MARY KAY O'CONNOR PROCESS SAFETY CENTER
TEXAS A&M ENGINEERING EXPERIMENT STATION

22nd Annual International Symposium
October 22-24, 2019 | College Station, Texas

How to Treat Expert Judgment? With certainty it contains uncertainty!

Hans J. Pasman* and William J. Rogers
TEES Mary Kay O'Connor Process Safety Center
Artie McFerrin Department of Chemical Engineering
Texas A & M University, College Station, 77843 Texas, USA
*Presenter Email: hjpasman@gmail.com

Abstract

To be acceptably safe one must identify the risks one is exposed to. It is uncertain whether the threat really will materialize, but determining the size and probability of the risk is also full of uncertainty. When performing an analysis and preparing for decision making under uncertainty, quite frequently failure rate data, information on consequence severity or on a probability value, yes, even on the possibility an event can or cannot occur is lacking. In those cases, the only way to proceed is to revert to expert judgment. Even in case historical data are available, but one should like to know whether these data still hold in the current situation, an expert can be asked about their reliability.

Anyhow, expert elicitation comes with an uncertainty depending on the expert's reliability, which becomes very visible when two or more experts give different answers or even conflicting ones. This is not a new problem, and very bright minds have thought how to tackle it. But so far, however, the topic has not been given much attention in process safety and risk assessment. The paper has a review character and will present various approaches with detailed explanation and examples.

Keywords: Probability, imprecision, expert elicitation, Dempster-Shafer theory, Fuzzy sets, and logic

1. Introduction

It seems so easy: once in case in an analysis, *e.g.*, a LOPA, the complexity of the problem structure has been solved, as a next and final step data must be filled in, such as the probability of failure on demand of a critical component, *e.g.*, a pressure relief valve. However, one knows that manufacturer data are too optimistic, while the OREDA data base [1] may provide an answer not valid under the conditions the valve is applied in an actual process. In the best case, there is in the plant some historical data available on similar valves, but those are of a different brand creating another uncertainty. As a last resort, a few people familiar with the installation and employed already a considerable number of years at the plant may be asked to provide an estimate. This reverting to expert judgment can have different forms. Easiest for the expert may

be to give a linguistic grading term ('high' or 'low') or just a single figure, but knowing there will be uncertainty the interviewer may invite the experts to specify an interval with both a lower and a higher bound. Even more sophisticated is to ask for an estimate of a mean value and a confidence or credibility interval. Experts may also be asked about different information, *e.g.*, whether a run-away in case of a certain batch-process is thought possible or plausible under the expected condition ranges.

In a wider context, questioning experts serves to support optimum decision making. There are many methods developed for that purpose, for example Saaty's [2] method of Analytical Hierarchy Process (AHP) developed in the 1970s, in which participants must make pairwise selection between alternatives on the basis of a number of criteria. More sophisticated is multi-attribute utility theory to make a decision based on value judgments of multiple, competing objectives. However, this paper will treat the use of expert judgment in risk assessment questions.

Already decades ago, statisticians engaged in intense deliberations on how to deal with uncertainty. It all starts with deep thoughts about the concept of probability. Oldest is the so-called frequentist idea in which an experiment is performed of which the outcome can take different forms or values. A classic case is the throw of a dice, or determining a sample of a large population of red and white balls in an urn and predicting the fraction of red. The objective is to determine a distribution of outcomes enabling prediction of future action outcomes or composition of the population without counting them all. This developed to a collection of distribution functions in which results could be fitted and to significance levels with confidence limits dependent on the variability of draws and the size of the sample. Later, the Bayesian approach gained strength, in which all previous information/knowledge can be cast into a prior distribution, while new evidence is represented by a likelihood function, and the result is an update of the prior to a posterior distribution based on the normalized co-occurence of the prior and the likelihood. The Bayesian model provides many more possibilities in solving problems than frequentist statistics and has become the leading approach for evidence-based testing of a hypothesis or an event.

In many situations, though, when asked in human dialogue to make an estimate or a prediction of the probability of an event occurrence, a person can produce such a probability value based on intuition supported by experience. Such a probability estimate is called a subjective or imprecise probability value. It fits well with the Bayesian approach, which works with information of all uncertainty levels and propagates the uncertainties. With new evidence in the likelihood, the prior is updated to a posterior result with lower uncertainty than the prior. Besides a development of a strict probabilistic approach by Cooke [3] in the 1980s that we shall consider later, in the 1960s and 70s, two different less strict theories evolved dealing with uncertainty:

- Dempster-Shafer theory of evidence, developed by Dempster [4] and later Shafer [5] on belief and plausibility. Also with the Dempster rule, replies from different responders can be combined. Strict probability theory requires that the probability an event will occur and the probability it will not occur, sum to unity, 1. A human making an estimate will often not guarantee, though, that the complement of the answer given, fulfills that requirement. He/she may not know, hence is ignorant, or at least he/she is unsure it will.

- Fuzzy set and logic by Zadeh [6, 7] in which an interviewer will obtain an imprecise, hence fuzzy, answer represented by a membership function with the value 1 at the given estimate and with value zero at a minimum below which and a maximum value above which the estimate value is believed to be not possible. The membership function can have any shape between 0 and 1.

Over the years, these theories have been further developed and applied for various purposes, such as Dempster-Shafer in sensor fusion for target identification in a cluttered environment and Fuzzy set and logic in classification problems. Fuzzy logic enables solving the problem of combining unsharp values of characteristic properties of related items or kinds to a desired final result, such as the known example of food quality wasn't bad, service was perfect, price is okay, hence tip can be decent. Different values of characteristics will infer different outcomes. Both Dempster-Shafer theory and Zadeh's fuzzy set and logic are widely applied to facilitate decision making in uncertain situations.

Uncertainty is usually categorized by two types, although distinction is not always very clear and a variable can contain both types of uncertainty at the same time:

- Aleatory uncertainty, by which due to lack of accuracy/precision of observational means, in general random variability from ranges of conditions, an outcome cannot be established accurately, and,
- Epistemic uncertainty, which is a consequence of lack of knowledge about the subject due to the amount and quality of the data.

In particular, the epistemic uncertainty will be addressed here.

By the end of 1980s, Klir [8] in the Cambridge Debate on Uncertainty, wrote a clear synthesizing paper *against* the claim that probability, as traditionally defined, the standard approach, is the only concept to describe uncertainty. He did this with a counterclaim that one must go beyond only probability. Klir starts off with distinguishing two types of uncertainty:

- Vagueness, encompassing: fuzziness, haziness, cloudiness, unclearness, indistinctiveness, sharplessness and indefiniteness, and,
- Ambiguity, comprising of: non-specificity, variety, generality, diversity, divergence, equivocation, incongruity, discrepancy, dissonance, disagreement.

He continues by mentioning that *imprecision* can relate to both vagueness and ambiguity, while in the latter non-specificity and disagreement are again different. After analyzing the matter in much detail, Klir [8] concludes that probability conceptualizes "uncertainty strictly in terms of conflict among degrees of belief allocated to mutually exclusive alternatives"; in other words, a probability P of an event or quantity intrinsically holds the contrast that the probability the event will *not* occur or the quantity will be different, will be the complement $1 - P$. Hence, the standard approach to probability covers only part of uncertainty.

Uncertainty is also related to less precise information and in that context Zadeh [9], after having launched in 1965 the fuzzy set theory mentioned above, published in1977 the possibility theory. The latter states that where imprecision is inherent to natural language, in case the meaning of information is the objective and not its measure, one can speak of "possibilistics", rather than probabilistics, and even of a possibility distribution as a counterpart to the probability one. All of this found its way to Artificial Intelligence techniques.

More recently in 2011, Helton and Johnson [10] summarized the alternative representations of epistemic uncertainty at increasing structure and quantification as follows: (1) Interval analysis, just providing a low and high boundary with no information in between (uniform distribution with all values equally likely); (2) Possibility theory, consisting of a set of possible elements to each of which a likelihood value can be attached together forming a possibility distribution, which is related to the Fuzzy set approach; (3) Evidence theory (Dempster-Shafer), which specifies a limited number of focal elements, while each element is given a measure of credibility (basic assignments or basic belief assignments summing to 1, confusingly also called basic probability assignment); (4) Probability theory, involving element probabilities in a fully developed structure embodied by a probability density function.

At this probability end of the spectrum, Cooke developed so-called *structured expert judgment* attempting to make the process of interrogating experts on a probability value, in which there will be always disagreement among experts, as transparent as possible. This is realized by introducing a set of strict methodological rules. These rules lead among others to calibration of the experts and to the individually scoring of performance-based weights.

In the remainder of the paper, we shall restrict ourselves to the more practical aspects of expert estimation. In Section 2 we shall describe the Dempster-Shafer approach in more detail, and in Section 3 the Fuzzy set and logic one, both with some examples. In Section 4 Cooke's method will be described, and in Section 5 similarities and differences, also in required effort, will be summarized, followed by Section 6 with conclusions.

