is assumed that the degradation state can only be known through periodic inspections; this is a very common assumption in the models proposed in the literature (e.g., [1], [2], [6], [8]).

2.4 Decision module

The optimal policy is found by evaluating and comparing the availability and/or the cost corresponding to different combinations of values of the decision variables. In this work, only one decision variable is considered, which is the time span between two successive inspections. Other variables, typically considered in the CBM optimization models, such as the action rules, the thresholds that define the degradation levels, etc. are here considered fixed.

3 EXPERT ELICITATION

The expert's elicitation plays a crucial role in this work, since our maintenance model completely relies on his/her statements. Several techniques have been propounded in the literature to obtain an un-biased assessment from the expert (e.g., [26]-[28]). These procedures, however, are quite general, whereas we need a more precise and detailed scheme of interview. Generally speaking, the information elicited from the expert can be dived into two parts:

- information to build the model;
- assessment of the numerical quantities for the simulation.

The first part enables the analyst to conceive the general structure of the model, whereas in the second part of the elicitation the expert assesses all the relevant numerical quantities needed for the simulation, such as the failure rates associated to the degradation states. The objective is to gather all the relevant information within the framework presented in Section 2.

To be practical, we define a series of interviews, each one focused on a key aspect of the modeling framework:

1) Degradation states. The expert is asked to define a set of discrete degradation states based on symptoms, which can be quantitatively measured at the inspections.
2) Influencing factors. The expert is asked to assess if, and to what extent, the aging process of the component under study is influenced by the ambient conditions in which it operates.
3) Effect of the influencing factors on the effective age. In this interview, the link between the IFs and the age evolution is established; in particular, for each degradation state, the expert is asked to assess the transition time toward the next degradation state.
4) Maintenance. The factors which can influence the outcome of a repair are identified, and the corresponding effects on the degradation state are modeled.
5) Parameter evaluation. In this interview, the expert is asked to evaluate all the numerical quantities needed for the simulation (failure rates, maintenance costs, and so on).

Following the general ideas exposed in [26]-[28], we acknowledge the importance of informing the expert about the desired results of an interview. However, since probably the expert has never thought of the degradation process in terms of transition times, effective age and so on, we propose to divide each interview in several steps to guide the expert in the elicitation process. In particular, each interview is structured in four parts.

First, the analyst shares with the expert the objectives of the interview. Depending on the main goal of the meeting (degradation states, IFs, etc.) several objectives can be defined in order to help the expert to express his/her opinion in a structured way. Second, for each objective, the expert's opinion is elicited. Then, a third, modeling phase follows; this is the core issue of the process, in which the analyst has to find a suitable model fairly representative of the information received within the general framework. If additional details are needed, a fourth part follows, in which a second elicitation takes place to complete the model. At the end of each interview, the analyst verifies that the main goal has been achieved and that he/she and the expert share common view and terminology for the following interview.

Naturally, the expert involved in the elicitation must be preliminary trained in the main features and interpretations of Fuzzy Set theory.

4 INTEGRATION OF THE FUZZY MODEL AND MONTE CARLO SIMULATION

One relevant issue in the quantification of the model proposed is the integration of the FL model, which provides the failure rate associated to the degradation state, and the Monte Carlo (MC) module, which simulates the stochastic failure behavior of the component and the changes in the IFs caused by random external events. The output provided by the simulation module is the estimation of the component mean unavailability and total cost over the mission time, which can be taken in practice to measure the performance of the maintenance policy to be assessed.

The solution proposed in [8] to embed the fuzzy module in the MC scheme is adopted in this work. Briefly, the mission time is suitably discretized in intervals (bins) of length *Dt* and a counter is introduced for each discrete time. Then, a large number of trials or histories (i.e., random walks of the system from one configuration to another) are simulated and the instances of the health state of the system and maintenance cost are collected in every trial, for every *Dt*; finally, the collected values are opportunely averaged to provide an estimation of the desired quantities.

Figure 11 sketches the algorithm devised in this work to simulate MC trials of the component life histories; starting from the current time *t*, the simulation time is moved forward to instant *t**, which is the minimum of the following time instants:

- t_{IF}: change of one or more influencing factors, e.g., a variation of the living condition in which the component operates;
- t_m: next scheduled inspection. The interval between two inspections is the decision variable of the problem; that is, the optimization problem amounts in finding the value of this variable that minimizes unavailability and cost.
- t_d: change of degradation state. As shown in Section 2, this happens when the effective age crosses one of the thresholds w_1, $w_2, \ldots, w_{n-1}$, which define the set of degradation states;
- t_f: failure of the component.

Then, the effective age is updated at $w=w(t^*)$. If a change in degradation state has occurred, then the failure rate is correspondingly updated. The fuzzy module is run again to estimate the new age speed ([29]), which remains constant until the occurrence of the next event (Eq. 1).

Notice that this ‘event-driven’ approach, in which the MC history jumps from one event to the next, evaluating the age speed only if needed, drastically reduces the computational time, if compared to the case considered in [8]. In that work, in fact, the presence of quantities like ‘the average vibration’ compels to run the fuzzy module at each time bin, since this improves the estimation of the integral quantities (e.g., the mean).

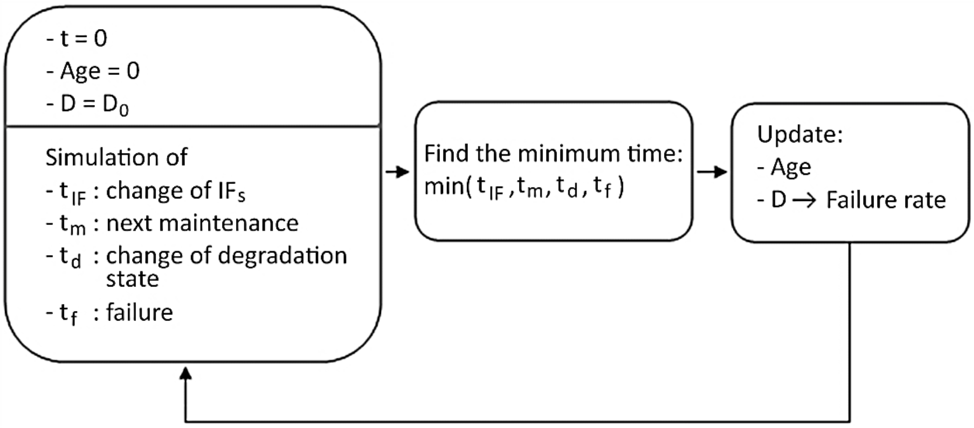

Figure 11: Integration of the FL model and MC simulation in practice.

Averaging all the relevant quantities upon the entire mission time, we can associate to each Inspection Interval (II) a set of useful indicators such as the mean unavailability of the component, its maintenance costs per year, the average probability to find it in a certain degradation state, etc. These quantities constitute the basis to assess the best maintenance policy, as shown in the case study.

5 CASE STUDY

The objective of the present case study is to optimize the maintenance of a Medium-Voltage Circuit-Breaker (MVCB), which protects a short circuit network for testing various devices such as breakers, isolators, disconnector switches, etc. (

Figure 12). The breaker is placed just after the generator, and its main function is to interrupt the short circuit current when required.

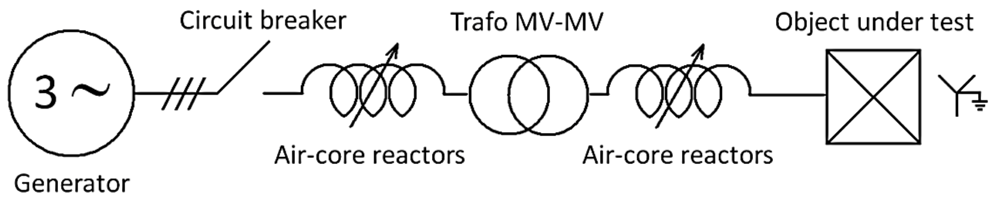

Figure 12: Overview of the Medium Voltage test network

Every time the MVCB interrupts the current, an electric arc takes place between its contacts, which consequently, wear off. If the degradation state is very bad, then the breaker loses its interrupting capability; in this case, the arc is not readily extinguished and serious damages can occur before other emergency systems clear the fault. A CBM policy is performed to prevent the wear from reaching a critical value. This is based on periodic visual inspections, in which the breaker is disassembled and the degradation state of the contacts controlled. In particular, three alternative scenarios are possible upon inspection:

- Contacts are as good as new: no action is taken and the breaker is put back together.
- Contacts are worn but the interrupting capability is still good: in this case, manual polishing is performed to smooth the surface of the contacts and decrease their electrical resistance.
- Contacts are heavily worn: the interrupting capability is at risk and they must be replaced with new ones.

Since maintenance actions, as well as failures, have a cost, the best inspection time is a non trivial issue which can be effectively addressed by means of a MC simulation over a long time span, e.g., 15 years. In the following, we walk through the steps of the expert elicitation process and the associated construction of the model. The information and the data correspond to the results of and actual elicitation process carried out with an expert.

5.1 Interview 1: Degradation states

1) General purposes of the interview

Define one or more measurable indicators of the contact degradation, and identify a set of discrete degradation states according to the value of the indicators.

2) Elicited information

The contact resistance is a good indicator of the contacts degradation. This quantity has never been measured, but it is possible to conceive a practical procedure to get this information at every inspection. Three degradation states can be defined:

- *D=1*: contacts are as good as new.
- *D=2*: contact resistance is affected by arc wear, and maintenance (contact polishing) can effectively reduce it.
- *D=3*: contacts must be replaced.

3) Modeling

The elicited information does not require a modeling session.

4) Model review

From now on, it is agreed with the expert to denote the three degradation states with “1”, “2” and “3” respectively.

5.2 Interview 2: Influencing factors

1) General purposes of the interview

Describe all the IFs which have an impact on the degradation of the contacts. For each IF, partition the universe of discourse with a suitable number of fuzzy sets.

2) Elicited information

The only variable that characterizes the environment is the interrupted current in each test: the higher the interrupted current, the heavier the wear by arc erosion. Most of the current settings range from 10 to 50 kA.

3) Modeling

The only IF is the interrupted current in each test. Since the contact degradation is more severe if the interrupted current is high, we choose to introduce three fuzzy sets for the environment, namely "Low", "Medium" and "High". For the sake of simplicity, triangular membership functions are considered [12] (Figure 13).

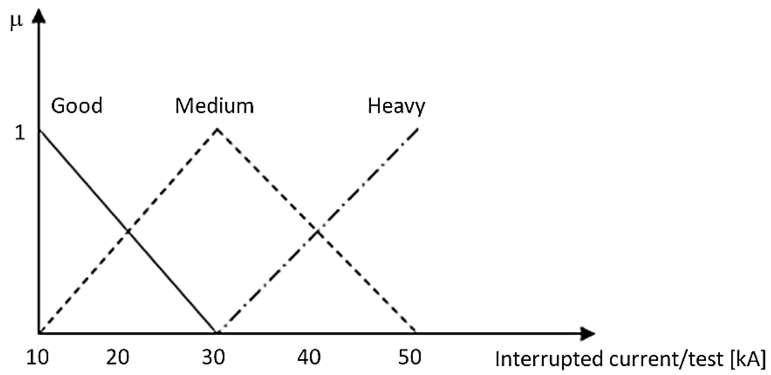

Figure 13: Fuzzy set definition for the environment, according to the interrupted current.

4) Model review

The expert agrees with the membership function showed in Figure 13, which should be kept in mind every time the words "Good", "Medium" and "Heavy" are used.

5.3 Interview 3: Effect of the influencing factors on the effective age

1) General purposes of the interview

For every combination of degradation state and working condition, estimate the time of transition toward the next degradation state. For the definition of the degradation states and working conditions, the expert refers to the results of the previous interviews.

2) Elicited information

When the contacts are new, or as good as new ($D=1$), an initial arc wear takes place. This process is rapid (two or three months) and fairly independent on the interrupting conditions. On the other hand, when the contacts are in degradation state *D=2* the value of the interrupted current heavily influences the transition time toward state *D=3*, in which the contacts ability to interrupt the current is compromised.

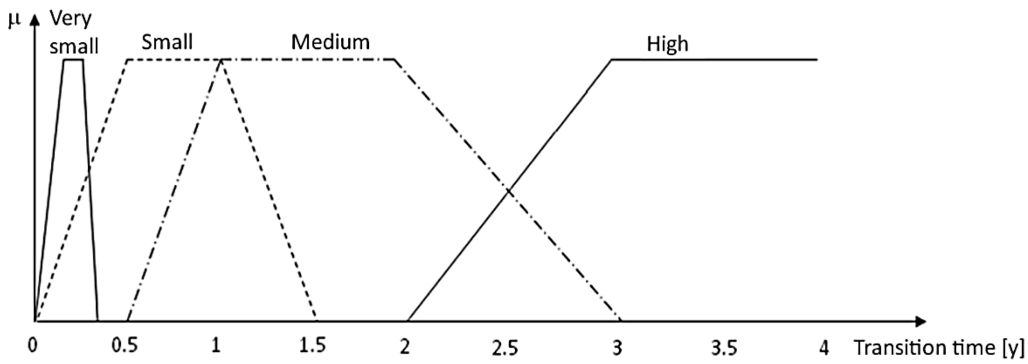

Figure 14: Fuzzification of the transition time.

