Fuzzy semi-quantitative approach for probability evaluation using Bow-Tie analysis

H. Abdo & J.-M. Flaus
CNRS, G-SCOP, University of Grenoble Alpes, Grenoble, France

F. Masse
INERIS, Parc Technologique, Verneuil-en-Halatte, France

ABSTRACT: The International Organization for Standardization (ISO) imposes the evaluation of the probability of accidents during risk analysis with the consideration of uncertainty. However, quantitative probability analysis can be too expensive and lead to unreliable estimation. This is due to imprecision and lack of data where unjustifiable assumptions should be added, while quantitative information is lost by using a qualitative probability analysis. This paper proposes a fuzzy semi-quantitative approach to address data uncertainties as an alternative for losing and adding information. A fuzzy-based approach is used for handling vagueness and imprecision in the input parameter frequencies. The application of this approach is demonstrated using the casestudy of a Loss of Containment Scenario (LOC) in a chemical facility.

1 INTRODUCTION

Probability analysis may be qualitative or quantitative depending on the circumstances (Leitch 2010). Data for performing a probability analysis is either from historical incident data or expert elicitations (Abdo & Flaus). Qualitative probability analysis uses a scale of qualitative expressions (low, medium, high, etc.) to describe an event’s probability or frequency. The advantage of the qualitative methodology is its simplicity of applying and understanding by the relevant personnel. Expert judgments represent an important source of data to apply this methodology which are subjective (verbal expression) in nature. This subjectivity represents a disadvantage when quantitative or more precise information is available. However, the quantitative approach uses a numerical scale with real values to describe the event frequencies. A disadvantage of the quantitative approach is that the imprecision and lack of such data can affect the quality of the analysis (Abdo & Flaus 2015). The quantitative methodology is too expensive and complex to be performed in terms of time and cost since statistical and empirical data are needed. It can lead to probability underestimation if uncertainty is not taken into consideration (Abdo & Flaus 2016). For these disadvantages, a semi-quantitative approach represents a better alternative based on the available information.

INERIS (the French National Institute for Industrial Environment and Risks) has developed an interval semi-quantitative Bow-Tie analysis to quantify the probability of risk based on the available information (De Dianous, Deust, Bouissou, Farret, & Chaumette 2007). This approach is mainly based on the INERIS expertise and the results of the European project ARAMIS ((Hourtolou & Salvi 2003); (Hourtolou & Salvi 2004)). It uses historical accident data or expert elicitations if the former is not available. It is easy to use, effective and implicitly takes uncertainty into consideration.

This study highlights the limits presented in the interval semi-quantitative approach at first. Then it handles these limits by introducing the concept of fuzzy numbers instead of intervals. Fuzzy numbers are used to represent subjectivity in expert judgments and consider the quantitative data if it exists. The rest of this section provides an introduction to Bow-Tie analysis and the interval semi-quantitative approach as a preface to the proposed approach.

1.1 Bow-Tie analysis

Bow-Tie analysis is a very prominent method to identify and analyze the probability of risk (Ferdous, Khan, Sadiq, Amyotte, & Veitch 2012). It presents a combination between Fault Tree Analysis (FTA) and Event Tree Analysis (ETA). FTA and ETA respectively describe the relationship between the undesirable event, its causes and its consequences for a systematic representation of
hazard ((Sadiq, Saint-Martin, & Kleiner 2008); (Ferdous, Khan, Sadiq, Amyotte, & Veitch 2009)). Figure 1 presents a schematic diagram of the Bow-Tie analysis, the definition of each term is detailed in Table 1.

Evaluating probability of an accident using Bow-Tie analysis is performed by: (i) characterizing input data, (ii) propagating these characterizations through the Bow-Tie analysis. Characterizing input data aims to qualitatively, quantitatively or semi-quantitatively represent the information provided either by experts or derived from historical data with the consideration of uncertainty. Propagating the characterizations through the Bow-Tie aims to calculate the probability of ERC, PhDs and AMs. This is done by solving the AND and OR gates, considering the existence of risk barriers and the occurrence of secondary events. The next section presents a brief details on the semi-quantitative approach used by INERIS to perform a probability analysis.