2. Dempster Shafer Theory (DST) of evidence

In various publications Shafer [5, 11] explains the original idea of belief functions and evidential reasoning in case a human makes a statement about an event, fact, or value. It encompasses belief, doubt, plausibility, disbelief, and ignorance, all associated with uncertainty. For example, if a person asserts that a certain event took place or is going to take place, it does not mean that there is no space to believe it did not occur or is not going to occur. The statement can gain strength, hence support, by evidence – a Bayesian element[1] -, while the reliability of the person making the statement can be estimated by a different person knowing the one making the statement (where reliability and unreliability sum to 1). This reliability is called weight of rather mass, m. Generalizing and following the practical application to reliability engineering by Rakowsky [13], hypotheses or information pieces of a data source, *e.g.*, experts' estimates on states or events, may constitute a set of elements, A (a frame, formally called a *frame of discernment* or of *disjoint states*). For example, suppose the set A contains three of each other independent pieces of evidence or elements A_1, A_2, and A_3, then, the power set[2] Ω of $A = \{\{\emptyset\},\{A_1\},\{A_2\},\{A_3\},\{A_1, A_2\},\{A_1, A_3\},\{A_2, A_3\},\{A_1, A_2, A_3\}\}$. Hence, with n elements the number of sub-sets is 2^n, or with three information pieces or so-called *focal*

[1] Dempster and Shafer asserted compatibility of their approach with the Bayesian updating of existing information with new evidence, and Shafer even called their theory a "generalization of the Bayesian theory of subjective probability judgment", evoking interesting and very clear comments by Judea Pearl [12]. The latter was in 2011 awarded with the ACM Turing award in computer science for introducing probabilistics in artificial intelligence and because of his foundational work on causality and on Bayesian networks.

[2] Power set Ω is the set of all sub-sets, mathematically formulated as $A \subseteq \Omega$. The number of subsets is 2^{Ω}.

elements, $2^3 = 8$. Each subset can be assigned a mass, called *basic assignment*[3] or basic belief assignment, depending on judgment of its trustworthiness. Pearl [12] called the assignment the "probability of provability". The sum of masses must be 1 (mathematically, $m : 2^{\Omega} \rightarrow [0,1]$). The $\emptyset$ null set represents the uncertainty whether any more information on the subject exists; if it is thought certain there is no more, $m(\emptyset) = 0$.

Due to subjectivity, it is uncertain which piece of evidence has the highest reality value. Also, because of multi-value statements and other complications, the explanatory wording by different scholars of the Dempster-Shafer Theory (DST) implied relations that can slightly differ and may confuse. At this stage, the belief function or belief structure shall be introduced. In this, *belief* represents the lower bound of mass associated with a focal element or a collection of those, supporting the expert claim or trustworthiness of the source, and plausibility represents the upper bound. In contrast to many papers, we shall start with *disbelief* in a subset A_i out of set A, or belief in not $\bar{A}_i$. It is easy to see that this is the complement of the sum of all supporting mass or evidence $m(A_j)$ of any focal element A_j intersecting with A_i including A_i itself, hence contributing to the plausibility of A_i, while excluding null sets: $bel(\bar{A}_i) = 1 - \sum_{A_j \cap A_i \neq 0} m(A_j)$ (to be clear j includes i). Thus, plausibility of A_i can be defined as $pl(A_i) = 1 - bel(\bar{A}_i)$. Next will be the definition of belief, or the degree in which A_i is believable: $bel(A_i) = \sum_{A_j \subseteq A_i;\, A_j \neq \emptyset} m(A_j)$, hence the mass of focal elements A_j, which are *subsets* of A_i and A_i itself. Its complement is *doubt* as represented in Figure 1. It always holds that $bel(A_i) \leq pl(A_i)$, while the two are not additive nor add to 1.

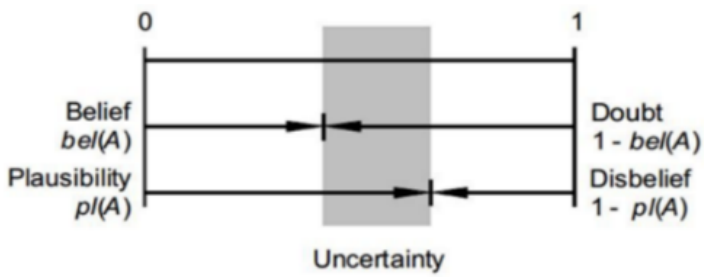


Figure 1. Representation of the various qualifying functions given evidence or data and the margin of uncertainty or ignorance according to Rakowsky [13].

Summarizing, $m(A)$ measures the degree of belief in claim A; $bel(A)$ measures the total degree of evidence support by the non-null focal element subsets of the set of interest A and that of A itself, so it represents a minimum of support, while $pl(A)$ measures the total support of all elements intersecting with A, and not null, hence a maximum of support. The difference represents the uncertainty. In continuation of the example above with j = 1 to 3:

$$bel(A) = \sum_{A_j \subseteq A;\, A_j \neq \emptyset} m(A_j),\ e.g.,\ bel(A_1, A_2) = m(A_1) + m(A_2) + m(A_1, A_2) \quad (1)$$

$$pl(A) = \sum_{A_j \cap A \neq 0} m(A_j), \quad e.g.,\ pl(A_1, A_2) = m(A_1) + m(A_2) + m(A_1, A_2) + m(A_1, A_3) + m(A_2, A_3) + m(A_1, A_2, A_3) \quad (2)$$

[3] Shafer [5,11] and others called m basic probability assignment or *bpa*, but later, *e.g.*, Rakowsky [13] warned that actually it should not be confused with a probability value, so it is preferable to speak of basic assignment.

If a second person confirms the statement made but the reliability of this person is estimated differently, Dempster's combination rule (intersection, AND, joint) to be considered as updating just as in the Bayes theorem, calculates the final weight:

$$m_{12}(A|A \neq \emptyset) = \frac{1}{1-k}\sum_{B\cap C=A} m_1(B)m_2(C);\ \ m_{12}(\emptyset) = 0;\ \ k = \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$$

where $m_{12}(A)$ is the combined reliability of common focal elements A, $m_1(B)$ is the reliability of focal elements B of the first person and $m_2(C)$ are those of C of the second, while k represents the effect of conflicting elements, hence, those that do not intersect, $B\cap C = \emptyset$. Summing intersecting elements, $B\cap C \neq 0$, though, is often an easier way to derive k via its complement.

Example 1: asking experts about a possibility of event occurrence
Applying the rule is not in every case straightforward. The following is an example inspired by one of Shafer [10] applied to the process safety area of interest. When asking two experts independently of each other whether for a certain batch process a run-away is possible or not, both respond positively. So, in the combination there are two focal elements with a runaway possibility. There is no conflict, but the older expert is estimated to have a reliability m_1 = 0.8 and the younger one m_2 only 0.6. When the joint unreliability (1 – 0.8)(1 – 0.6) = 0.08 is calculated, $k = 0$, as there is no intersection with a null element. Hence, the reliability of the joint opinion m_{12} increases to 1 – 0.08 = 0.92. If conflicting opinions appear and the younger states that runaway is not possible, in the combination will be an intersection with a null set, hence k = 0.8×0.6 = 0.48. The unreliability of the younger person, (1 – 0.6), leaves the possibility of runaway still open, so that the extent of belief m_{12} of the statement runaway is possible decreases to 0.8·(1 – 0.6)/(1 – 0.48) = 0.62, and in analogy, that of runaway will not be possible 0.23. If, for example, probabilities of an event are estimated and two fully reliable experts differ strongly of opinion but coincidentally agree with low probability on the same alternative possibility, the rule no longer provides a correct answer as the rule yields mass = 1 for that low probability; later Yager [14] and others suggested improvements, see for more details also Sentz and Ferson [15].

The above example was on an event possibility, but DST can be applied also to failure rates. We shall follow Rakowsky [13], Simon et al. [16] and Khakzad [17] to see how we can apply DST in daily life of risk assessment to objectivize expert estimates of state probability and belief mass values and using those in fault and event trees, formally modeled in evidential network representing belief and plausibility but solved by Bayesian network algorithms. More reading and applications can be found in [15], [18] and [19]. Although the DST allows analyzing relatively complex situations, as examples the simplest scenarios will be considered here.

Example 2: an expert estimating probability of successful functioning and failure
Assume a component that will have one failure mode, so it has only two states it can be in: functioning successfully (S) or failed (F). The frame of discernment is: $\{\{\emptyset\},\{S\},\{F\},\{S,F\}\}$. The component is either in S or F -state (mutual exclusiveness), so the first and last term can be eliminated as the knowledge with the two middle elements about which state the component is in, is complete; however, if the latter is not the case, the set $\{S,F\}$ will express the epistemic uncertainty or ignorance about the state the component is in. Anyhow, given the expert is fully trusted, the probabilities provided can be taken equal to the masses, because these probabilities

represent the best extent of belief. For example, when asked, an expert estimates the successful functioning probability over the next year to be 0.7, and the failure probability in that same year to be 0.2 (two in ten). The sum of masses over the singletons must equal 1, so the mass of $\{S, F\}$ = 0.1. In this case, $bel(S)$ and $pl(S)$ are equal to 0.7 and 0.8, whereas $bel(F)$ and $pl(F)$ will be 0.2 and 0.3. Simon et al. [16] and Khakzad [17] show how epistemic uncertainty can be propagated as a separate 'state' using the Bayesian network (BN) infrastructure. In Figure 2 an example is presented of a fault tree modeled as a BN.