Following the definition of the environment ("Good", "Medium" and "Heavy") that has been done in the first interview, the expert makes the following statements:

- If the environment is "Good", then the contacts would last three to four years.
- If the environment is "Medium", then the contacts would last one to two years.
- If the environment is "Heavy", then the contacts would last less than one year.

3) <u>Modeling</u>

The effective age *w(t)* is introduced and two thresholds (w_1 and w_2) are defined in order to identify three degradation states. These thresholds are quite arbitrary, since the relevant parameter to estimate the age speed is the transition time from one degradation state to the next one, as discussed in Section 2. This freedom is removed by requiring that the age evolves according the calendar time if the environment is in its least “stressing” condition. For example, the expert believes that the contacts working in a “Good” environment take three to four years to reach the degradation state *D=3* since the time they entered in degradation state *D=2* (that is if the interrupted current is 10 kA). This leads to set the transition time $t_{2\rightarrow 3}$ at 3.5 years, i.e., 5600 working hours. With regards to the transition time $t_{1\rightarrow 2}$, the expert estimates it to be fairly independent of the interrupted current and gives two or three months as a rough estimation. In this case, the threshold is set at 2.5 months (i.e., 400 working hours).

Now we have to associate to each (fuzzy) environmental working condition a certain age speed by means of the transition time from one degradation state to the next one. The first step is the fuzzification of the linguistic variable “transition time” according the expert’s estimations. We start with the transition time from *D=1* to *D=2*, which takes two to three months.

	Environment		
	Good	Medium	Heavy
Transition time *D=1* →*D=2*		Very small	
Transition time *D=2* →*D=3*	High	Medium	Small

Table 2: Expert’s estimation of the transition time

The fuzzy set “Very Small” is defined by means of a trapezoidal membership function, which have the interval [2,3] months as kernel (i.e., the set of values with membership degrees equal to 1) and a support that ranges from zero to five months (Figure 14). This is equivalent to add an uncertainty of two months to the information provided by the expert.

When the component is in *D=2*, the transition time depends on the environment. Also in this case, the expert’s estimations is represented by trapezoidal membership functions whose kernel extend between the given limits. The support of each function adds an uncertainty of one year for estimated values larger than two years and six months for values smaller than one year. Following this procedure, we define three fuzzy sets, “Small”, “Medium” and “High”, which by construction can be directly assigned to the environmental conditions “Heavy”, “Medium” and “Good”, respectively (Table 2).

4) Model review

No review was needed.

5.4 Interview 4: Maintenance

1) General purposes of the interview

Describe the maintenance actions and the various degradation states in which the contacts may be left after a repair. Identify all the factors which can influence the quality of a maintenance action and, for each factor, partition the universe of discourse with a suitable number of fuzzy sets. Establish a set of fuzzy rules to evaluate the likelihood that a repair will leave the contact in a certain degradation state.

2) Elicited information

An inspection of the contacts is scheduled at constant intervals. The maintenance actions depend on the degradation state in which the contacts are found at inspection:

- $D = 1$: No actions
- $D = 2$: Contact polishing to reduce the contact resistance
- $D = 3$: Contact replacement

The effect of contact polishing is to reduce the contact resistance and, thus, the degradation state. The original contact resistance cannot be restored, due to the initial arc wear. The number of previous maintenances may impact on the quality of a repair. In fact, if the number of maintenances already done is high (i.e., ≥ 6), the thickness of the contact is reduced and the contact resistance cannot be lowered anymore. Furthermore, the degradation state of the contact cannot be increased by a bad maintenance action.

3) Modeling

After a repair, two possible situations can occur:

- ineffective repair: the age is left as it was before the inspection;
- effective repair: the age is lowered to the beginning of $D=2$.

As told by the expert, the number of previous maintenances is considered as a factor that can influence the quality of a repair. The universe of discourse is partitioned by means of two fuzzy sets, “Low” and “High” respectively, as shown in Figure 15.

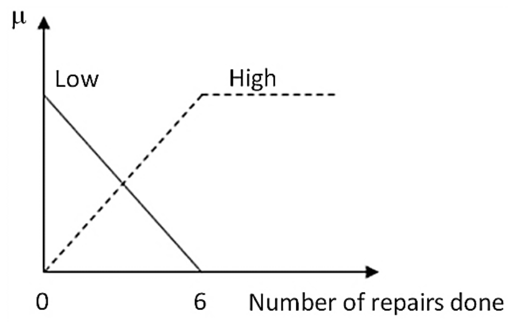

Figure 15: Fuzzification of the number of repairs already done.

Table 3 summarizes the expert’s assessment for what concerns the likelihood that a repair either will reduce the degradation state (L_2) or leave it as it is (L_0).

	N = repairs already done	
	Low	High
L_0	Low	High
L_2	High	Low

Table 3: Likelihood of the two possible outcomes of a repair, as a function of the number of repairs already done.

Denoting by N the number of repairs already done, the following rules are provided by the expert:

1) *If* N is Low *then* L_2 is High.
2) *If* N is Low *then* L_0 is Low.
3) *If* N is High *then* L_2 is Low.
4) *If* N is High *then* L_0 is High.

As explained in Section 2, the likelihood of an event is fuzzified (Figure 16).

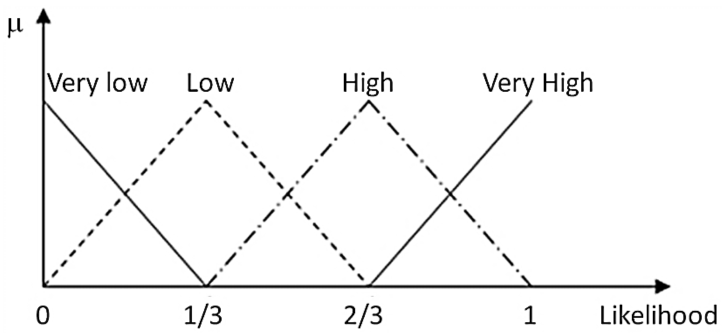

Figure 16: Fuzzification of the likelihood for an event to happen.

4) Model review

No review was needed.

5.5 Interview 5: Parameter evaluation

1) General purposes of the interview

The goal of this last interview is to assess all the numerical quantities that characterize the case study. This means:

- Quantitative modeling of each IF.
- Failure rates associated to each degradation state.
- Duration and cost of each maintenance action.

2) Elicited information

The interrupted current depends on the actual configuration of the short-circuit network which is changed, on average, once every ten (working) days. During this change, a calibration of the short-circuit current is done by adjusting the variable reactors and the transformer windings. The final value of the current depends on both the object under test at the end of the network and the client's needs. Most of the current settings range from 10 to 50 kA; lower short-circuit currents are more likely than higher ones. On average, a test every 8 hours is performed.
No statistics for the failure rate of the component is available. The only information available is:

- $D = 1$: failure is considered nearly impossible;
- $D = 2$: failure is possible, even if none has been observed in 30 years;
- $D = 3$: failure is likely within one week (5 working days).

Maintenance duration and costs, as estimated by the expert, are reported in Table 4.

Quantity	*Expert's assessment*
Inspection duration [h]	8
Inspection cost [€]	200
Maintenance duration [h]	8
Maintenance cost [€]	200
Scheduled Replacement duration [h]	16
Scheduled Replacement cost [€]	1000
Replacement duration after failure [h]	40
Replacement cost after failure [€]	5000

Table 4: Summary of the parameters used for the simulation.

3) Modeling

In this particular case, we have one IF, i.e., the interrupted current in each test. In order to model its value to run the MC simulation, the following assumptions are made:

- each test is performed regardless of the results of the previous tests. This suggests to use an exponential distribution to model the instant of time at which a test is done; following the expert, the mean time between two tests is set to 8 hours;
- each change of the network configuration, from which the value of the interrupted current depends, is performed regardless of the previous one. This suggests to use an exponential distribution to model the instant of time at which the network parameters are changed; following the expert, the mean time between two network changes is set to 80 hours (i.e., 10 working days);
- the probability distribution function of the value of the short-circuit current (which is a consequence of the network configuration) is assumed to be triangular (Figure 17). This seems to be the simplest way to take into account that lower short-circuit currents are more likely than higher ones.

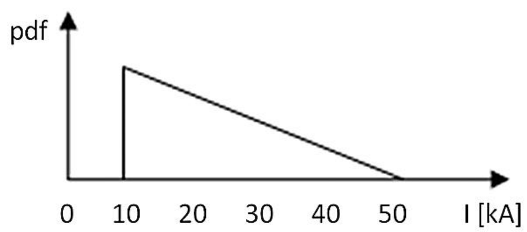

Figure 17: Probability density function for the short-circuit current of a test.

With regards to the estimation of the failure rates, the expert statements lead to the following values:

- *D = 1*: $\lambda = 1.0e\text{-}6\ h^{-1}$ or larger;
- *D = 2*: a conservative estimation for the mean time to failure could be 30 years. Given that one year has 10 working months, 1 month has 20 working days and 1 day has 8 working hours, we find $\lambda = 1/(30*10*20*8) = 2.1e\text{-}5\ h^{-1}$;
- *D = 3*: $\lambda = 1/(5*8) = 2.5e\text{-}2\ h^{-1}$.

4) Model review

No review was needed.

6 RESULTS AND DISCUSSION

Figures 18 to 23 show the results of the CBM model built according to the procedure described above. These results are provided by MC simulation of 5000 trials with the mission time of the component set to 15 working years. The computational time for each simulation is about 15 seconds (Intel Pentium, 1.73 GHz).

The average unavailability reaches a minimum in correspondence of II=3500 working hours, i.e., 21.8 working months. As shown below, this minimum corresponds to the situation in which an inspection takes place just before the component enters the third degradation state. A deeper insight can be gained by decomposing the average unavailability in its three components:

1) Average unavailability due to repairs (Figure 19) which obviously decreases for higher IIs. In particular, we see that it is totally useless to schedule periodic maintenances with periods greater than 4500 h, since the breaker always fails before the inspection is done.
2) Average unavailability due to preventive replacements (Figure 20). The peak at 4000 h reveals that the component is often found in the third degradation state, for this II; thus, it is preventively replaced to avoid failures.
3) Average unavailability due to replacements after failure (Figure 21). As expected, this quantity increases for higher IIs: the more rare the inspections, the smaller the probability to fix the component before failure.

Figure 22 shows the total costs; also in this case, the minimum is reached at 3500 hours. The relative flatness of both the total unavailability and total costs between

1500 h and 4000 h gives a certain freedom to choose the II: other criteria not included in this analysis can be taken into account if these lead to a value of the II close to 3500 h.

Figure 23 shows the instantaneous unavailability in every time bin of length 80 h (10 working days), for an II of 3500 h; in this figure, the temporarily loss of availability of the breaker during each scheduled inspection is highlighted.

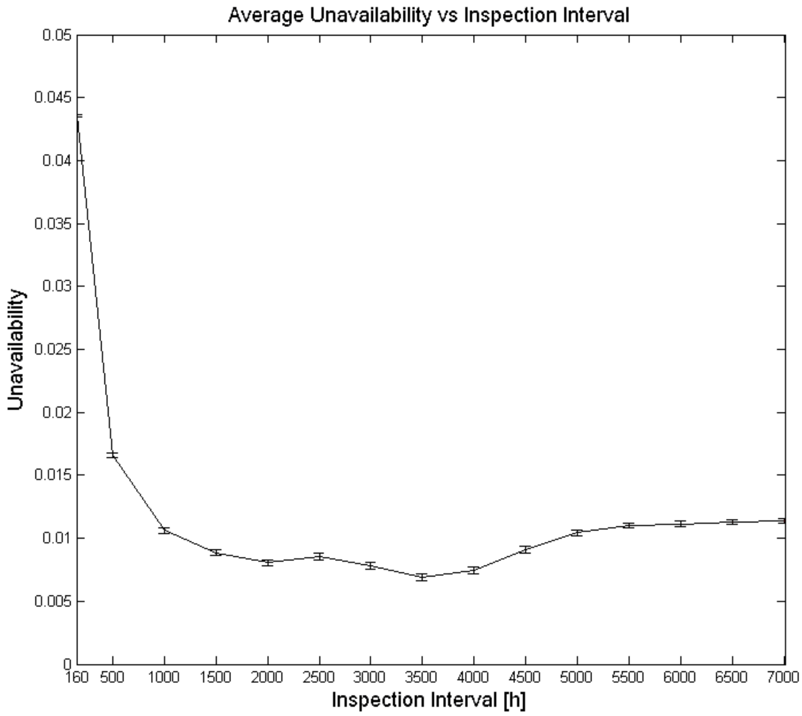

Figure 18: Average unavailability as a function of the II.