1.2 Interval semi-quantitative approach

Before starting, it should be noted that there is a difference between probability and frequency. But in a simplified manner, these two notions of probability and frequency coincide when frequencies are low (less than 1 time every 10 years). The semi-quantitative scale used by INERIS to characterize the frequency of basic events is presented in Table 2. A class of frequency in terms of an interval is given to each basic event as input. This class is derived from experts, or by translating the quantitative data into a class (e.g. an event with a frequency equals $4 \times 10^{-4}$ is of class F3). This translation is performed based on Equation 1 below:

$$\text{Class}(f) = -\text{Ent}(\log(f)) - 1.$$  

where, $\text{Ent}$ is the integer part, log is the logarithm base 10 and $f$ is the frequency value of the event.

The propagation rules are defined and detailed in (INERIS 2015), and will be presented with additions to suit the fuzzy approach in section 3.4.

The second part of the paper highlights the limitations in the interval semi-quantitative approach. The third part examines the proposed methodologies to deal with the limitations of this approach. Section four presents a case study and compares the fuzzy with the interval semi-quantitative approach and the traditional quantitative approach. Finally, a number of conclusions are drawn in Section 5.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Signification</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>EI</td>
<td>Basic event</td>
<td>Direct cause of a loss of containment or physical integrity</td>
</tr>
<tr>
<td>E</td>
<td>Event</td>
<td>Physical integrity caused by the occurrence basic events</td>
</tr>
<tr>
<td>ERC</td>
<td>Undesirable event</td>
<td>The unwanted event such as a loss of containment, etc.</td>
</tr>
<tr>
<td>ES</td>
<td>Secondary event</td>
<td>Characterize the source term of an accident, such as ignition</td>
</tr>
<tr>
<td>PhD</td>
<td>Dangerous phenomenon</td>
<td>Physical phenomena that can cause major accidents, explosion, dispersion, fire</td>
</tr>
<tr>
<td>AM</td>
<td>Major accident</td>
<td>Damages caused by the effects of a PhD on people, environment or goods</td>
</tr>
<tr>
<td>Risk barrier</td>
<td>Measures taken place to reduce the probability of undesirable event and the effects of accidents</td>
<td></td>
</tr>
<tr>
<td>AND, OR gates</td>
<td>Describe the relationships between events</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Abbreviations, significations and definitions of elements listed in the Bow-Tie diagram.

Table 2. Determining the frequency classes based on the semi-quantitative approach.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-2</td>
<td>$10^{1.5} \text{year} \leq \text{frequency} \leq 10^{2.5} \text{year}$</td>
</tr>
<tr>
<td>F-1</td>
<td>$10^{0.5} \text{year} \leq \text{frequency} \leq 10^{1.5} \text{year}$</td>
</tr>
<tr>
<td>F0</td>
<td>$10^{0} \text{year} \leq \text{frequency} \leq 10^{1} \text{year}$</td>
</tr>
<tr>
<td>F1</td>
<td>$10^{1.5} \text{year} \leq \text{frequency} \leq 10^{2.5} \text{year}$</td>
</tr>
<tr>
<td>FX</td>
<td>$10^{2} \text{year} \leq \text{frequency}$</td>
</tr>
</tbody>
</table>

Figure 1. Elements of a “Bow-Tie” diagram.
2 PROBLEM STATEMENT

We shall now introduce the limitations presented in the interval semi-quantitative approach and discuss the methodology behind the proposed solution. These limitations are presented by converting the statistical data into classes:

- The discreteness of the frequency classes makes the conditions on the border between two intervals not well defined. Vagueness on the extent of half the range of the interval to which category it belongs is presented. The same class is given to different frequencies even if the difference between them is too remarkable (see Figure 2(a), the same class \( F - 1 \) is given to events \( E_1 \) and \( E_2 \) where their frequencies are 11 and 99 per year respectively);