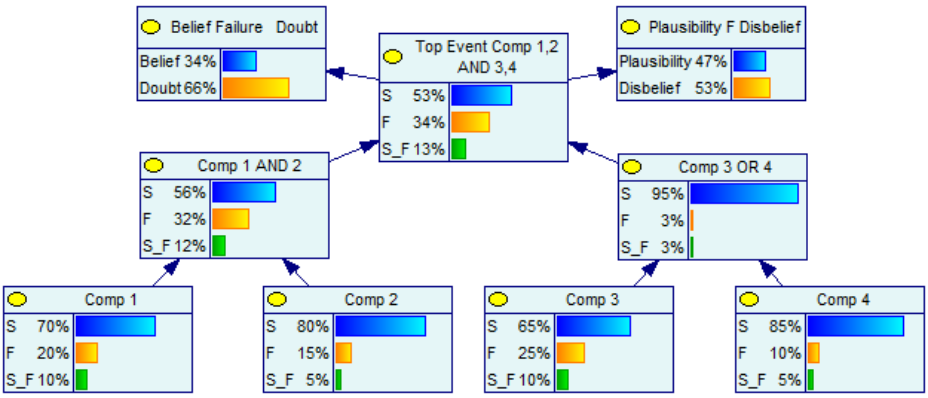


Figure 2. Evidential network showing the result of estimates of functional success probability$\{S\}$ over one year, failure probability$\{F\}$ (for demonstration assumed high) and remaining epistemic uncertainty$\{S, F\}$ of two parallel components in AND and two in OR position solved by Bayesian network (GeNie, BayesFusion LLC), following the reasoning of Simon et al. [16] and Khakzad [17].

Example 3: two (or more) experts are asked to estimate probability values

Alternatively, two experts A and B can be asked to provide probability values. Fusing the values they give, shall be performed using the combination rule as shown by Rakowsky [13]. A simple example is worked out in the following Tables 1-3:

Table 1. Expert focal element input data

	Expert A	*m*	*Expert B*	*m*
$\{S\}$	A1	0.7	B1	0.8
$\{F\}$	A2	0.2	B2	0.15
$\{S, F\}$	A3	0.1	B3	0.05

Table 2. Combination table of singleton intersections

	A1	A2	A3
B1	$\{S\}$	0	$\{S\}$
B2	0	$\{F\}$	$\{F\}$
B3	$\{S\}$	$\{F\}$	$\{S, F\}$

Table 3. Support results for the combined singletons;

$\{S\}$			$\{F\}$			$\{S, F\}$		
R1	A1∩B1	0.560	R4	A2∩B2	0.030	R7	A3∩B3	0.005
R2	A1∩B3	0.035	R5	A2∩B3	0.010			
R3	A3∩B1	0.080	R6	A3∩B2	0.015			
Σ		0.675	Σ		0.055	ΣΣ	=	0.735
$m\{S\}$ = 0.675/0.735 = 0.918			$m\{F\}$ = 0.055/0.735 = 0.075			$m\{S, F\}$ = 0.005/0.735 = 0.007		

The belief and plausibility of the resulting focal elements is: $bel(\{S\})$ = 0.675; $pl(\{S\})$ = 0.925; $bel(\{F\})$ = 0.075; $pl(\{F\})$ = 0.082; $bel(\{S,F\})$ = 0.005; $pl(\{S,F\})$ = 1. This procedure can be applied also if one expert provides interval values instead of point values.

From the above reasoning for a binary system, use can be made to solve more complex situations with multiple modes of failure, such as ternary (Rakowsky [13]) and quaternary systems (Khakzad [17]).

Example 4: different ways to interview experts

DST can also be applied in a slightly different sense. Curcurù et al. [20] describe two examples, both used with respect to failure rates in a fault tree:

- Example 4.1: an analyst sets a basic assignment value for an upper and a lower bound of trustworthiness, whereas experts produce an interval of failure probability values that to their opinion corresponds to the basic assignment bounds.
- Example 4.2 is the other way around: if there are historical data available, the analyst proposes an upper and lower bound value and (two) experts may judge trustworthiness by providing a basic assignment value for applying the data in an actual, concrete case.

Because both cases will be developed similarly, we shall consider example 2 more closely, and select for that the fault tree of two basic components A and B in parallel, connected by an AND gate. In Table 4 the input data of basic assignment values m and corresponding failure probability values for the two components and the two experts are presented. For the failure probabilities Curcurù et al. [20] did not define a time span, which could, *e.g.*, be one year.

In Table 5 the basic assignments of the two experts are aggregated; in Table 6 the AND intersection is realized by solving the interval multiplication. The results of Table 6 are plotted in Figure 3.

Table 4. Input data according to the example given by Curcurù et al. [20], arranged differently.

Expert 1					Expert 2				
	Lower bound		Upper bound			Lower bound		Upper bound	
m	0.9	0.1	0.9	0.1	*m*	0.7	0.3	0.7	0.3
Comp. A	2E-03	0	4E-03	1	Comp. A	2E-03	0	4E-03	1
m	0.9	0.1	0.9	0.1	*m*	0.6	0.4	0.6	0.4
Comp. B	3E-03	0	5E-03	1	Comp. B	3E-03	0	5E-03	1

0 = component will never fail
1 = it will always fail

Table 5. Application of the combination rule on the *m*-values: $(1-(1-m_1)(1-m_2))$.

Expert 1 + 2				
	Lower bound		Upper bound	
m	0.97	0.03	0.97	0.03
Comp. A	2E-03	0	4E-03	1
m	0.96	0.04	0.96	0.04
Comp. B	3E-03	0	5E-03	1

Table 6. AND intersection resulting in probability of failure of the top event

AND gate => Comp. A × Comp. B = interval multiplication:
$[x_1, x_2] \cdot [y_1, y_2] = [\min(x_1y_1, x_1y_2, x_2y_1, x_2y_2), \max(x_1y_1, x_1y_2, x_2y_1, x_2y_2)]$

	Lower bound		Upper bound	
m	0.931	0.001	0.931	0.001
x_1y_1, x_2y_2	6E-06	0	2E-05	1
m	0.039	0.029	0.039	0.029
x_1y_2, x_2y_1	0	0	4E-03	5E-03

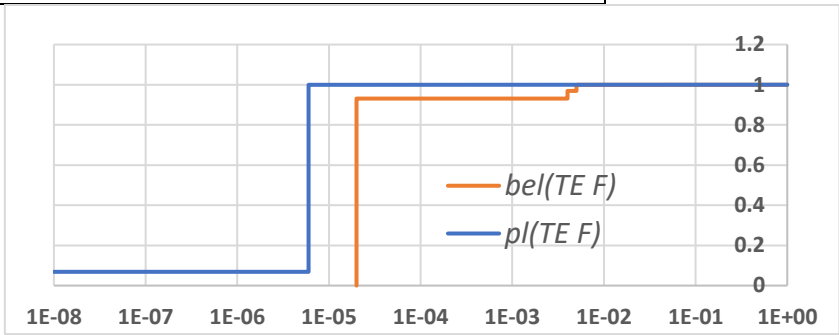


Figure 3. Belief (lower bound) and plausibility (upper) as by Curcurù et al. [20], applying bel and pl definitions (here only plausibility values had to be summed, from the smallest up).

In case of OR-gate failure Component A and B probabilities must be added and their co-occurrence subtracted from the sum. Example 4.1 can be evaluated similarly as example 4.2.

3. Fuzzy sets and logic, type 1 and type 2

Type-1 fuzzy sets

Zadeh's 1965 fuzzy sets and logic [6] for dealing with uncertainty have become quite known, while mid-1990s Klir and Yuan [21] showed their applicability. In the late 1990s and in this century Mendel [22] broadened and deepened the concept. Application is relatively straightforward and wide-spread, and it is used in risk assessment to estimate values or to express even linguistic grades of consequence severity or event frequency. Oldest is what in hindsight is called Type-1 fuzzy set, which originated in classifying types where it is not always possible to sharply describe characteristics or criteria. To use the words of Zadeh: due to imprecision the class will have a "continuum of grades of membership". In other words, the class or fuzzy set A is characterized by a membership function associating each point of A on the real line (X) on which the set extends from the extremes x_a to x_b, with a membership value (μ) in the interval [0, 1]. At x_0, $\mu = 1$, while at x_a and x_b, $\mu = 0$. The membership function can have any shape, but this is often assumed to be a triangle, trapezoid, or Gaussian, see Figure 4. Hence, $A = \{(x, \mu_A(x)|x \in X\}$ or in words: A is a function of x and μ, given variable x is part of the universe of discourse X, or $A = \int_{x \in X} \mu_A(x)/x$, where the quotient symbol or slash means group associating all elements in X with $\mu_A(x) > 0$, and the integral $\int$ can be replaced by a summation Σ in case values are discrete. A horizontal cross-section is called an α-cut, as shown in Figure 4, and is defined as a crisp set (not a fuzzy set) of all membership grades larger than alpha: $A_\alpha = \{x|\mu_A(x) \geq \alpha\}$.