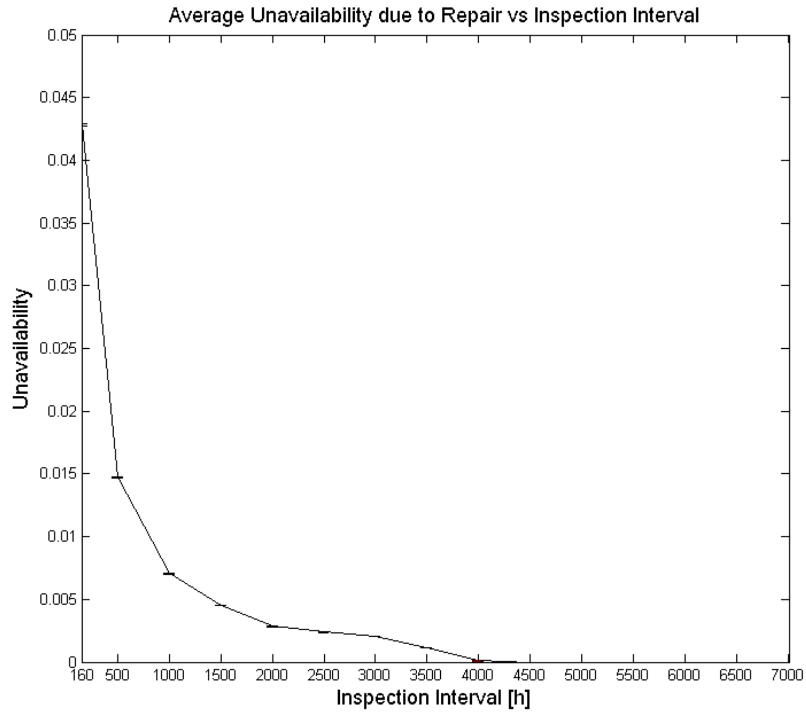

Figure 19: Average unavailability due to repairs as a function of the II.

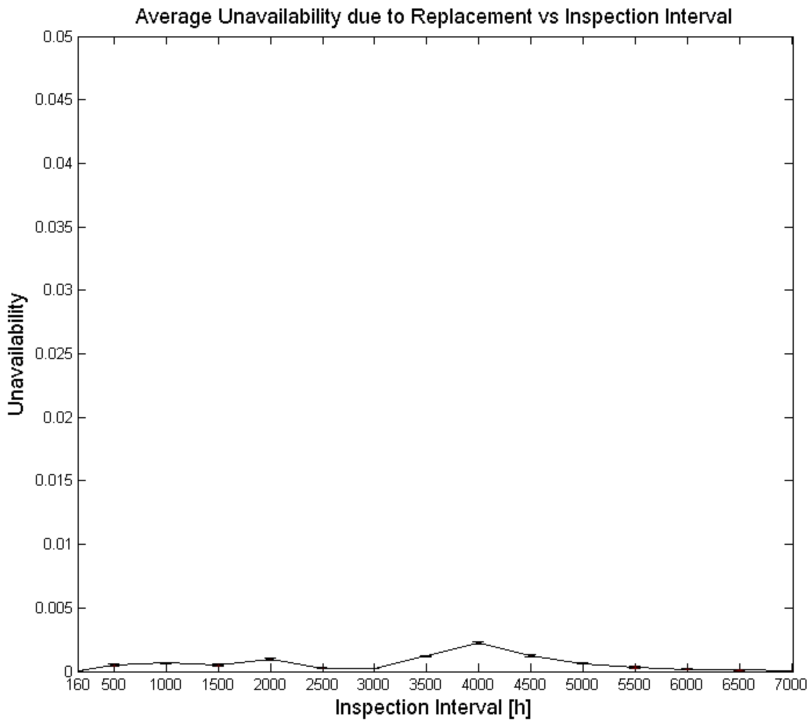

Figure 20: Average unavailability due to preventive replacements as a function of the II.

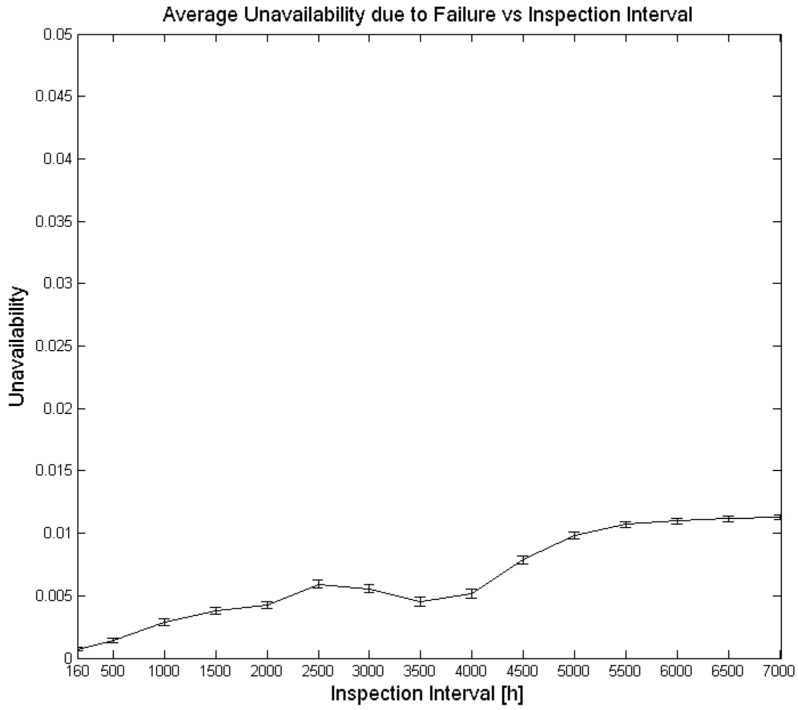

Figure 21: Average unavailability due to replacements after failure as a function of the II.

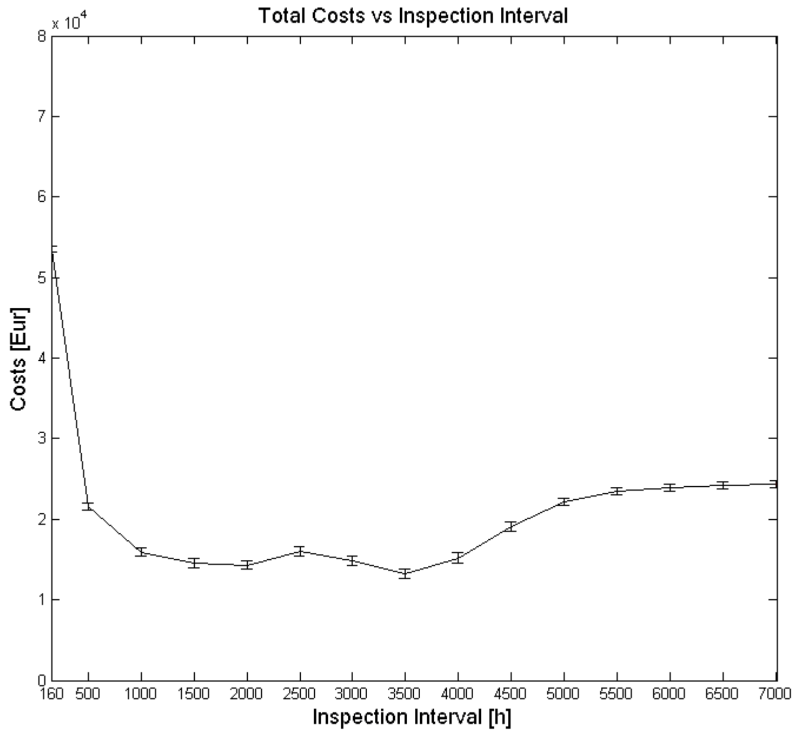

Figure 22: Average cost as a function of the II.

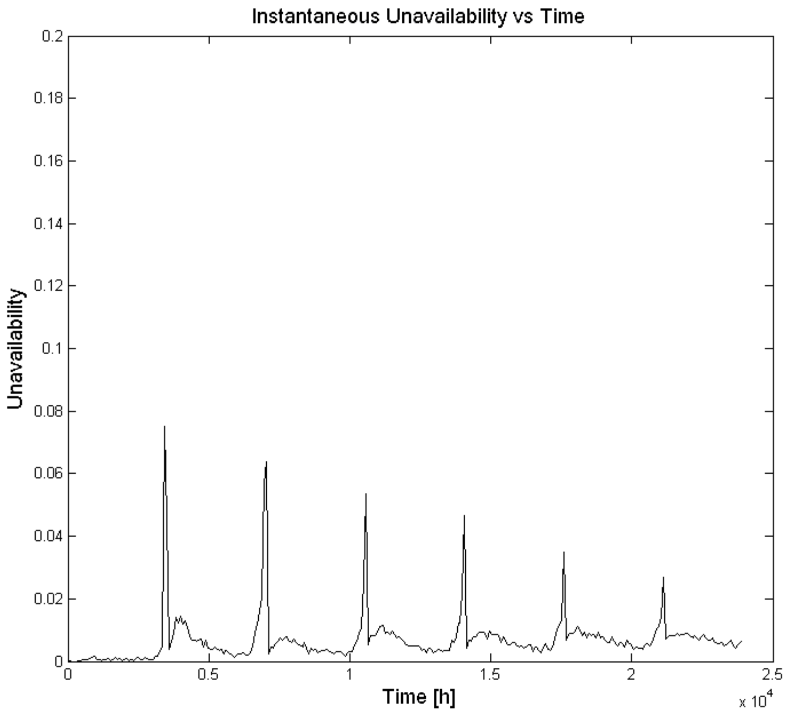

Figure 23: Instantaneous unavailability with an II= 3500 h (best choice).

6.1 INFLUENCE OF THE WORKING ENVIRONMENT

The influence of the working environment is investigated by means of three different kinds of simulations. Namely, in each simulation the interrupted current is forced to assume a fixed value, corresponding to a “Good”, “Medium” and “Heavy” environments, respectively. From Figure 13, it clearly appears that the values most representative of these sets are 10, 30 and 50 kA, respectively. Figure 24 shows the result of this comparison. The absolute minima of the average unavailability in case of “Good”, “Medium” and “Heavy” environment are located at 1500, 3000 and 5500 hours, respectively. These values correspond to a situation in which the periodic controls are performed just before the component enters the third degradation state and fails, thus saving the larger unavailability associated to the corrective actions.

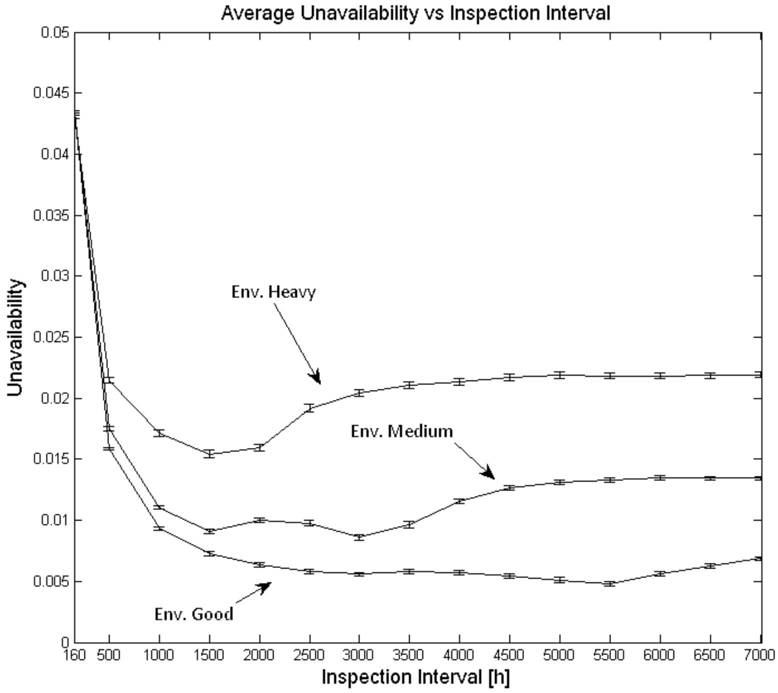

Figure 24: Comparison between the average unavailability in case of working environment always "Good", "Medium" and "Heavy", respectively.