- The interval representation of frequency classes can lead to probability underestimation. Figure 2(b) presents an example of an OR gate of two inputs \( E_{11} \) and \( E_{12} \) with quantitative frequencies equal \( 9.5 \times 10^{-2} \) and \( 9.6 \times 10^{-2} \) respectively. Suppose that these quantitative information are certain, the quantitative output frequency of the OR gate is equal to \( 1.05 \times 10^{-1} \). The translation of this quantitative output frequency using the interval scale results in a class \( F_0 \). While a frequency class equals \( F_1 \) is obtained based on the interval semi-quantitative approach. The output class of an OR gate based on the interval semi-quantitative approach is presented in Eq. 4;

- If we were wrong in determining the statistical data, i.e. an error factor may affect the statistical value. This error or deviation can lead to a different class and thus to a different result. Figure 2(c) shows an example of an OR gate with two input events \( E_{11} \) and \( E_{12} \). The occurrence frequencies of the both input events are equal to 9 times/year. Suppose we have an error factor due to the lack of information or a measurement error, and the frequencies can be higher by a factor of 2 times/year (e.g. 11 times/year). This deviation changes the class of events \( E_1 \) and \( E_2 \) from \( F - 1 \) to \( F - 2 \). The output of the OR gate is affected by this error where its class is changed. Thus, a small deviation can lead to a great change in the output.

However, a fuzzy-based approach is proposed in order to address the issues highlighted in this section.

Figure 2. Limitations presented in the traditional semi-quantitative approach.
3 METHODOLOGY

In this section, we present the added improvements on the INERIS approach. The characterization of input data in terms of fuzzy numbers and the propagation rules to calculate the occurrence probabilities of ERC and outcomes are discussed in the rest of this section.

3.1 Preliminary

Fuzzy set was introduced by the pioneer Zadeh in 1965 as a tool to characterize imprecise variables as well as to represent experts' knowledge in a mathematical tool (Zadeh 1965). Fuzzy variable is associated with a possibility distribution or membership function in the same manner as random variable is associated with a probability distribution. Consider a fuzzy subset $F$ of the universal set $X$, thus the membership function is $\mu_F : X \rightarrow [0,1]$. Where, for a given $x \in X$, the membership degree $\mu_{F,x}(x)$ represents the degree of compatibility of the value $x$ with the concept expressed by $F$. Triangular fuzzy number is the most popular which can be expressed as a triplet $[a, b, c]$, see (Abdo & Flaus 2016) for more details.

3.2 Define event frequency using fuzzy numbers

Fuzzy numbers are used to express the linguistic frequencies as shown in Figure 3. Since the frequency classes follow a logarithmic scale, the grades $FX, X \in \mathbb{N}$ are not triangular fuzzy numbers. $FX$ is divided in three parts (see Figure 4) and derived with its membership function using Equations 2 and 3, respectively.

$$FX = [a, b, c, d, e] = \left[ \frac{10^{-\frac{x-a}{b-a}} + 10^{-\frac{x-c}{b-c}}}{2}, \frac{10^{-\frac{x-c}{e-d}} + 10^{-\frac{x-d}{e-d}}}{2} \right]$$

(2)

Each value on this scale has its own possibility degrees to which classes it belongs. In the next section, we will define how to represent each type of input data based on the fuzzy semi-quantitative approach.

3.3 Representation (fuzzification) of input data

Statistical accident data and expert elicitations are both used in this approach. Thus, precise values and verbal expressions represent the input of the analysis. These inputs are fuzzified and represented as follows:

- The crisp or precise values derived from statistical data are mapped on the universe of discourse. This fuzzification process gives two membership degrees to each crisp value. Figure 3 shows an example of event A that belongs to classes F3 and F2 with membership degrees equal 0.6 and 0.4 respectively. It should be noted that uncertainties attached to statistical values are considered to be covered by the fuzzy classes.

- Experts are asked to give a verbal expression in terms of a frequency class to such an event if statistical data is not provided. This class is taken to be the input for the event.

3.4 Propagating fuzzy frequencies through the Bow-Tie

This section aims to set the fuzzy rules for propagating the input frequencies through the Bow-Tie analysis. Propagating inputs is achieved by solving the gates between the events, and aggregating the frequencies of risk barriers and the secondary events. The Fuzzy semi-quantitative rules are presented as follows:

- Treatment of OR and AND gates (section 3.4.1 and 3.4.2 respectively);
- Treatment of secondary events (section 3.4.3);
- Treatment of security barriers (section 3.4.4).