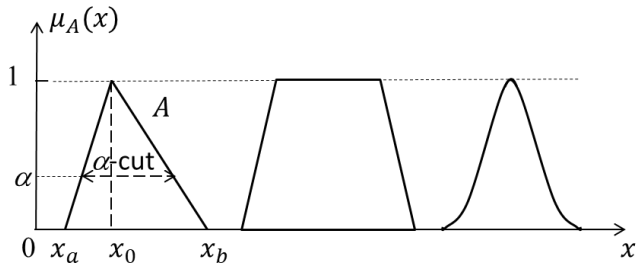


Figure 4. Fuzzy set shapes

It turned out that the approach could be applied to a large variety of problems in which a value, a type of object, a characteristic, or property could not be sharply defined but yet could be expressed as a fuzzy set, called fuzzy number, and processed mathematically. For example, in case of consequence severity and event frequency yielding risk or when the result of failing basic components in a parallel or series configurations must be determined, arithmetic operations on the fuzzy numbers shall be performed. Taking again the fault tree as in the previous section on DST, and calculating the failure probability of the system of two parallel basic components (AND-gate) the fuzzy numbers of experts 1 and 2 (Table 7) must first be averaged (Table 8) and then multiplied (Table 9). In principle, the latter means multiplying the membership values at all α-cut levels, which boils down to multiplying the values of x_a, $x_{0,}$ and x_b of the respective input fuzzy sets, if these have similar, e.g., triangular shape. Results are depicted in Figures 5a and b. In case of an OR-gate both fuzzy numbers must be added and their product subtracted.

Table 7. Input values of two experts of estimated failure probabilities of components A and B

	Expert 1			Expert 2		
	x_a	x_0	x_b	x_a	x_0	x_b
Comp. A	2E-03	3.00E-03	5E-03	1E-03	3.00E-03	4E-03
Comp. B	3E-03	3.50E-03	4E-03	2E-03	4.00E-03	5E-03

Table 8. Averaged input of failure probabilities

	x_a	x_0	x_b
Comp. A	1.50E-03	3.00E-03	4.50E-03
Comp. B	2.50E-03	3.75E-03	4.50E-03

Table 9. Resulting fuzzy number failure probability top-event in AND-gate configuration of components A and B

x_a	x_0	x_b
3.75E-06	1.13E-05	2.03E-05

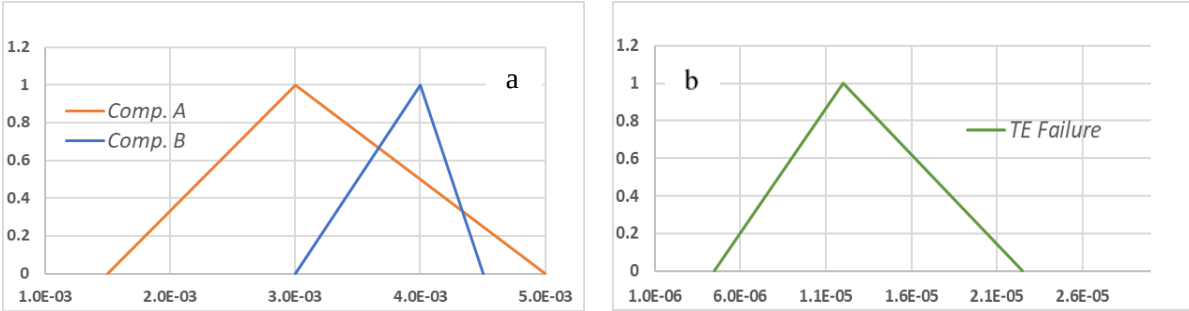


Figure 5a and b. *Left*, the averaged fuzzy sets of the two components A and B, and *right*, the resulting top event fuzzy number in case of AND-gate configuration of the two.

If in contrast inference is required of one or more fuzzy sets on the same or on different universes, rule-based logic must be applied to derive the resulting shape of the inferred fuzzy system or function, *e.g.*, in its simplest form:

IF $x = A$ THEN $z = C$; or more extended: IF $x = A$ AND $y = B$ THEN $z = D$,

where the first is a single-antecedent−single-consequent rule, and the more common second is a two-antecedent−single-consequent rule. The AND connection can also be OR or NOT. Antecedents and consequent can be numeric or also frequently linguistic. If antecedent or their combination is numeric, a numeric consequent can be implied if the result cannot be determined arithmetically. Obviously, the arguments have a conditional relation. If the consequent will be a fuzzy set, the inference procedure has been developed by Mamdani [23], and if it is a function of x and y, or a constant, Sugeno, or more precisely Takagi, Sugeno and Kang [24, 25] inference shall be applied. The Sugeno inference operation increased the applicability of fuzzy sets greatly, *e.g.*, by enabling improved control systems (robots).

Given two AND related antecedent fuzzy sets at the same universe, their intersection/conjunction produces the minimum area they have in common; if their relation is an OR union/disjunction yields the area they have not in common, hence the complement. It shows the min and max principle. Of practical use, *e.g.*, in control, is the case of antecedents not necessarily at the same universe. If a value of both antecedents must be combined to infer a consequent value being a fuzzy set (Mamdani), the smallest antecedent membership grade is implemented as the membership grade of the given consequent (so called t-norm operator) producing the firing level of the rule. The result is the remains of the consequent set below the α-cut at the firing level and the sought value is the centroid of the remaining part, obtained by defuzzification. By defuzzification of the resulting fuzzy set the arithmetic mean position of all the points, the centroid, or any other central measure is computed. In case two rules apply the result can consist of an amalgam of different consequents of which a centroid can be calculated (see Figure 6). Looping through the whole of antecedents in discrete steps is also possible and will construct a complete consequent. Often, inputs are provided in linguistic terms, such as high medium low or good medium bad, or a more extended range of terms. This kind of input but also numeric ones can lead to arrays of fuzzy sets to cover a range (see for an example Figure 7); the finer the higher the ‘granulation’. In contrast, for Sugeno inference the consequent in the form of a function or constant is multiplied with the firing level and no defuzzification is needed. In case of an OR-relation the maximum (t-conorm) shall be obtained, so the largest membership grade is selected.

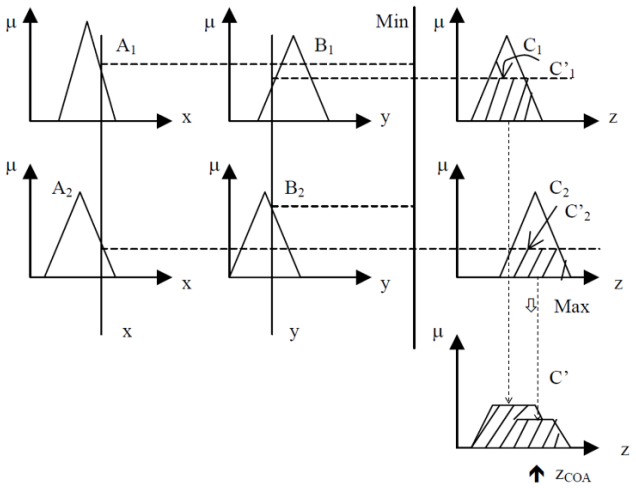


Figure 6. Two rule, two antecedent−single consequent in AND configuration with Mamdani inference as, *e.g.*, occurs in a control situation. Indicated x and y set values 'translate' into z_{coa}, the centroid of area, slightly modified after Castillo and Melin [29].

The above description is a bare minimum in which mathematical equations have been avoided. Wierman [26] or Mendel [22] present details and derive the many equations, which for extended cases tend to become rather complicated. Mendel [22] also suggests software that can be used within Matlab®, which itself has an app for rule-based logic of Type-1 fuzzy sets.

Applications in risk assessment are several, *e.g.*, Markowski et al. [27, 28]. The membership function of values of various types of variables to be used in the assessment, such as consequence severity or event frequency including human factor influences; the latter in many situations must be estimated by experts as proper and sufficient observations are lacking.

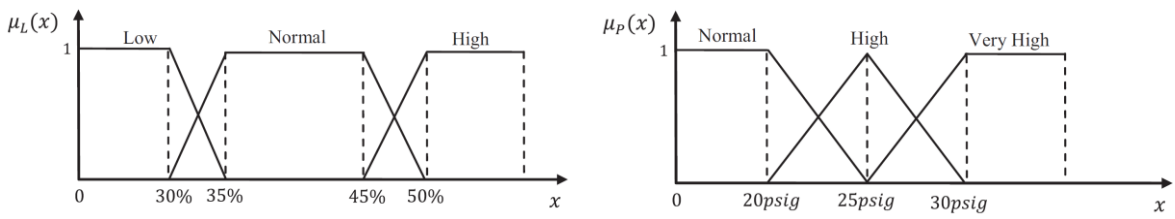


Figure 7a and b. Two example membership function arrays of reactor liquid level (*left*) and reactor pressure (*right*), both based on interviewing operators in order to obtain early warning of a possible run-away through an automated system applying Bayesian network, after Naderpour et al. [30]. Note the shoulder fuzzy sets at both ends.