For example, in case of "Heavy" environment, the transition time $t_{1\to2}$ from $D=1$ to $D=2$ is always 400 h, regardless of the ambient conditions (in agreement with the expert's assessment of the third interview of Section 5). On the contrary, the transition time $t_{2\to3}$ is found running the fuzzy module when the environment is "Heavy", which corresponds to a "Small" transition time (Figure 14). The defuzzification of this set leads to the crisp value of 1200 hours, i.e., 7.5 months. This meets the assessment of the expert, who says that in "Heavy" conditions the equipment takes less than one year to reach the third degradation state.

To sum up, the contacts take 1600 hours to reach the third degradation state starting from the "as good as new" condition in a "Heavy" environment. After this time, the breaker will shortly fail to interrupt the current and need to be replaced. This explains why the result of the simulations finds the minimum of the mean unavailability at1500 hours: it is just before the component experiences a failure. This also explains the sharp minimum of the mean unavailability around II=1500 h, which is observed if the environment is always “Heavy”. In this case, repairs play an important role in saving the unavailability due to the corrective maintenance. On the other side, frequent maintenance actions are ineffective, and lead to raising the mean unavailability of the component.

Finally, notice that if the environment is always “Good”, then the minimum of the unavailability is achieved in correspondence of a larger II. This is due to the fact that the entrance in *D=3* occurs later if the component works in nominal conditions.

6.2 INFLUENCE OF REPAIR EFFICIENCY

One of the most important features of our maintenance model is that the outcome of a repair is affected by an aleatory uncertainty. In particular, the two possibilities for the result of a repair when the contacts are in *D=2* are (see Section 5):

1) ineffective repair: the age is left as it was before the inspection;
2) effective repair: the age is lowered to the beginning of *D=2*.

Figure 25 shows a comparison between the average unavailability corresponding to the two extreme cases of repairs always effective and always ineffective, respectively. In both cases, the contacts are replaced if they are found in *D=3* at the inspection. This explains the increase of the unavailability for higher IIs even when ABAO maintenance actions are performed. The two curves coincides for IIs larger than 4000 h. Indeed, in this situation the contacts are always found in *D=3* and preventive replacements are done instead of repairs. However, between 500 and 4000 hours we can appreciate the importance of performing effective maintenance, with a significant reduction of the total, average unavailability.

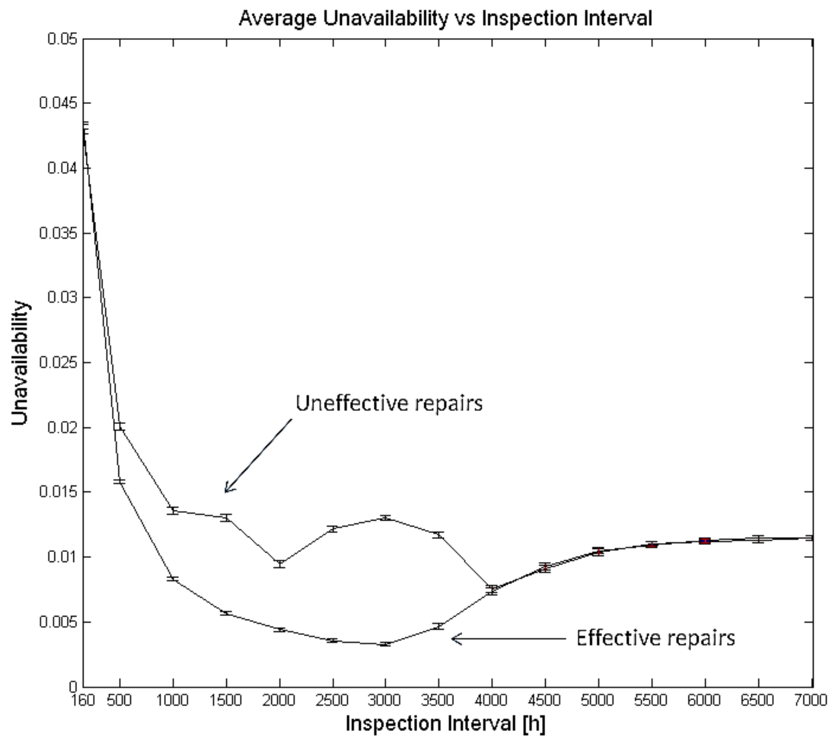

Figure 25: Comparison between the average unavailability in case of repairs always effective and ineffective, respectively.

6.3 FUZZY VS CRISP

In this work, we have resorted to FL to elicit the expert's knowledge about:

- the influence of the working conditions on the component degradation process;
- the effects of the maintenance actions on this process.

These two issues, described respectively in Sections 2.1 and 2.2, are implemented in two fuzzy modules which can be depicted as in

Figure 26; for example, in the case of the maintenance module, the (crisp) input is the number of maintenance actions already done in the past, the (crisp) output is the probability that a repair will result in an effective reduction of the component age, whereas the expert's knowledge is summarized in four linguistic rules (see §2.2) which are the basis of the fuzzy inference process.

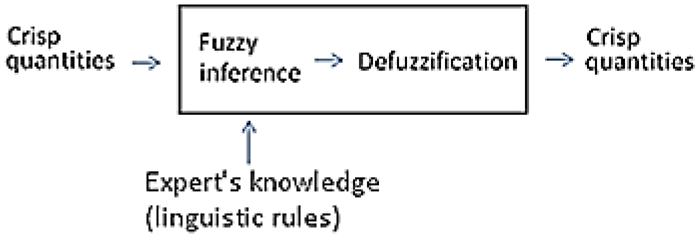

Figure 26: Graphic representation of a generic fuzzy module.

It is interesting to compare the results presented in the previous Sections with that of a new model in which all fuzzy modules are “reduced” to crisp modules (see Figure 27); then, the fuzzy rules are substituted by standard logical expressions, which summarize the expert’s assessments. For example, let us first consider the degradation module; it receives in input the value of the interrupted current I in each test, which is a measure of the quality of the working conditions, and returns the age speed. As explained in §2.1, we first calculate the degree of activation of the three fuzzy sets describing the environment (Good, Medium and Heavy) according to the value of the interrupted current; then, on the basis of the linguistic rules given by the expert, we infer a fuzzy set describing the transition time from one degradation state to the following one.

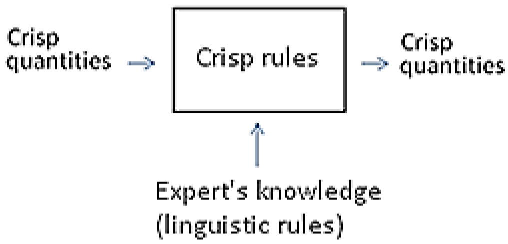

Figure 27: Reduction of a fuzzy module.

Defuzzification follows, and the crisp value of the transition time is used to calculate the age speed. In order to "reduce" this fuzzy module, we first transform the three fuzzy sets Good, Medium and Heavy, partitioning the universe of discourse (see §2.1), in three *crisp*, ordinary sets with the same name. One of the possible choices is to define:

1) Good: $I < 20$ kA
2) Medium: $20 \leq I \leq 40$ (3)
3) Heavy: $I > 40$ kA

Now we have to assign to each set a crisp value for the transition time from *D=2* to *D=3,* according to the linguistic rules given by the expert, which are:

1) **If** the environment is Good **then** the transition time is Large
2) **If** the environment is Medium **then** the transition time is Medium (4)
3) **If** the environment is Heavy **then** the transition time is Short

To do this, the “Large”, “Medium” and “Small” fuzzy sets shown in Figure 14, i.e., the consequents of the fuzzy rules, are defuzzified; then, the crisp version of the fuzzy rules (4) becomes:

1) **If** $I < 20$ kA **then** the transition time is 5172 h (3.2 y)
2) **If** $20 \leq I \leq 40$ **then** the transition time is 2628 h (1.6 y) (5)
3) **If** $I > 40$ kA **then** the transition time is 1200 h (0.75 y)

Once the transition time $t_{2\to3}$ is known, the age speed is:

$$\Psi = \frac{w_2 - w_1}{t_{2\to3}} \tag{6}$$

being w_1 and w_2 the two thresholds on the age, as explained in §2.1. Figure 28 shows the results of the inference process, which relates the value of the interrupted current (i.e., the environment) to the value of the age speed, by fuzzy rules and crisp rules, respectively. We see that the age speed inferred by crisp rules (dashed line) overestimates the age speed deduced by fuzzy rules (solid line) for currents higher than 40 kA. This is due to our choice to include all the currents above 40 kA in the “Heavy” environment (see Equation 3) and assign to this set an age speed that comes from the defuzzification of the “Small” transition time, i.e., the shortest one in this degradation state. When using fuzzy inference, instead, the age speed reaches its maximum only in correspondence of an interrupted current of 50 kA, which is only value that fully activates the (fuzzy) “Heavy” environment (Figure 13).

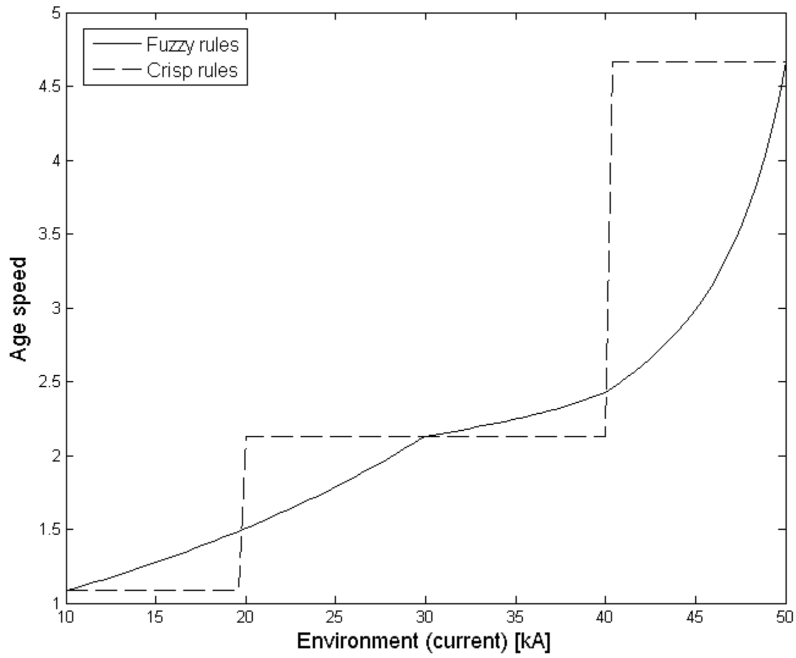

Figure 28: Comparison between the inferred age speed using fuzzy rules (solid line) and crisp rules (dashed line).

The same procedure described above is applied to “reduce” the maintenance module (see §2.2), which receives in input the number of maintenance actions already done, N, and gives back the probability of a reduction of the effective age to the beginning of $D=2$, P_2, and the probability of performing an ineffective action which leaves the age as it is, P_0. In §2.2 the universe of discourse of N was partitioned in the two fuzzy sets “Low” and “High”. Now we define two ordinary, crisp sets as:

- Low: $N \leq 3$
- High: $N > 3$

As before, we apply the fuzzy inference rules given by the expert first when N is completely Low, and then when N is completely High. After defuzzification we find the following crisp rules:

1) **If** $N \leq 3$ **then** $P_0 = 0.33$, $P_2 = 0.66$
2) **If** $N > 3$ **then** $P_0 = 0.66$, $P_2 = 0.33$

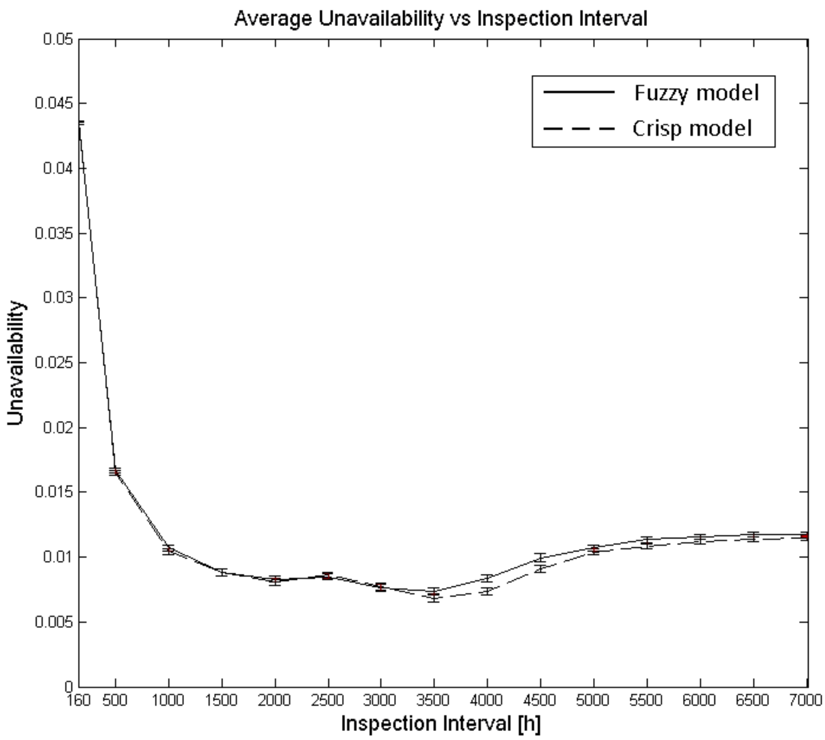

Figure 29: Comparison between fuzzy (solid line) and crisp inference (dashed line).