A limitation of Type-1 fuzzy set is that different responses of the experts must be averaged. The authority of or confidence in experts can be expressed by weight factors, but expert weighting is often not done. Due to all this, uncertainty in inputs will be smaller than when responses are included in full, and uncertainty in the output will be lost when after defuzzification only the crisp value is retained. In part, these limitations can be avoided with Type-2 fuzzy sets.

Type-2 fuzzy sets

After his 1965 announcement of the development of fuzzy sets, Zadeh [7] also proposed in 1975 the concept of Type-2 fuzzy sets. The essence of the concept is that the membership function itself is fuzzy. This opens the possibility of improvement, because the uncertainty in

various inputs can be retained and are not in part reduced by averaging. Development of practical implementation of the idea took a long time, see Mendel [22] and his many IEEE articles. Therefore, during the last decade Type-2 fuzzy set quickly became popular as a tool to support multi-criteria decision making. Operations with Type-2 fuzzy sets are not simple. In fact, each input of a Type-1 fuzzy set contains in 2-dimensional form the uncertainty an estimating expert perceives about the subject. In Type-2 fuzzy set an additional uncertainty is added, so the aggregated uncertainty becomes 3-dimensional in representation and evaluation becomes rather intricate. As an in-between concept Interval Type-2 fuzzy set (IT2 FS) has been developed, which allows uncertainty to be modeled 2-D again, because over intervals there is no preference for any value, hence an interval will be covered by a uniform distribution. To represent this uncertainty mean and standard deviation of the uniform distribution are equated to those of a T1 FS triangle. Below it will be explained in more detail how that works. Fortunately, the IT2 FS requirement of estimates to be formulated as intervals $[a, b]$, provides in fact the most convenient way experts can make estimates. In Figure 8a and 8b is shown how individual responses as T1 FSs are wrapped in an IT2 FS upper and lower 'membership function'.

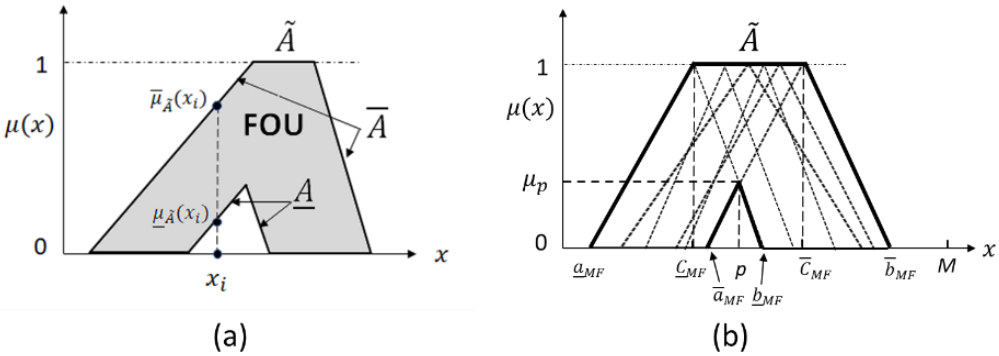


Figure 8a and b. IT2 FS $\tilde{A}$ in case of primary triangular T1 FSs; (a) showing the footprint of uncertainty (FOU) area and maximum and minimum envelope borders $\overline{A}$ and $\underline{A}$; (b) showing a few embedded T1 FSs and all parameter symbols involved; M means here the maximum bound on the value of x (after Liu and Mendel, 2008).

The way this concept can be used in risk assessments will be shown by some examples.

Example 1. Estimating failure rate values.
Although an estimate can be made directly in numerical values, it is often given in linguistic grades, such as '*very high, high, medium, low, very low*' or even wider. Anyhow, in such case the expert must indicate as well on a continuous, usually logarithmic numerical scale (Figure 9) what is meant by these terms.

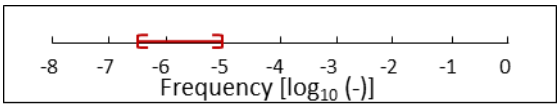


Figure 9. Example of a logarithmic scale of probability values per year (frequency) of an interval [-6.5, -5] or rather [0.32E-06, 1E-05] per year according to the expert equivalent to '*very low*'.

For creating interval fuzzy set the expert generated intervals $[a, b]$ must be converted first into symmetric triangular fuzzy sets. The principle is to equate the mean, $m_{MF} = (a_{MF} + b_{MF})/2$,

and standard deviation, $\sigma_{MF} = (b_{MF} - a_{MF})/2\sqrt{6}$, of a T1 FS to the mean, $(a+b)/2$, and standard deviation, $(b-a)/2\sqrt{3}$, of the uniform distribution over the interval (Liu and Mendel [31]). This yields: $a_{MF} = ½(a+b) - (b-a)/\sqrt{2}$ and $b_{MF} = ½(a+b) + (b-a)/\sqrt{2}$. In case of left and right shoulder fuzzy sets, slightly different equations apply.

Table 10 Responses on a pressure relief valve of three experts and the transformed values

Per 10^7 hours		**Expert 1**		**Expert 2**		**Expert 3**	
a, min	**b, max**	**2.0**	**7.0**	**5.0**	**10.0**	**3.0**	**8.0**
a_{MF}	b_{MF}	1.0	8.0	4.0	11.0	2.0	9.0
m_{MF}	s_{MF}	4.5	1.4	7.5	1.4	2.0	9.0
[a, b] in fits		200	700	500	1000	300	800

Suppose we interview three experts about the failure rate of a particular pressure relief valve. Their responses of failure per 10^7 hours (in fits ×100) are summarized in Table 10 together with the transformation and aggregation to obtain the FOU. For the aggregation, the maximum and minimum a_{MF} and b_{MF}-values of the three experts are selected. Referring to Figure 8b for the symbols, the C_{MF}-values are obtained as $\underline{C}_{MF} = (\underline{a}_{MF} + \underline{b}_{MF})/2$ and $\overline{C}_{MF} = (\overline{a}_{MF} + \overline{b}_{MF})/2$, while p and μ_p follow from: $p = \frac{\underline{b}_{MF}(\bar{C}_{MF}-\bar{a}_{MF})+\bar{a}_{MF}(\underline{b}_{MF}-\underline{C}_{MF})}{(\bar{C}_{MF}-\bar{a}_{MF})+(\underline{b}_{MF}-\underline{C}_{MF})}$ and $\mu_p = \frac{\underline{b}_{MF}-p}{\underline{b}_{MF}-\underline{C}_{MF}}$. This way the full uncertainty implied by the responses is embodied in the FOU. It means that the broader the base of the IT 2FS and the smaller the lower FOU envelope the larger the uncertainty.

Applying the procedure to construct the IT2 FS yields Figure 10. The centroid bounds have been computed according to the Karnik-Mendel method (Mendel [22]) and are c_l = 3.6 and c_r = 8.2 with a mean C of 5.88.

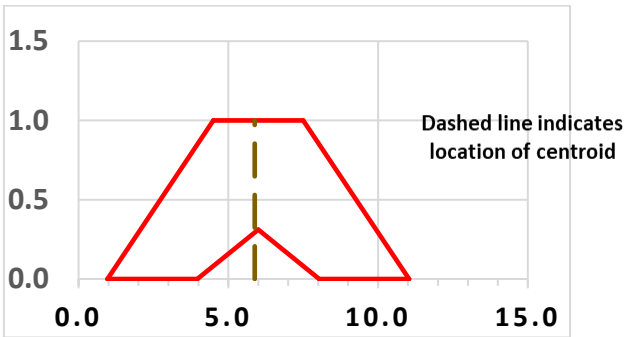


Figure 10. IT2 FS constructed based on the responses of three experts.

Just for comparison, if by means of a Bayesian network the three uniform distributions of the data are convoluted a mean of 5.83 is found with a standard deviation σ of 0.823 (hence 6σ, which covers 99.7% of a normal distribution, would have as bounds 3.4 and 8.3)

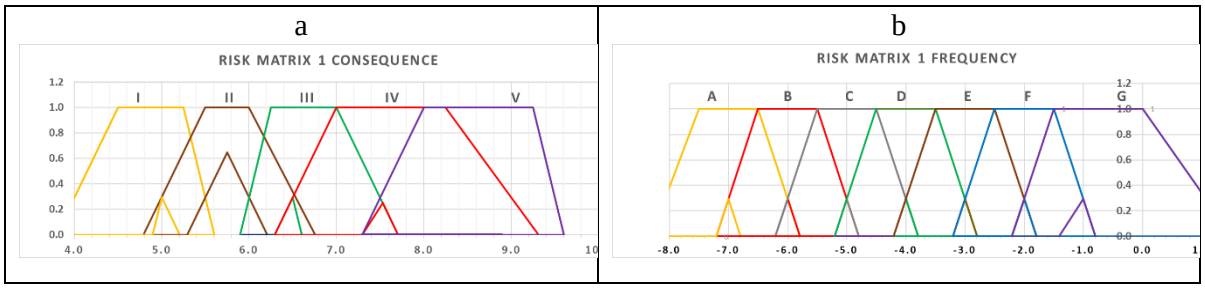
Tables 11a and b. Mock-up response values of three experts

$\log_{10}$ scale **Consequence [$]**	Expert 1 a, min	Interval b, max	Expert 2 a, min	Interval b, max	Expert 3 a, min	Interval b, max
Negligible (I)	5.00	5.50	4.00	5.00	4.50	5.20
Low (II)	5.50	6.50	5.00	6.00	5.25	6.50
Moderate (III)	6.50	7.50	6.00	6.50	6.50	7.50
High (IV)	7.50	9.00	6.50	7.50	7.50	8.50
Catastrophic (V)	9.00	9.50	7.50	8.50	8.50	9.20

$\log_{10}$ scale **Frequency [/yr]**	Expert 1 a, min	Interval b, max	Expert 2 a, min	Interval b, max	Expert 3 a, min	Interval b, max
Remote (A)	-8.00	-6.50	-8.00	-7.00	-7.00	-6.00
Unlikely (B)	-6.50	-5.50	-7.00	-6.00	-6.00	-5.00
Very Low (C)	-5.50	-4.50	-6.00	-5.00	-5.00	-4.00
Low (D)	-4.50	-3.50	-5.00	-4.00	-4.00	-3.00
Medium (E)	-3.50	-2.50	-4.00	-3.00	-3.00	-2.00
High (F)	-2.50	-1.50	-3.00	-2.00	-2.00	-1.00
Very High (G)	-1.50	1.00	-2.00	-1.00	-1.00	1.00

Example 2. Converting a linguistic risk matrix into a quantitative one

Interval Type-2 fuzzy set (IT2 FS) can also be used to convert linguistically graded consequence severity or event occurrence frequency into numerical risk. To that end experts must be asked to give estimates of linguistic terms as intervals on a continuous scale, which due to the wide range covered will be logarithmic. Another more complex case is the conversion of a linguistic risk matrix with given acceptance criteria (acceptable, tolerable acceptable, tolerable not acceptable, not acceptable) into a quantitative one[4]. The procedure creates an array of severity and frequency IT2 FSs, which form the antecedents for a Sugeno inference mentioned earlier. We shall not go into details. For the inference special software has been developed. The inputs are collected in Tables 11a and b, the arrays of IT2 FSs are presented in Figures 11a and b, the original and converted three-dimensional matrix of this example is shown in Figures 12a and b and a projection of the 3-D on the ground plane in Table 12.



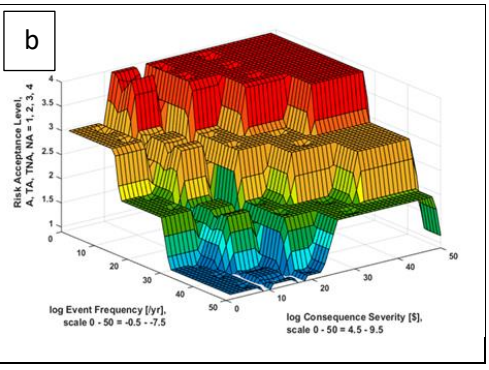
Figures 11a and b. IT2 FSs of consequence severity and event frequency of the three experts.

a

Frequency		Consequence I Negligible	II Low	III Moderate	IV High	V Catastrophic
G	Very high	TNA	NA	NA	NA	NA
F	High	TNA	TNA	NA	NA	NA
E	Medium	TA	TNA	TNA	NA	NA
D	Low	TA	TA	TNA	TNA	NA
C	Very low	A	TA	TA	TNA	TNA
B	Unlikely	A	A	TA	TA	TNA
A	Remote	A	A	A	TA	TA

NA: Not Acceptable; TNA: Tolerable Not Acceptable; TA: Tolerable Acceptable; A: Acceptable

b



Figures 12a and b. The original linguistic matrix and the converted three-dimensional one.

Table 12. Projection of the converted 3-D matrix on the X-Y plane

$\log_{10}$ Consequence [$]	4.50	5.00	5.50	6.00	6.50	7.00	7.50	8.00	8.50	9.00	9.50

[4] This example has been part of a paper submitted to Process Safety and Environmental Protection journal.

$\log_{10}$ Frequency [/yr]				Risk Acceptance Level:			1 = A; 2 = TA; 3 = TNA; 4 = NA				
-7.50	1.00	1.00	0.95	0.86	1.15	2.00	2.00	2.00	2.00	2.00	3.00
-6.80	1.00	1.00	1.00	1.72	2.00	2.00	2.00	3.00	3.00	3.00	3.00
-6.10	1.00	1.00	1.29	1.72	2.00	2.00	2.00	3.00	3.00	3.00	3.00
-5.40	1.00	1.15	1.74	1.72	2.00	3.00	2.99	3.00	3.00	3.00	3.00
-4.70	2.00	2.00	2.00	2.58	3.00	3.00	3.00	4.00	4.00	4.00	2.75
-4.00	2.00	2.00	2.36	2.58	3.00	3.15	3.15	4.00	4.00	4.00	2.45
-3.30	2.00	2.00	2.62	2.58	3.00	4.00	3.99	4.00	4.00	4.00	2.15
-2.60	3.00	3.00	3.00	3.43	4.00	4.00	4.00	4.00	4.00	4.00	2.00
-1.90	3.00	3.00	3.72	3.43	4.00	4.00	4.00	4.00	4.00	4.00	2.00
-1.20	3.00	3.00	3.81	3.43	4.00	4.00	4.00	4.00	4.00	4.00	2.00
-0.50	0.00	0.00	0.00	1.00	1.00	0.96	1.53	1.77	1.73	1.75	1.99

Because the experts provided intervals also at the ends of the ranges, judging the range beyond as unrealistic, there are no shoulder sets. Therefore, the risk acceptance level figures decrease at the edges.

4. Probabilistic approach of expert estimation

Expert elicitation applying a probabilistic approach has a long history. It had already started in the 1970s with trying to solve problems in nuclear risk assessment (Rasmussen [32]). Perhaps the latest is applying it in the field of climate change (Oppenheimer et al. [33]) claiming that *structured expert judgement* is applied "in order to facilitate characterization of uncertainty in a reproducible, consistent and transparent fashion." That is "experts quantify their uncertainty on potentially observable variables of interest", "and on calibration variables from their field whose true values are known post hoc" (seed variables). Hence, the elicitation is on numeric values. The example given is on atmospheric dispersion modeling. One of the co-authors of [33] is Cooke, [3], who is an expert in probabilistic approach to expert elicitation. Cooke calls his method the "classical model", because it is based on calibration measurement and classical statistical testing [34]. The principle is that the experts provide numerical answers at 50% and 5% probability. Oppenheimer et al. [33] summarize characteristics of expert elicitation as follows:

"(i) experts' statistical accuracy and informativeness (the ability to concentrate high probability in small regions) is very uneven, ranging from informative and statistically accurate to very overconfident;

(ii) both equal-weight and performance-based combinations of individual experts' distributions generally result in improved statistical accuracy, and for equal weighting this improved accuracy is often purchased at the expense of very wide confidence bands;

(iii) statistical accuracy and informativeness are often antagonistic - the most informative experts are also the least accurate - although many expert panels contain accurate and informative individuals; and

(iv) performance weighting yields better performance, both in- and out of- sample, than weighting schemes not based on performance."

The procedure described by Cooke and Goossens [35], here summarized and maybe oversimplified, is that experts are asked to provide estimates of ≤5%, >5% and ≤50%, >50% and ≤95%, and >95%, forming 4 inter-quantile intervals (or bins) as an impression of the perceived width of the distribution of answers. If desired, elicitation can be refined by asking more quantile fractions. Each expert will answer N questions. He/she will also be considered as a statistical hypothesis, so that the N realizations will produce for each expert a sample distribution $s(e)$, or rather for each interval, $s_i(e)$. If there is a distribution where $s_i(e)$ can

occur with probability p_i, the amount of conditional information[5] received is $NI(s_i(e)|p_i) = -Ns_i(e)\ln(s_i(e)/p_i)$, or the total amount is $NI(s(e)|p) = -N\sum_{i=1}^{4} s_i(e)\ln(s_i(e)/p_i)$.

With large N, $2NI(s(e)|p)$ becomes chi square (χ^2) distributed with 3 degrees of freedom. This can be seen when expanding the natural logarithm in the relation in a Taylor series and focusing on the dominant terms. The operation results in the χ^2-statistic, based on the summed squared differences of predicted and observed values, for goodness of fit of s to p holding for a 'well-calibrated' expert. If the statistic is low, the probability that the fit is good, is high. This probability provides an expert's *calibration score*. When the expert is asked to produce quantiles instead of numbers in a bin, a transformation of the answer is needed to allow for the non-equal masses, but the method remains basically the same.

Besides a calibration score of each expert, an *information score* is determined. Information scores measure the width of the distribution the expert presents. An expert's information score is the average relative information with respect to the background [34], *e.g.*, a uniform distribution of which a relevant so-called intrinsic range is considered. This range is taken such that it spans all quantile answers plus an overshoot. It is selected by the analyst, because it holds for all experts. Information score of expert e is: $(1/N)\sum_1^N I(f_{e,i}|g_i)$, in which N is as before, $f_{e,i}$ is the density of expert e for item i, and g_i is the background density for item i, all under the assumption that variables are independent. Information scores increase when the quantiles become narrower, as an expert can be supposed to be surer of his/her answer. However, if quantiles come very close it may indicate overconfidence.