Figure 29 shows the results of the MC simulation in which the two fuzzy modules have been reduced (dashed line). We see that the unavailability is in good agreement with that obtained with the fuzzy model described in the previous sections (solid line). This proves that the two models are equivalent on average, after a high number of histories simulated by the MC module. But we can do a step further and see that this correspondence still holds at a single-trial level. To do this, we have to compare the time evolution of the age speed during *one single history* using fuzzy and crisp inference, respectively.

For this comparison to be meaningful, however, the time history of the working conditions has to be the same. For this reason, we first run a simulation and “save” the resulting evolution of the environment (which, in our case, includes the network changes, the value of the interrupted current and the instant of time of a short-circuit test), the random component failure and the results of the maintenance actions. Then, we use this particular history to perform two MC simulations using fuzzy and crisp inference, respectively.

Figure 30 shows the results of such a comparison for what concerns the age speed. The component is in the first degradation state up to 500 h, where the age speed does not depend on the working conditions and the effective age evolves as the calendar time.

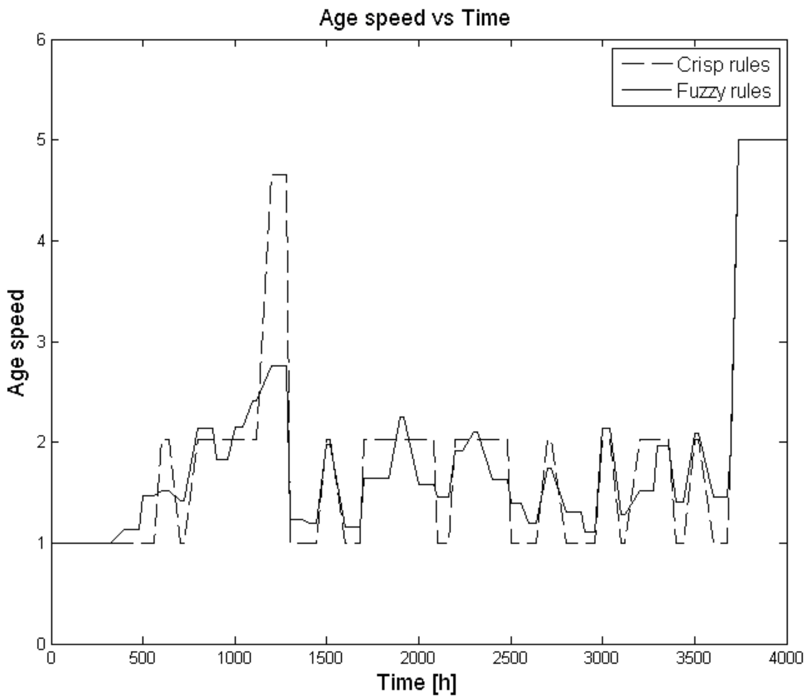

Figure 30: Time history of the age speed during a single trial using “crisp” inference (dashed line) and fuzzy inference (solid line).

Then, from 500 h to 3700 h the component enters the second degradation state. At 1200 h we clearly see that the age speed inferred with crisp rules overestimates the age speed inferred with fuzzy rules, as discussed above. In fact, at this time the interrupted current happens to be 45 kA, which is not a fully “Heavy” environment according to the fuzzy approach (Figure 13) but it surely is according to the crisp one (Equation 3).

However, apart from this hundred of hours, it is clear that the crisp rules approximate the fuzzy ones quite well, provided that the working conditions are the same. At 3700 h the component enters the third degradation state, where, once again, the age speed does not depend on the environment and it is set to an arbitrary value of 5. At this point, the component will fail in a few hundreds of hours and will be replaced.

6.4 TIME VS NUMBER OF OPERATIONS

In this work we have considered the variable “time” to be the main underling factor of the degradation process. However, in many industrial applications it could be more appropriate to consider other variables such as the number of operations. This is the case, for example, of electrical contactors which have to endure a high number of open/close movements. Our model is easily adapted to this kind of situations, in which the continuous variable “time” in the MC simulation is replaced by the discrete variable “number of operations”. The main consequences of such a change are:

- Periodic inspections are scheduled after a fixed number of operations instead of constant time intervals.
- The component failure rate λ becomes the probability of failure at the next operation given that the component has been working so far. Denoting with p such a probability, it is easily seen that $1/p$ is the average number of operations until a working component fails.
- The occurrence of a failure is randomly sampled from a discrete distribution $f(n)$ which gives the probability that a working component will fail at the *n-th* operation, i.e., $f(n) = p(1-p)^n$.

7 CONCLUSION AND OPEN ISSUES

In this work we have proposed a complete method to: i) elicit from one expert, in a structured way, his/her knowledge about a given component; ii) use this information to build a maintenance model; iii) simulate the model to optimize the maintenance policy of a generic electrical component. One key feature of the model lies in its capability of accounting for the actual working conditions in which the component operates. This capability comes from the fact that the aging process of the component is deduced step-by-step by means of five interviews, in which the expert's experience on components working in different conditions with different aging behaviors is captured and made exploitable for modeling. In this respect, we have largely resorted to FL to represent and propagate the imprecision associated to the qualitative sentences of the expert.

Future research should focus on four main issues. First, a sensitivity analysis to identify, among the great amount of data elicited from the expert, those most influential on the model outcome and, in the end, the choice of the best maintenance policy. For example, from Figure 24 it is evident that in the present case study, the environment heavily influences the mean unavailability of the component. Therefore, this part of the elicitation process should be analyzed step-by-step to identify the 'hot spots'.

Second, the failure rates could be deduced within a FL framework to take into account their intrinsic uncertainty. In our approach, in fact, the expert is asked give the failure probability of the component for each degradation state, but it is unlikely that he/she will be able to assess this quantity based on what happened in the past, i.e., relying on statistically significant data. Even if such data existed, it would impossible to tell in which degradation state the component was when failure occurred. Rather, the expert will use his/her experience to provide a reasonable estimation, as the case study on the MVCB has shown. In that case, the failure rate λ of the component in "heavy" working conditions (see Figure 13), for example, is said to be “likely within one week” (40 working hours). In this work we neglected the uncertainty which clearly stems from this statement and considered λ as a crisp number, i.e., $\lambda = 1/40 = 2.5e\text{-}2\ h^{-1}$. To improve the modeling, λ could be regarded as a fuzzy number, rather than a crisp quantity, possibly resorting to the likelihood fuzzification of a future event proposed in the maintenance module (see Figure 6).

Third, since the case study considered is made up of a single component affected by only one degradation process, the potential of the framework needs to be tested on a multi-component and multi-degradation processes system.

Finally, further research should be done to avoid crisp thresholds on the effective age. As explained in §2.1, they have been introduced to define the discrete set of degradation states $D=1, 2, \ldots, n$ and avoid to use directly the effective age as a continuous indicator of the component degradation. This, in fact, would have been in contrast with the human way of reasoning of the expert, who tends to perceive the component aging as a discrete process. However, crisp thresholds could be inadequate to represent the expert’s assessments. To understand why, recall that the two basic functions of the fuzzy module are:

- provide the value of the age speed according to the living conditions.
- forecast the instant of time at which the age will cross the next threshold and the component will increase its degradation state.

The only available input of the module is the expert’s knowledge, which is elicited thanks to statements like “If the environment is Good, then the transition time between D=1 and D=2 is Large”. The main problem, now, is to understand how to represent the uncertainty of such an assessment, i.e., how to “transfer” the ambiguity of the expert’s knowledge on the two quantities used by the fuzzy module, namely the age speed and the thresholds on the effective age. A first step is to recognize that when the expert says “If the environment is Good, then the transition time between D=1 and D=2 is Large”, he is actually making a synthesis of two uncertainties:

- the “Good” environment may be not exactly defined, or, if it is, the expert may be unable to characterize it by means of crisp, certain values;
- the degradation states D=1 and D=2 may not have a precise definition.

Therefore, the uncertainty contained in the “Large” transition time is made up of two contributions, and the model should be able to split and treat them separately. The first possibility, which is the one used in this work, is to assume a crisp separation between the various degradation states. Then, the uncertainty expressed by the expert about the transition time is completely “transferred” to the working conditions (see Figure 31), that is, to the age speed. This is equivalent to say that the uncertainty of expert’s assessment about the transition time is entirely due to the uncertainty of the working conditions. Note that in the current approach, the age speed is subsequently defuzzified for sake of computational ease.

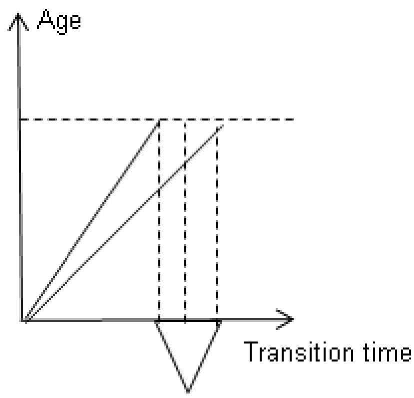

Figure 31: If the thresholds which separate the degradation states are crisp, the uncertainty of the transition time is completely modeled with an uncertainty of the working conditions.

The second possibility is to assume that the working conditions are exactly defined. Then, the uncertainty of the expert’s assessment about the transition time is modeled with a corresponding uncertainty on the threshold between the degradation states (see Figure 32).

A third possibility is to "distribute" the uncertainty of the transition time among the threshold on the effective age *and* the working conditions (see Figure 33). These three approaches have both advantages and drawbacks, and certainly do not exhaust all the possible ways to treat the expert's assessments; for this reason, this issue should be carefully investigated in the future.

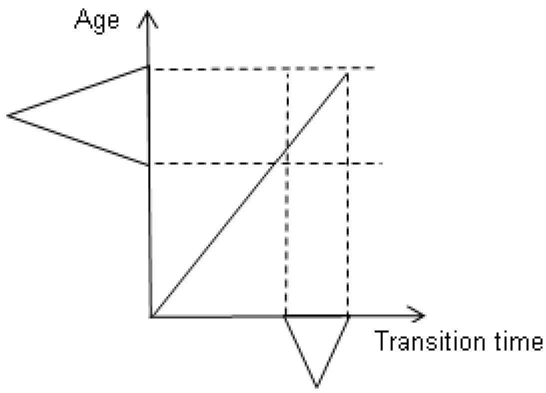

Figure 32. If the working conditions are known exactly, the uncertainty of the transition time is completely modeled with an uncertainty of threshold on the effective age that separates the degradation states.

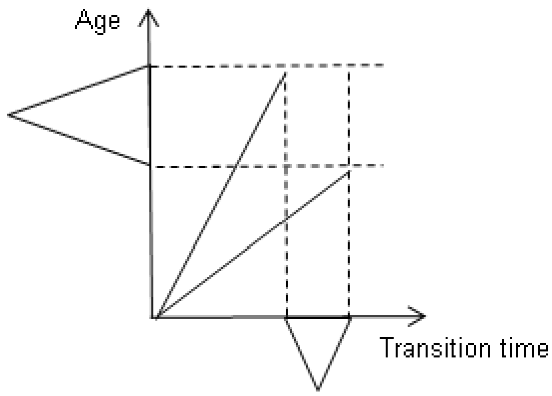

Figure 33. The uncertainty of the transition time is modeled with an uncertainty of threshold between the degradation states *and* an uncertainty of the working conditions.

ACRONYMS AND SYMBOLS

Acronyms	
ABAO	As Bad As Old
AGAN	As Good As New
ALM	Accelerated Life Models
CBM	Condition Based Maintenance
IF	Influencing Factor
II	Inspection Interval
MC	Monte Carlo
MVCB	Medium Voltage Circuit Breaker
PHM	Proportional Hazard Models
PM	Preventive Maintenance
UoD	Universe of Discourse

Symbols	
D_i	i-th degradation state
λ	Failure rate
L_i	likelihood of the event O_i
N	Number of maintenances already done
O_i	i-th outcome of a repair action
p	Failure probability at the next operation
t	Calendar time
$t_{j \to j+1}$	Transition time between the degradation states D=j and D=j+1

w	Effective age
w_k	Threshold on the effective age between the degradation states $D=j$ and $D=j+1$
Ψ	Age speed

REFERENCES

[1] Grall, A., Bérenguer, C., Chu, C. 1998. Optimal dynamic inspection/replacement planning in condition - based maintenance. Safety and Reliability. S. Lydersen, G. Hansen & H. Sandtorv (eds), Trondheim, Balkema, pp: 381-388.