The 'decision maker' (DM) combines the calibration and the information scores in linear pooling. To do this the best way causes again a number of intricacies, because experts can try to manipulate their quantile distances. The challenge is to find a set of weights, each weight being an expert's product of calibration and information score, such that the linear pool under these weights maximizes the product of calibration and information scores and optimizes performance. The best expert gets weight of 1. The choice is between a *global* weight decision maker (scores global over all seed variables) (GWDM) and an *item* weight (IWDM) discriminating per seed variable, which both should be better than the *equal* weight one (EWDM), which consists of simple averaging. GWDM is proportional to the product of the calibration and the information scores, given that the calibration exceeds a "significance level" cut-off α, below which the expert weight is set to zero. The significance cut-off α is determined by where it maximizes the product scores. In other words, good answers are rewarded and bad ones are eliminated. IWDM does the same but per questioning the item/seed variable. The performance optimized combination DMs often turned out to be better than EWDM [34].

The data base in [34] contains 45 panel runs on different topics and reports the DM performances. The results are obtained using the software EXCALIBUR (which appears not to

[5] This is according to the Shannon entropy theory. Information content I on randomness of random variable x_j representing an observable event j occurring with probability p_j is: $I(p_j) = -\log_b(p_j)$, or if b is taken as the mathematical constant e, base of the natural logarithm, $-\ln(p_j) = \ln(1/p_j)$. If base is e, information quantity is measured in the unit nats (but when $b = 2$ in bits). Derivation is by considering if two independent events occur, the information on the joint probability: $I(p_1p_2) = I(p_1) + I(p_2)$, hence in general, $I(p^a) = aI(p)$ analogous to $\log(p^a) = a\log(p)$. In N repeats, event j will on average occur Np_j times, so total $NI(p_j) = -\sum_1^n Np_j\ln(p_j)$. In case of conditional $I(s|p)$ to be understood as amount of randomness in s given p, so that $I(s|p) = -s\ln(s/p)$. If $s = p$, then $I = 0$, in other words the sample is exactly the value of the true distribution.

be downloadable anymore). Each study takes quite an effort (1 – 3 months). Further cases are described in [35]. Hanea [36] gives a detailed account of applying the method to the problem of escape time from a burning building applying Bayesian network to relate parameters and interviewing four experts each from a different domain: fire prevention, fire development, fire safety of buildings, and people behavior in evacuation. For calibration, 7 questions were asked on fire statistics and 10 on parameter values of interest, such as alarming system reliability, people flow through exits, waiting time, number of exits and distance to exits in public buildings. EWDM appeared to be slightly lower and less informative than the other two DMs. Hanea reports details, achievements, but also encountered problems of various nature. Bolger and Rowe [37] criticize the unequal weighting of the classical model for several reasons, *e.g.*, because of the influence on experts from a psychological behavior point of view and the usual limited set of seed variables, and they would prefer equal weighting. On the other hand, if bias by group think, polarization, failure to share information, dominance, and dogmatism can be reduced by experienced leadership, group discussions (behavioral aggregation) can also help to resolve the problem. This evoked commentary by Cooke [37] and a response by Bolger and Rowe [37], which altogether sharpens the contours of the problem area of applying expert elicitation.

The mentioned climate change study [33, 33A] focuses on Gaussian dispersion modeling and on the uncertainty induced by crosswind dispersion. Eleven experts have been interviewed on 36 calibration variables; the EWDM appeared in this case to be slightly higher than the performance-based ones, though. The combined uncertainty by the experts turned out to deviate significantly from those implied by 5 different choices of crosswind dispersion model parameters, and it therefore gained credibility. The result was used for probabilistic inversion to improve the model.

5. Similarities and differences in methods

The probabilistic approach again shows that when less uncertainty is desired, efforts to delimit uncertainty increase strongly. One does not get more value cheap! In the probabilistic approach with the attempts to 'calibrate' experts, the afore mentioned expert elicitation complexities appear quite pronounced. However, intrinsically in the Dempster-Shafer and fuzzy set approaches of expert judgment, the above observed problems of expert weighting will be present as well. In DST it is the analyst who by assigning mass gives experts a weight, but that weight also is subjective, while in fuzzy set expert weighting is hardly used. It all increases the uncertainty of the results produced by the methods .

It is not easy to make a choice between Dempster-Shafer method and fuzzy set, although for simpler questions regarding reliability and other observable data the more rigid mathematical treatment of information in the Dempster-Shafer approach makes it attractive and preferential over fuzzy set. In case of bowtie causal structure, thanks to the rigor, DST failure data can be propagated in an *Evidential network*, similar to a Bayesian network. For that, the epistemic uncertainty term is treated as an additional mode. Of course, fuzzy set Type-2 is a substantial improvement over Type-1, so it is worth the additional effort, but it remains approximative. Ferdous et al. in [39] and earlier papers compared for bowtie analysis the lack of failure data by having experts cast their subjective linguistic estimates in a Type-1 fuzzy set and in addition in a DST format showing inconsistency among experts and their fractions of ignorance.

Apart from the methods described, there are other less known ones. Much like DST, Credal sets and *Credal networks* try to accommodate imprecise and incomplete probability, see Cozman [40], or more recently Piatti et al. [41]. A credal set concerns a closed convex set[6] of probability measures that must sum to 1, but which can vary within that constraint. In a finitely generated credal set a number of probability value combination sets (each set summing to 1) exist that can be represented graphically by a polytope reflecting the sets of values by its vertices. Credal sets are often applied to binary variables because it can define upper and lower probability set functions, $\overline{P}(A)$ and $\underline{P}(A)$. It can be modeled by networks that are similar to Bayesian networks and Evidential ones but due to the constraint differ on a few fundamental aspects. In a given credal network, there are several credal sets called extensions that meet the constraints set by the network. Inference is possible. Only to a strong extension being still a joint credal set, which forms a joint density distribution over the variables X_i given parents pa as in a Bayesian network, $P(X) = \sum_i P(X_i|pa(X_i))$, allowing directional separation (so called d-separation), the Bayesian network mathematical infrastructure can be applied. Natural extension is the largest set complying with the network and must be solved in a different way (for which also software is available [42]).

In the fuzzy set Type-2 examples the expert was always asked to indicate on a continuous quantitative scale an interval what he/she thought the grading term covered, enabling the fusing with others. Herrera and Martinez [38] developed a method where the linguistic term can be retained and in which by adding a value between -0.5 and +0.5 in a fusing process loss or gain can be accounted for. It offers a solution in highly complex situations of interdependence, cascading, and indirect causation in which even an indication on a quantitative scale is not possible. The extent to which the grading term holds, appears from the value added.

Zadeh [43] is also pioneer of possibility theory. A quotation from [43] is: “A thesis advanced in this paper is that the imprecision that is intrinsic in natural languages is, in the main, possibilistic rather than probabilistic in nature”; and later “when our main concern is with the meaning of information-rather than with its measure-the proper framework for information analysis is possibilistic rather than probabilistic in nature”. If something is possible it does not need to be probable. Possibility theory forms an interface with probability theory, DST, and fuzzy set. It all has been formulated mathematically (Zadeh [43] and Klir [8], [44]). A state of affairs and applications are presented by Dubois and Prade [45]. An application in risk assessment is to find bounds on probability predicted by possibilistic considerations. So far, the approach did not result in convincing applications offering more than we have seen earlier with DST and IT2 FS. The same is true for the 2011 Zadeh introduction of the Z-number [46]. This number consists of two elements, in which the first informs about the constraint of a real variable in terms of ‘about’, ‘close-to’, etc., and the second on the reliability of the first in terms of ‘sure’, ‘likely’ and such like. Thus, a typical expression is: ‘the driving time of College Station to Houston downtown is about 2 hrs, usually’. Mathematically, the constraint can be expressed as a fuzzy number and the reliability as a probability value. One recent fault tree application has been published by Yadi et al. [47] in case of elicitationof experts unsure of exact failure rate values as an alternative to methods mentioned before in this paper.

[6] A convex set is a set of elements from a vector space such that all the points on the straight line between any two points of the set are also contained in the set: $x = \lambda a + (1-\lambda)b$ for all λ from 0 to 1 (Watkins Th., www.sjsu.edu/faculty/watkins/convex.htm).

All the methods described in this paper have found their way into Artificial Intelligence.

6. Conclusions

Methods are available to objectivize imprecise and subjective estimates of mostly binary variable values that in principle are observable, but because of long observation lead times or other reasons of inaccessibility, must be obtained by interviewing experts and eliciting their opinions.

Most forceful on the experts and at the same time using most of statistical background knowledge is the so-called structured expert judgment using the “classical model”. However, the traditional method is effort intensive and therefore costly.

The Dempster-Shafer approach seems for the type of failure data or consequence model parameter value questions we encounter in risk assessment the best option, because epistemic uncertainty thereby obtains as a separate focal element its own more realistic place and can be included in bowtie and other analyses making use of Bayesian network.