[2] Bérenguer, C., Grall, A., Castanier, B. 2000. Simulation and evaluation of condition-based maintenance policies for multi-component continuous-state deteriorating systems. Foresight and Precaution, Cottam, Harvey, Pape & Tait (eds), Rotterdam, Balkema, pp. 275-282.

[3] Wang, W. 2008. Condition-based Maintenance Modelling. Complex System Maintenance Handbook. Pp.11-132. K.A.H. Kobbacy & D.N.P. Murthy Eds. Springer Series in Reliability Engineering.

[4] Lam, C. and Yeh, R. 1994. Optimal maintenance policies for deteriorating systems under various maintenance strategies. IEEE Transactions on Reliability. Vol. 43, pp. 423-430.

[5] Kopnov, V.A. 1999. Optimal degradation process control by two-level policies. Reliability Engineering and System Safety. Vol. 66, pp. 1-11.

[6] Ponchet, A., Fouladirad, M., Grall, A. 2009. Imperfect condition-based maintenance on a finite time span for a gradually deteriorating system. Safety Reliability and Risk analysis: Theory, Methods and Applications. Proceedings of Esrel 2009, Praha (Czech Republic). Briš et al. Eds.

[7] Marseguerra, M., Zio, E., Podofillini, L. 2002. Condition-based maintenance optimization by means of genetic algorithms and Monte Carlo simulation. Reliability Engineering and System Safety. Vol. 77, pp. 151-165.

[8] Baraldi, P., Compare, M., Despujols, A., Zio. E. 2011. Modeling the effects of maintenance on the degradation of a water-feeding turbo-pump of nuclear power plant. Accepted on Journal of Risk and Reliability.

[9] Baraldi, P., Zio, E., Compare, M., Rossetti, G. and Despujols A. 2009. A novel approach to model the degradation of components in electrical production plants. Proceedings of the European Safety and Reliability Conference ESREL 2009, Praha, Czech Republic.

[10] Bojadziev, G. and Bojadziev, M. 1996. Fuzzy Sets, Fuzzy Logic, Applications. Advances in Fuzzy Systems – Applications and Theory. Vol. 5.

[11] Dubois, D. and Prade, H. 1996. What are fuzzy rules and how to use them. Fuzzy Sets and Systems. Vol. 84, Issue 2, pp. 169-185.

[12] Dubois, D. and Prade, H. 1992. Gradual Inference rules in approximate reasoning. Information Science. Vol. 61, pp. 103-122.

[13] Van Leekwijck, W., Kerre, E. 1999. Defuzzification: criteria and classification. Fuzzy Sets and Systems. Vol. 108, pp. 159-178.

[14] Martorell, S., Sanchez, A., Serradell, V. 1999. Age-dependent reliability model considering effects of maintenance and working conditions. Reliability Engineering and System Safety. Vol. 64, pp. 19-31.

[15] Vesely, W.E.. 1987. Risk evaluation of aging phenomena: The linear aging reliability model and its extensions. NUREG/CR-4769, US NRC. Washington.

[16] Samanta, P.K., Vesely, W.e., Hsu, F., Subudhi, M. 1991. Degradation modeling with application to aging maintenance effectiveness evaluations. NUREG/CR-5612, US NRC. Washington.

[17] Zille, V., Bérenguer, C., Grall, A., Despujols, A., & Lonchampt, J. 2007. Modelling and performance assessment of complex maintenance programs for

multi-component systems. Proceedings of the 32nd ESReDA seminar and 1st ESReDA-ESRA seminar, 6-9 may 2007, Alghero, Italy. pp. 127-140.

[18] Singpurwalla, N.D. 1995. Survival in Dynamic Environment. Satistical Science. Vol. 10, No. 1, pp. 86-103.

[19] Wang, H. 2002. A survey of maintenance policies of deteriorating systems. European Journal of Operational Research. Vol. 139, pp. 469-489.

[20] Wang, H. and Pham, H. 1999. Some maintenance models and availability with imperfect maintenance models. Annals of Operations Research. Vol. 91, pp. 305-318.

[21] Barlow, R.E. and Proschan, F. 1965. Mathematical Theory of Reliability, John Wiley, New York.

[22] Doyent, L. and Gaudoin O. 2004. Classes of imperfect repair models based on reduction of failure intensity or virtual age. Reliability Engineering and System Safety. Vol. 84, pp. 45-56.

[23] Brown, M. and Proschan, F. 1983. Imperfect Repair. Journal of Applied Probability. Vol. 20, pp. 851-859.

[24] Clavareau, J. and Labeau, P.E. 2009. A model of imperfect preventive maintenance efficiency based on the concepts of elasticity and effective age. Safety Reliability and Risk analysis: Theory, Methods and Applications. Proceedings of Esrel 2009, Praha (Czech Republic). Briš et al. Eds.

[25] Aja Fernàndez, S. Un nuevo marco matricial para la implementaciòn de inferencia borrosa aplicados al procesado de informaciòn no numèrica, Ph.D. Thesis.

[26] Bolado, R., Badea, A., Poole, M. 2009. PAMINA - Performance Assessment Methodologies in Application to Guide the Development of the Safety Case, N. M2.2.A.3.

[27] Bolado, R., Duro, L., Grivé, M., Cormenzana, J. L., Cunado, M. A., Bennett, D. G. 2010. Estimation via expert judgement of the solubility limit distributions for five elements in the near field in a nuclear waste repository in granite. Reliability, Risk and Safety, 2010 Taylor & Francis Group, London.

[28] Sandri, S.A., Dubois, D., Kalfsbeek, H.K. 1995. Elicitation, Assessement, and Pooling of Expert Judgements Using Possibility Theory. IEEE Transactions on Fuzzy Systems. Vol.3, pp.313-335.

[29] Marseguerra, M., and Zio, E. 2002. Basics of the Monte Carlo method with application to system reliability. LiLoLe-Verlag GmbH. Hagen, Germany.

[30] Zio, E. 2007. An introduction to the basics of reliability and risk analysis. Series in Quality, Reliability and Engineering Statistics, Vol. 13.

[31] Baraldi, P., Librizzi, M., Zio, E., Podofillini, L., Dang, V.N.: Two techniques of sensitivity and uncertainty analysis of fuzzy expert systems. Expert Syst. Appl. 36(10): 12461-12471 (2009)

[32] Mamdani, E.H., "Applications of fuzzy logic to approximate reasoning using linguistic synthesis," *IEEE Transactions on Computers*, Vol. 26, No. 12, pp. 1182-1191, 1977.

[33] Sugeno, M., "Fuzzy measures and fuzzy integrals: a survey," (M.M. Gupta, G. N. Saridis, and B.R. Gaines, editors) *Fuzzy Automata and Decision Processes*, pp. 89-102, North-Holland, NY, 1977.

APPENDIX: BASIC CONCEPTS UNDERLYING FUZZY LOGIC THEORY

Introduction

Let X be a variable defined in the set U_X which contains the possible values of X and is usually called the Universe Of Discourse (UOD) of X. In a given UOD U_X an element x is considered not thoroughly characterized in terms of relevant attributes, i.e. it is not possible to affirm for sure anything about the element. Due to the lack of perfect characterization, the task of assigning this not well defined element to some subset of U_X shows two different kinds of uncertainty treated by the *Fuzzy Measure Theory* and by the *Fuzzy Set Theory* respectively [10].

The *Fuzzy Measure Theory* deals with imprecise information where imprecision is defined as the uncertainty in the attribution of x to a particular member of a countable set constituted by disjoint elements. A countable set is a set whose members can be labelled with positive integers.

The *Fuzzy Set Theory* deals with vague information where vagueness is defined as the uncertainty associated with linguistic or intuitive information. Vagueness is a concept related to non-measurable issues and involves situations in which the transitions among linguistic statements have not sharp boundaries. This lack of sharp boundaries between the sets of the UOD represents a fundamental feature of the Fuzzy Set Theory; the various sets overlap and there is a continuous transition from one set to the next one as described by the respective overlapped characteristic functions here called *membership functions*. Correspondingly a given $x \in U_X$ may simultaneously belong to several sets with different degrees of membership: this feature clearly distinguishes the Fuzzy Set Theory from the Probability Theory which operates on crisp disjoint sets.

Fuzzy sets defined over the universal set

Assume that the UOD U_X pertaining to the variable X has been divided in a sequence of n_x subsets $X_l \subset U_X$, $l=1,2,...,n_X$, In the standard theory the subsets X_l are mutually exclusive and a given x may belong to only one of them. Correspondingly these are called *crisp sets* and the membership of a generic x to a set X_l is specified by the rectangular characteristic function χ_{X_l} which is unity or zero according to whether x belongs to X_l or not.

In the fuzzy context the situation is quite different: the subsets X_l of the UOD U_X pertaining to a linguistic variable are not necessarily exclusive so that a given $x \in U_X$ may belong to more that one of them with different grades of membership $\mu_{X_l}(x)$, $l = 1,2, \dots, n_X$. The subsets are not identified by fixed boundaries but instead by linguistic terms (called *words*) which characterize the subset.
Fuzzy set theory aims at quantifying the meanings of the words attached to the subsets X_l within the framework of the set theory. According to the fuzzy set theory, membership functions (MFs), $\mu_{X_l}(x)$, are defined over the sets X_l, $l = 1,2, \dots, n_X$, to represent the degree with which each element $x \in U_X$ is included in one or more sets X_l.

As opposed to the characteristic functions in the standard set theory which as above said are disjoint and rectangular, the shape of a MF is entirely subjective and the various MFs may also overlap. Generally, the appropriate range of each variable, X (i.e., the universe of discourse of that variable, U_X), is a priori established and then divided in subsets (FSs), X_l, characterized by linguistic terms and MFs, $\mu_{X_l}(x)$. The shape of the MFs can be triangular, trapezoidal or quasi-Gaussian.

Summarizing, the fuzzy logic deals with linguistic variables whose arguments are words also called *fuzzy values* (e.g. *negative*, *approximately zero*, *positive*). Each of these words refers to a subset of the universe of discourse and the degree of membership to the word of the crisp values within the subset is analytically specified by the membership functions.

Finally, it is crucial to point out that the MF $\mu_{X_l}(x)$ must not be interpreted as the probability that x belongs to X_l. In fact, while the probability theory deals with well defined events each belonging to only one of several disjoint sets the fuzzy theory concerns concepts which vary gradually over several FSs.

Basic operations between fuzzy sets

As for crisp sets, basic operations between two FSs such as *intersection*, *union* and *complementation* are defined. However, since FSs are defined by their MFs, the set C resulting from an operation between two sets A and B, both defined in the same UOD U_X, must be equipped with a MF $\mu_C(X)$, defined for each $x \in A \cup B$ and obtained by means of an operation between A and B. Correspondingly, it turns out that the set C is constituted by all points of sets A and B whatever the basic operation may be. The basic operations between FSs are:

1. Intersection (the AND operator in crisp set theory);
2. Union (the OR operator in crisp set theory);
3. Complementation (the NOT operator in crisp set theory).

Intersection

Considering a generic *t-norm*, T, as a function of two arguments, non-decreasing in each argument (i.e., if $b < d$, then $aTb < aTd$), commutative, associative and such that $xT0 = 0$ and $xT1 = x$, $\forall x \in [0,1]$, the intersection of two FSs is defined by means of any *t-norm* as:

$$ATB \rightarrow \mu_{ATB}(x) = \mu_A(x) T \mu_B(x)$$

The *t-norms* usually adopted are: the $\wedge$ operator (i.e., the *minimum* operator), the algebraic product and the bounded product, respectively given by:

$$ATB \rightarrow \mu_{A\wedge B}(x) = \mu_A(\mathrm{x}) \wedge \mu_B(x)$$
$$ATB \rightarrow \mu_{A\cdot B}(x) = \mu_A(x) \cdot \mu_B(x)$$
$$ATB \rightarrow \mu_{A\otimes B}(x) = \mu_A(x) \otimes \mu_B(x) = \max(0, \mu_A(x) + \mu_B(X) - 1)$$

The *minimum* norm yields the maximum result among the various *t-norms*.

Union

Considering a generic *s-norm* (also called *t-conorm*), S, with the same properties of the *t-norm* and such that $xS0 = x$ and $xS1 = 1$, $\forall x \in [0,1]$, the union of two FSs is defined by means of any *s-norm* as:

$$ASB \rightarrow \mu_{ASB}(x) = \mu_A(x) S \mu_B(x)$$

The *s-norms* usually adopted are the ∨ operator (i.e., the *maximum* operator), the algebraic sum and the bounded sum, respectively given by:

$$ASB \rightarrow \mu_{A\vee B}(x) = \mu_A(\mathrm{x}) \vee \mu_B(x)$$
$$ASB \rightarrow \mu_{A+B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x)$$
$$ASB \rightarrow \mu_{A\oplus B}(x) = \mu_A(x) \oplus \mu_B(x) = \min(1, \mu_A(x) + \mu_B(X))$$

The *maximum* norm yields the minimum result among *s-norms*.

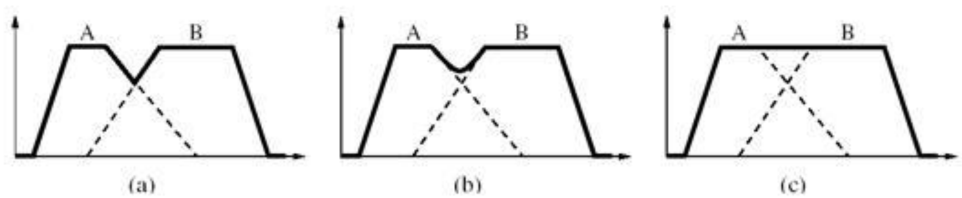

Figure 34. Example of *s-norm* operators: maximum (a), algebraic sum (b) and bounded sum (c).

Complement

Any complement operation, $compl$, on a fuzzy set A must satisfy the conditions $compl(0) = 1$ and $compl\big(compl(a)\big) = a$, $\forall a \in [0,1]$, and is defined as:

$$\overline{\mathrm{A}} = \mu_{\overline{\mathrm{A}}}(\mathrm{x}) = 1 - \mu_A(x) = \bar{\mu}_A(x)$$

When the adopted *t-norm* and *s-norm* are $\wedge$ and $\vee$, respectively, the operations are called *standard fuzzy operations*. In this case, if the values of the MFs are restricted to {0,1}, the standard fuzzy operations coincide with the corresponding crisp operations in terms of classical characteristic functions. This aspect shows that the standard fuzzy operations are the generalization of their classical counterparts.

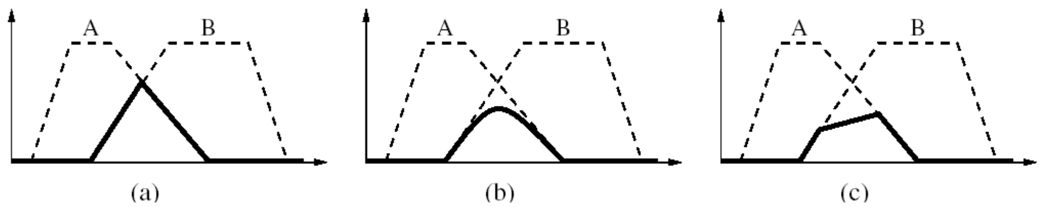

Figure 35. Example of *t-norm* operators: minimum (a), algebraic product (b) and bounded product (c).

Other operations defined for the Fuzzy Sets are the *bounded difference*, the *concentration*, *intensification* and *dilatation*.

Typologies of fuzzy systems

Fuzzy systems operate by means of a set of *if-then* rules which are defined through some *antecedents* and *consequents*, suitably related by fuzzy *connections*. Three main types of fuzzy systems have been implemented by far: Mamdani [32], Takagi-Sugeno-Kang (TSK) [33] and singleton systems. The difference between these systems lies in the structure of the rule consequent part. A brief explanation and discussion about the advantages and limitations of each system follows.

Mamdani systems

The structure of the r-th rule proposed by Mamdani is founded on the Multiple Input Single Output (MISO) approach, in which the single output is constituted by a linguistic term (i.e., a fuzzy set) and a respective membership function:

$$if\ X_1\ is\ X_{1l_1^r}\ and\ X_2\ is\ X_{2l_2^r}\ and\ \ldots\ and\ X_n\ is\ X_{nl_n^r}\ then\ O\ is\ O_v^r$$

where X_q, $q = 1,2, \dots, n_i$ are the input variables, O is the single output variable, $X_{ql_q^r}$ is the l_q antecedent, $l_q = 1,2, \dots, n_{X_q}$, of the q-th input variable pertaining to the r-th rule and O_v is the v-th consequent (i.e., the linguistic term) associated to the output variable in the r-th rule.

The main advantage of Mamdani systems is their high interpretability, due to the clear semantic meaning associated to both the antecedents and the consequent of the rule. On the contrary, the strictness of the linguistic terms which prevents those systems form providing precise results, and the elevate computing time to perform the inference operations render the Mamdani systems not much suitable for practical purposes.

Takagi-Sugeno-Kang (TSK) systems

The TSK approach proposes to adopt a linear affined function of the input variables as the consequent of the rule. The structure of the generic r-th rule can be written as

$$\textit{if } X_1 \textit{ is } X_{1l_1^r} \textit{ and } X_2 \textit{ is } X_{2l_2^r} \textit{ and ... and } X_n \textit{ is } X_{nl_n^r} \textit{ then } O \textit{ is } \left(\boldsymbol{\psi}_r\right)^t \boldsymbol{X} + \boldsymbol{b}_r$$

where $\boldsymbol{X} = \left(X_1, X_2, \dots, X_{n_i}\right)$ are the input variables, $\boldsymbol{\psi}_r = \left(\psi_{r1}, \psi_{r2}, \dots, \psi_{rn_i}\right)$ and $\boldsymbol{b}_r$ are the vectors containing the TSK coefficients and the known term of the linear combination pertaining to the r-th rule, respectively. Other non-linear functions can be employed, too.

Due to their structure, TSK systems provide a higher precision at the expense of a reduced interpretability, since the consequents do not represent a linguistic term anymore. In practice, the TSK models present a black-box structure similar to that of Artificial Neural Networks, thus losing part of the potential advantages offered by the fuzzy models.

Singleton systems

In these systems, the consequent of the rules is constituted by a crisp constant value, o_r, thus the generic r-th rule is expressed as follows:

$$if\ X_1\ is\ X_{1l_1^r}\ and\ X_2\ is\ X_{2l_2^r}\ and\ ...\ and\ X_n\ is\ X_{nl_n^r}\ then\ O\ is\ o_r$$

These systems represent a good compromise between the need of obtaining a precise model and the advantage of dealing with a transparent and interpretable set of fuzzy rules. Moreover, due to the discrete singleton representation, a less computing time is necessary to perform the inference process.

The fuzzy inference process

Conceptually, the inference process is constituted by the *facts*, representing the raw materials, the *reasoning mechanism*, corresponding to the fact transformation, and the inferred *conclusions*, forming the final product.

In a fuzzy context, the *knowledge* on which the fuzzy reasoning relies is translated into a set of rules, the fuzzy *if-then* rules in the Fuzzy Knowledge Base (FKB), or Rule Base (RB), each one representing an *implication* involving the appearance of a *fact* and the attainment of a *conclusion*.

Specifically, the fuzzy reasoning mechanism evolves, as shown in Figure 36 in three main sequential phases:

1. the fuzzification of the facts which converts the crisp variables into fuzzy sets;
2. the fuzzy data processing by means of the fuzzy inference engine which works on the fuzzy Rule Base;
3. the defuzzification of the inferred fuzzy conclusions to obtain a final crisp result.

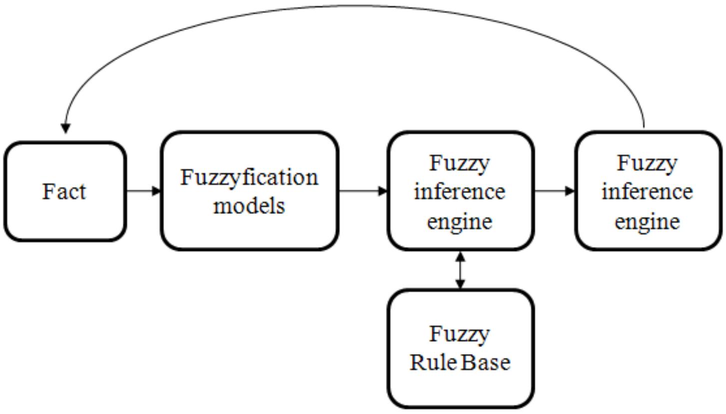

Figure 36. The fuzzy inference process.

The fuzzification module

This module is utilized for fuzzyfying the crisp values, in order to account for their inherent uncertainties, by means of a transformation which provides a mapping between the crisp input variables and the respective degrees of activation of the membership functions.

Let $x \in U_X$ be a crisp input variable and $\mu_{X_l}(x)$, $l = 1,2,\dots,n_X$, the set of membership functions defined over U_X. The mapping between x and $\mu_{X_l}(x)$ is a transformations $L_X: U_X \rightarrow [0,1]^{n_X}$ which corresponds to computing the degrees of membership of x with respect to the MFs, namely

$$L_X(x) = \left(\mu_{X_1}(x), \mu_{X_2}(x), \dots, \mu_{X_{n_i}}(x)\right)$$

and represents the *interface* between the crisp and fuzzy universes.

For example, the fuzzification of a crisp temperature provided by the interface L_T yields the following results (Figure 37):

$$L_T(\theta) = \left(\mu_{T_1}(\theta), \mu_{T_2}(\theta), \mu_{T_3}(\theta)\right) = (0.43,0.77,0.00)$$

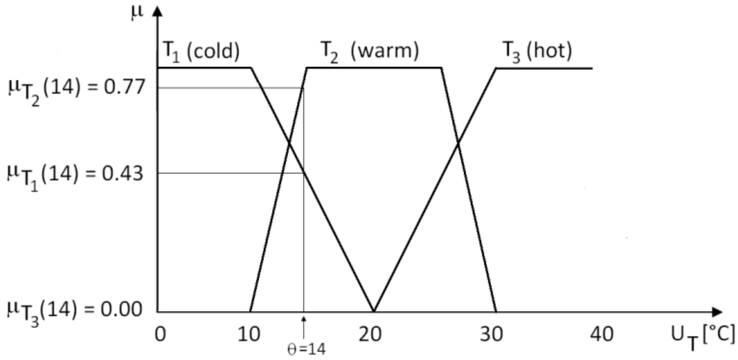

Figure 37. Fuzzification of a crisp temperature value by means of the interface L_T.

Fuzzy data processing

The fuzzy inference engine operates by means of the fuzzy *if-then* rules. This phase is characterized by the evaluation of the *degree of evidence* of the antecedents of each rule, by the computation of the *implication* brought by each rule and by the *aggregation* of the results thereby obtained.

Following the Mamdani approach above explained, a set of fuzzy rules may be represented by a *Decision Table* defined over the FSs relating to the input and output variables. For example, let X_1 and X_2 be two input variables furnished with $n_{X_1} = 3$ and $n_{X_2} = 4$ FSs, namely (X_{11}, X_{12}, X_{13}) and $(X_{21}, X_{22}, X_{23}, X_{24})$, respectively, and O the single output variable related to four FSs, (O_1, O_2, O_3, O_4). A possible *Decision Table* may be written as:

	$\boldsymbol{X_{21}}$	$\boldsymbol{X_{22}}$	$\boldsymbol{X_{23}}$	$\boldsymbol{X_{24}}$
$\boldsymbol{X_{11}}$	O_1	O_2	O_2	O_4
$\boldsymbol{X_{12}}$	---	O_1	---	O_3
$\boldsymbol{X_{13}}$	O_3	O_4	O_3	O_1

Table 5. A possible fuzzy *Decision Table*.

Obviously, a system with n_i inputs produces a n_i-dimensional decision table with a number of cells equal to the product of the FSs (i.e., the linguistic terms) associated to each input variable. Each cell describes a fuzzy rule, except the void cells which mean lack of the corresponding rule.

The activation of the antecedents

The *Facts* and the antecedents of the rules are linguistically interpreted as simple or compound conditional propositions expressed, respectively, in the form:

$$if\ X_1\ is\ X_{1l_1}\ and\ X_2\ is\ X_{2l_2}\ and\ ...\ and\ X_{n_i}\ is\ X_{n_i l_{n_i}}$$
$$if\ X_1\ is\ X_{1l_1}^{r}\ and\ X_2\ is\ X_{2l_2}^{r}\ and\ ...\ and\ X_{n_i}\ is\ X_{n_i l_{n_i}}^{r}$$

where for each input variable X_q, X_{ql_q} and $X_{ql_q}^{r}$ are the FSs of the *Fact* and of the antecedent of the *r*-th rule, respectively. The degree of activation of the antecedents of a rule is a crisp number which represents the *degree of evidence* with which the incoming *Fact* satisfies that rule (i.e. the so-called *consistency of the Fact*).