In case values of consequences and rates are given in linguistic terms of natural language, fuzzy set is most suitable. When the effect of linguistic graded variables shall be combined through inference of fuzzy sets (type-1 or type-2), it is possible to derive a concluding linguistic graded result (“Computing with words or perceptions” [48]). The same is true if variables are expressed as index values. In contrast to type-1 fuzzy set where expert opinion differences, whether or not weighted, must be averaged, type-2 can retain the uncertainty introduced by the response differences. Processng is effort intensive, though. Additionally, when the variable can be expressed on a continuous numeric scale, such as for consequence severity or event frequency, experts can be asked to indicate an interval (interval type-2) facilitating processing. The result by inference or arithmetic operation will then be numerical

A few other, less often or maybe not yet often applied methods are mentioned, which in specific complex cases can increase considerably the quality of a solution to a problem.

References

1. OREDA, 2015. Offshore and Onshore Reliability Data, sixth ed., SINTEF NTNU. <https://www.oreda.com/>.
2. Saaty, Th.L. Fundamentals of Decision Making: the Analytic Hierarchy Process. RWS 3 Publications, Pittsburgh, PA, 2006.
3. Cooke, R.M., 1991. Experts in Uncertainty, Opinion and Subjective Probability in Science. Oxford University Press, New York - Oxford, ISBN 0-19-506465-8.
4. Dempster A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics 38 (2), 325–339.
5. Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press.
6. Zadeh L.A., 1965. Fuzzy sets, Information and Control 8, 338–353.
7. Zadeh L.A., 1975. The concept of a linguistic variable and its application to approximate reasoning. Information Sciences 8, 199–249.
8. Klir G.J., 1989. Is there more to Uncertainty than some Probability Theorists may have us believe, Int’l Journal of General System 15, 347-378.
9. Zadeh L.A., 1978. Fuzzy sets as a basis for the theory of possibility. Fuzzy Sets Syst. 1, 3–28 (reprinted 1999, Fuzzy Sets and Systems, 100 Supplement, 9-34) (1978).

10. Helton J.C., Johnson J.D., 2011. Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty, Reliability Engineering and System Safety 96, 1034-1052
11. Shafer G., 1990. Perspectives on the Theory and Practice of Belief Functions, International Journal of Approximate Reasoning (4) 323-362; and 1992. Rejoinders to Comments, same journal (6) 445-480.
12. Pearl J., 1990. Reasoning with Belief Functions: An Analysis of Compatibility, International Journal of Approximate Reasoning 4, 363-389.
13. Rakowsky U.K., 2007. Fundamentals of the Dempster-Shafer Theory and its Applications to Reliability Modeling, International Journal of Reliability, Quality and Safety Engineering, 14 (6) 579–601.
14. Yager R., 1987. On the Dempster-Shafer Framework and New Combination Rules, Information Sciences 41, 93-137.
15. Sentz K, and Ferson, S., 2002. Combination of Evidence in Dempster- Shafer Theory, SANDIA Report 2002-0835; http://prod.sandia.gov/techlib/access-control.cgi/2002/020835.pdf.
16. Simon C., Weber P., Evsukoff A., 2008 Bayesian networks inference algorithm to implement Dempster Shafer theory in reliability analysis, Reliability Engineering and System Safety 93, 950–96.
17. Khakzad N., 2019, System safety assessment under epistemic uncertainty: Using imprecise probabilities in Bayesian network, Safety Science 116, 149-160.
18. Simon C, Weber P., 2009. Evidential Networks for Reliability Analysis and Performance Evaluation of Systems With Imprecise Knowledge, IEEE Transactions on Reliability 58 (1), 69-87.
19. Ferson S., Kreinovich V., Ginzburg L., Myers D.S., and Sentz K., 2003. Constructing probability boxes and Dempster-Shafer structures, SAND2002-4015, Unlimited Release, www.osti.gov/servlets/purl/809606.
20. Curcurù G., Galante G.M., La Fata C.M., 2012. Epistemic uncertainty in fault tree analysis approached by the evidence theory. Journal of Loss Prevention in the Process Industries 25, 667–676.
21. Klir G.J., and Yuan B., 1995. Fuzzy sets and fuzzy logic: Theory and applications. Upper Saddle River: Prentice Hall.
22. Mendel J.M., 2017. Uncertain rule-based fuzzy logic systems: introduction and new directions. 2nd Edition, Springer, ISBN 978-3-319-51369-0.
23. Mamdani E.H., 1974. Applications of fuzzy algorithms for simple dynamic plant. Proceedings of the IEEE 12, 1585–1588.
24. Takagi T. and Sugeno M., 1985. Fuzzy identification of systems and its application to modeling and control. IEEE Transaction on Systems Man & Cybernetics 15, 116–132.
25. Sugeno M., and Kang G.T., 1988. Structure identification of fuzzy model. Fuzzy Sets and Systems 28 (1), 15–33.
26. Wierman M.J., 2010. An Introduction to the Mathematics of Uncertainty, University of Creighton, https://www.creighton.edu/fileadmin/user/CCAS/programs/fuzzy_math/docs/MOU.pdf.
27. Markowski,A.S., Mannan M.S., Bigoszewska A., 2009. Fuzzy logic for process safety analysis, Journal of Loss Prevention in the Process Industries 22, 695-702.
28. Markowski A.S., Mannan M.S., Kotynia (Bigoszewska) A., Siuta D., 2010. Uncertainty aspects in process safety analysis, Journal of Loss Prevention in the Process Industries 23, 446-454.
29. Castillo O., Melin P., 2008. Type-2 Fuzzy Logic: Theory and Applications, Studies in Fuzziness and Soft Computing, Volume 223, Springer, e-ISBN 978-3-540-76284-3, p.17.
30. Naderpour, M., Lu, J., Zhang, G., 2015. An abnormal situation modeling method to assist operators in safety-critical systems, Reliability Engineering and System Safety 133, 33–47.
31. Liu F., Mendel J.M., 2008. Encoding words into interval type-2 fuzzy sets using an interval approach. IEEE Transactions on Fuzzy Systems 16 (6), 1503-1521.
32. Rasmussen N. C. et al., 1975. Reactor Safety Study: An Assessment of Accident Risks in US Commercial Nuclear Power Plants WASH-1400 (NUREG75/014), (US Nuclear Regulatory Commission).

33. Oppenheimer M., Little Ch.M. and Cooke R.M., 2016. Expert judgement and uncertainty quantification for climate change, Nature Climate Change, 6, 445-451
DOI: 10.1038/NCLIMATE2959.
Supplementary information can be downloaded at https://images.nature.com/original/nature-assets/nclimate/journal/v6/n5/extref/nclimate2959-s1.pdf
34. Cooke R.M., Goossens L.L.H.J., 2008. TU Delft expert judgment data base, Reliability Engineering and System Safety 93, 657–674.
35. Goossens L.H.J., Cooke R.M., Hale A.R., Rodić-Wiersma Lj., 2008. Fifteen years of expert judgement at TUDelft, Safety Science 46, 234-244.
36. Hanea D.M., 2009. Human Risk of Fire: Building a decision support tool using Bayesian networks, Dissertation TUDelft, Wöhrmann Print Service ISBN: 978-90-5638-227-8.
37. Bolger F. and Rowe G., 2015. The Aggregation of Expert Judgment: Do Good Things Come to Those Who Weight? Risk Analysis 35 (1), 5-11; with a Commentary by Cooke R.M., idem, 12-15; and a Response by Bolger F. and Rowe G, 2015. There is Data, and then there is Data: Only Experimental Evidence will Determine the Utility of Differential Weighting of Expert Judgment, Risk Analysis 35 (1), 21-26.
38. Herrera F. and Martinez L., 2000. A 2-Tuple Fuzzy Linguistic Representation Model for Computing with Words, IEEE Transactions on Fuzzy Systems 8 (6), 746-752.
39. Ferdous R., Khan F., Sadiq R., Amyotte P., Veitch B., 2013. Analyzing system safety and risks under uncertainty using a bow-tie diagram: an innovative approach. Process Safety and Environmental Protection 91, 1–18.
40. Cozman F.G., 2000. Credal networks, Artificial Intelligence 120, 199-233.
41. Piatti A., Antonucci A. and Zaffalon M., 2010. Building Knowledge-based Systems by Credal Networks: a Tutorial, in: Baswell, A.R. (Ed.), Advances in Mathematics Research, vol. 11. Nova Science Publishers, New York. http://people.idsia.ch/~zaffalon/papers/2010nova-cns.pdf.
42. Cozman F. G. JavaBayes – user manual, http://www.cs.cmu.edu/~javabayes/ .
43. Zadeh L.A., 1977. Fuzzy Sets as a Basis for a Theory of Possibility, reprint Fuzzy Sets and Systems 100 Supplement (1999) 9-34.
44. Klir G.J., 1999. On fuzzy-set interpretation of possibility theory, Fuzzy Sets and Systems 108, 263-273.
45. Dubois D., Prade H., 2011. Possibility Theory and its Applications: Where Do we Stand? Mathware & Soft Computing Magazine. 18 (1), 18-30.
46. Zadeh L.A., 2011. A Note on Z-numbers, Information Sciences 181, 2923–2932.
47. Yazdi M., Hafezi P., Abbassi R., A methodology for enhancing the reliability of expert system applications in probabilistic risk assessment, Journal of Loss Prevention in the Process Industries 58, 51–59.
48. Zadeh L.A., 2002, From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions, International Journal of Applied Math and Computer Science, 12 (3) 307–324.