Let X and X^r symbolize the *Fact* and the r-th rule antecedents, respectively. In general, the degree of evidence, $s_r(X, X^r)$, with which the *Fact* satisfies the antecedents of the generic r-th fuzzy rule is a crisp number obtained by the *t-norm* operation between the partial degrees of evidence, $s_r\left(X_{ql_q}, X_{ql_q}^{r}\right)$, with which each FS of the *Fact*, X_{ql_q}, pertaining to the input variable X_q, satisfies the respective antecedent, $X_{ql_q}^{r}$, in the r-th fuzzy rule:

$$s_r(X, X^r) = T_{q=l}^{n_i} s_r\left(X_{ql_q}, X_{ql_q}^{r}\right)$$

The partial degree of evidence for the input variable X_q is a crisp number calculated by means of the generic operator, Δ (usually the *minimum* or *maximum* operator), acting on the MF,

$$\mu_{X_{ql_q} T X_{ql_q}^{r}}(x_q) = \mu_{X_{ql_q}}(x_q) T \mu_{X_{ql_q}^{r}}(x_q) \qquad x_q \in U_{X_q}$$

constructed by any *t-norm* operation performed between the MF pertaining to the *Fact*, $\mu_{X_{ql_q}}$ and the MF of the r-th rule antecedent, $\mu_{X_{ql_q}^r}$:

$$s_r\left(X_{ql_q}, X_{ql_q}^r\right) = \Delta_{x_q}\left(\mu_{X_{ql_q} T X_{ql_q}^r}\right)$$

For example, let the MFs of the input X_1 and X_2 be the triangular ones in Figure 38.

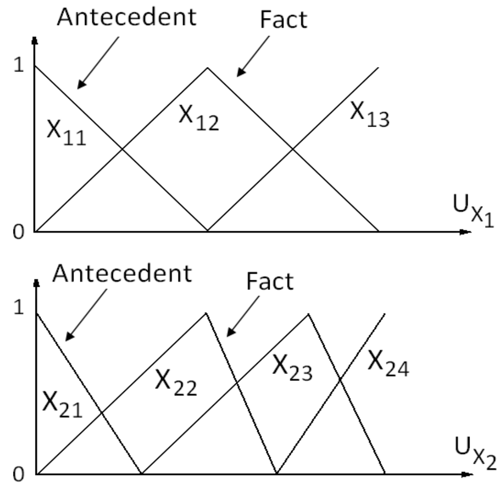

Figure 38. FSs and MFs for the input variables X_1 and X_2.

Moreover, let the Rule Base be constituted by only one rule, for example:

$$if\ X_1\ is\ X_{11}^1\ and\ X_2\ is\ X_{21}^1\ then\ O\ is\ O_1^1$$

and the *Fact* be expressed as: "*if* X_1 *is* X_{12} *and* X_2 *is* X_{22}", as indicated in Figure 38 and, finally, let the minimum operator, $\wedge$, be the adopted *t-norm* and the maximum operator, $\vee$, represent the operator Δ.

Graphically, the partial degrees of evidence are the maximum of the MFs, $\mu_{X_{12} \cap X_{11}^1}(x_1)$ and $\mu_{X_{22} \cap X_{21}^1}(x_2)$ (bold in for $x_1 \in U_{X_1}$ and $x_2 \in U_{x_2}$, respectively), obtained, for each input variable, by taking the minimum value of the superposition of the MFs of the *Fact* and of the rule's antecedent pertaining to that variable:

$$s_1(X_{12}, X_{11}^1) = \bigvee_{x_1 \in U_{X_1}} \left(\mu_{X_{12}}(x_1) \vee \mu_{X_{11}^1}(x_1)\right)$$

$$s_1(X_{22}, X_{21}^1) = \bigvee_{x_2 \in U_{X_2}} \left(\mu_{X_{22}}(x_2) \vee \mu_{X_{21}^1}(x_2)\right)$$

Finally, the consistency of the *Fact* is computed as the minimum of the partial degrees of evidence:

$$s_1(X, X^1) = s_1(X_{12}, X_{11}^1) \wedge s_1(X_{22}, X_{21}^1)$$

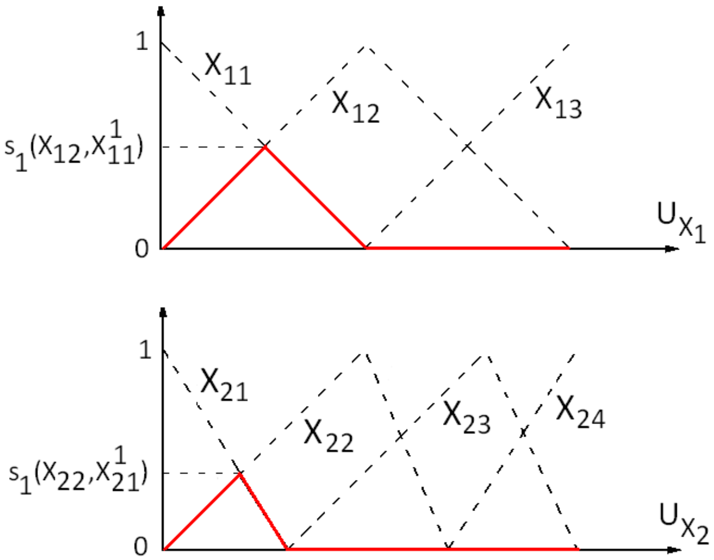

Figure 39. Graphic computation of the partial degrees of evidence.

So far we have considered the case in which the \emph{Fact} is expressed in the form:

$$if\ X_1\ is\ X_{1l_1}\ and\ X_2\ is\ X_{2l_2}\ and\ \ldots\ and\ X_n\ is\ X_{nl_n}$$

where X_{1l_1} ,…, X_{nl_n} are fuzzy sets. An example may be "*if* the humidity is *high then* the temperature is *low*". Anyway, in many practical situations, the input variables assume a crisp value:

$$if\ X_1\ is\ x_1\ and\ X_2\ is\ x_2\ and\ \ldots\ and\ X_n\ is\ x_n$$

being x_q the actual value of the *q-th* variable X_q. For example: "If the humidity is 67% and the temperature is 23°C". In this case, the partial degrees of evidence $s_1\left(x_q, X^r_{ql_q}\right)$ with which each value of the fact x_q, pertaining to the input variable X_n, satisfies the antecedent $X^r_{ql_q}$ of the *r-th* fuzzy rule is simply:

$$s_1\left(x_q, X^r_{ql_q}\right) = \mu_{X^r_{ql_q}}\left(x_q\right)$$

as shown in Figure 40.

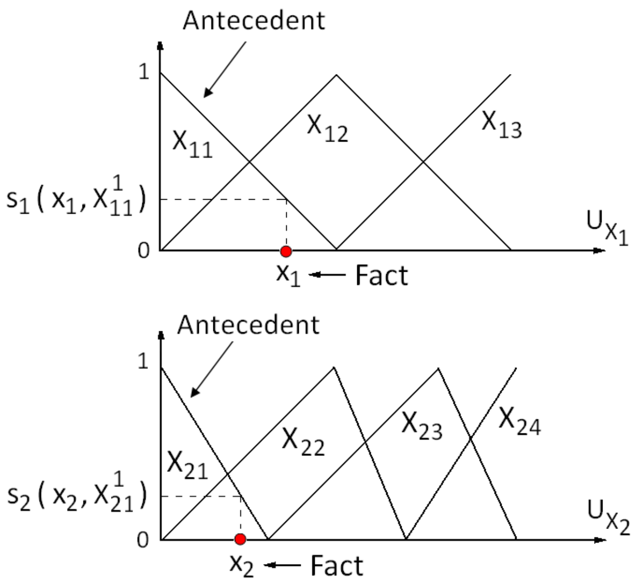

Figure 40. Graphic computation of the partial degrees of evidence when the *Fact* is crisp value

The implication

The fundamental task of the fuzzy reasoning is to obtain the membership function of the output variable, whether it is a MF, as for Mamdani systems, or it is represented by a single-point fuzzy set, the so-called *singleton*. The most important inference procedures are the General Modus Ponens (GMP) and the General Modus Tollens (GMT), which are the fuzzy analogous of the crisp evaluation of a function and of the inverse of a function, respectively. This Section will expound only the GMP inference procedure.

The term *modus ponens* means *method of affirming* since the conclusion is an affirmation. It is founded on a premise and for an incoming fact it furnishes a conclusion.

In general, let X_q, $q = 1,2,\dots,n_i$, be the set of input variables and X_{ql_q}, $l_q = 1,2,\dots,n_{X_q}$, the FSs pertaining to each variable, O be the single output variable and O_v, $v = 1,2,\dots,n_O$, the related FSs. The n_i-dimensional decision table is therefore constituted by

$$\rho = \sum_{q=1}^{n_i} n_{X_q}$$

fuzzy *if-then* rules and the inference engine therefore reads:

Premise 1: if X_1 *is* $X_{1l_1^1}$ *and* X_2 *is* $X_{2l_2^1}$ *and ... and* X_{n_i} *is* $X^1_{n_i l_{n_i}}$ *then* O *is* O^1_v *else*

Premise 2: if X_1 *is* $X_{1l_1^2}$ *and* X_2 *is* $X_{2l_2^2}$ *and ... and* X_{n_i} *is* $X^2_{n_i l_{n_i}}$ *then* O *is* O^2_v *else*

...

Premise ρ*: if* X_1 *is* $X_{1l_1^\rho}$ *and* X_2 *is* $X_{2l_2^\rho}$ *and ... and* X_{n_i} *is* $X^\rho_{n_i l_{n_i}}$ *then* O *is* O^ρ_v

Fact: if X_1 *is* X_{1l_1} *and* X_2 *is* X_{2l_2} *and ... and* X_{n_i} *is* $X_{n_i l_{n_i}}$

Conclusion: O *is* O'

The *premise* is the so-called *implication* and is defined over a FS which can be written, for the r-th premise, as:

$$\bigcap_{q=1}^{n_i} X_{ql_q^r} \rightarrow O_v^r$$

The connective "*else*" performs the *aggregation* between premises, as explained later. The lack of a rule, in correspondence of a possible combination of the FSs of the input variables, is handled by setting the consequent MF to zero.

Knowing, for each premise, the MFs of the antecedents and of the consequent, the *implication* relation and the *Fact*, the purpose of the fuzzy reasoning is to provide the MF $\mu_{O'}(o)$ (i.e., the linguistic term O' of the model output O).

The defuzzification module

The inferred MF of the single output variable, $\mu_{O'}(o)$, represents the linguistic term associated to the output variable. Most of the times, in order to be effective, the linguistic term must be converted into a crisp number.

For example, if dealing with the control of a plant, the output MF indicates the action to be taken by the controller in correspondence of the Fact, but it is always necessary to convert it into a number which representing the signal sent to the plant. This operation, called defuzzification, may be conducted in several ways. The methods most usually adopted are:

1. the Center Of Area (COA);
2. the Center Average (CA);
3. the Mean Of Maxima (MOM), also called the α-cut method).

In the *Center Of Area* method, the required crisp value, o_{COA}, is defined as the abscissa of the center of gravity of $\mu_{O'}(o)$. In the continuous case, the COA method yields:

$$o_{COA} = \frac{\int_{o \in U_O} o \cdot \mu_{O'}(o)\, do}{\int_{o \in U_O} \mu_{O'}(o) do}$$

The *Center Average* method is directly connected with the consistency of the *Fact*. In general, the crisp value, o_{CA}, is attained by means of any *t-norm* in the following way:

$$o_{CA} = \frac{\bigvee_r \left(\tilde{o}_r T \mu_{O'}^{(r)}(\tilde{o}_r) \right)}{\bigvee_r \left(\mu_{O'}^{(r)}(\tilde{o}_r) \right)}$$

where $\vee$ represents the *maximum* operator, $\tilde{o}_r$, is the abscissa corresponding to the maximum value of $\mu_{O'}(o)$ (i.e., the MF the consequent used by the r-th implication), and $\mu_{O'}^{(r)}(o)$ is the MF inferred by the r-th implication.

In the *Mean Of Maxima* method, the crisp output is obtained by selecting the most relevant output interval $[o_1^\alpha, o_2^\alpha]$ to be employed for the crisp output computation, where α is the ordinate at which the output MF is "cut". In the MOM method, α is chosen as the maximum of $\mu_{O'}(o)$ and the crisp output is calculated as:

$$o_{MOM} = \frac{\int_{o_1^\alpha}^{o_2^\alpha} o \cdot do}{\int_{o_1^\alpha}^{o_2^\alpha} do}$$