3.4.1 Treatment of OR gate

OR gate signifies that the output event occurs if either of the input events has occurred. Figure 5(a)
presents an OR gate with two input events $A$ and $B$. The output event $E$ occurs after the occurrence of $A$ or $B$. Based on the fuzzy approach, each one of $A$ and $B$ may be attached to one or two frequency classes depending on the type of input data (expert or statistical data). The frequency of the OR gate is calculated using the Cartesian product where the Class and the possibility degree of the table cases are calculated based on the two equations below:

$$\text{Class}(E) = \min[\text{Class}(E_1), \text{Class}(E_2)]$$ (4)

$$\text{Degree(Class}(E)) = \min[\text{Degree(Class}(E_1), \text{Degree(Class}(E_2))]$$ (5)

For example, let $A$ and $B$ belong to $[F_3, F_2]$ and $[F_4, F_3]$ with possibility degrees equal $[0.4, 0.6]$ and $[0.5, 0.5]$ respectively. The Cartesian product for the output is presented in Table 5(b). The output frequency is $[F_3, F_2]$ with possibility degrees of $[0.5, 0.6]$.

3.4.2 Treatment of AND gate

AND gate signifies that the output event $E$ occurs if the input events occur simultaneously. The classes and membership degrees of the output event are determined based on Eq. 6 and Eq. 7 respectively. An example is presented in Figure 8(a).

$$\text{Class}(E) = \max[\text{Class}(E_1), \text{Class}(E_2)]$$ (6)

$$\text{Degree(Class}(E)) = \max[\text{Degree(Class}(E_1), \text{Degree(Class}(E_2))]$$ (7)

It should be noted that, if the same class $F_j$ has different possibility degrees, then the maximum degree between them is selected for the $F_j$. As it is the case for class $F_3$ in the example, where 0.6 is taken as the output possibility degree.

3.4.3 Treatment of secondary events

PhD1 occurs if ERC occurs and ES occurs conditionally after ERC. PhD2 occurs after the
occurrence of ERC and no occurrence of ES. The input data here are the frequency classes of the ERC and conditional probability of ES (noted \( p \)). Output frequencies and degrees are calculated using the equations below:

\[
\text{Class}(\text{PhD}1) = \max\{\text{Class}(\text{ERC}) - \text{Ent}(\log(1 - p) - 0.5)\}; \text{see (INERIS 2015)}
\]

\[
\text{Degree}(\text{Class}(\text{PhD}1)) = \text{Degree}(\{\text{Class}(\text{ERC})\})
\]

\[
\text{Class}(\text{PhD}2) = \max\{\text{Class}(\text{ERC}) - \text{Ent}(\log(p) - 0.5)\}
\]

\[
\text{Degree}(\text{Class}(\text{PhD}2)) = \text{Degree}(\{\text{Class}(\text{ERC})\})
\]

Figure 7 presents an example.

3.4.4 Treatment of safety barriers

A security barrier operates after the occurrence of the event that this barrier is attached to. The output event \( E \) occurs if the input event \( E_i \) occurs and the attached security barrier does not operate. In the other case (if the security barriers operate), another event \( \bar{E} \) may occur.

The data needed for the treatment are the fuzzy frequencies of the event \( E_i \) and the confidence level of the security barrier. The fuzzy classes of the output events \( E \) and \( \bar{E} \) are calculated using the equations below:

\[
\text{Class}(E) = \text{Class}(E_i) + NC = F(x + NC); \text{(INERIS 2015)}
\]

\[
\text{Degree}(\text{Class}(E)) = \text{Degree}(\{\text{Class}(E_i)\})
\]

\[
\text{Class}(E) = \text{Class}(E_i)
\]

\[
\text{Degree}(\text{Class}(\bar{E})) = \text{Degree}(\{\text{Class}(E_i)\})
\]

Figure 7. Consideration of a ES within the fuzzy semi quantitative approach.

(a) The fuzzy output of dangerous phenomena after the occurrence of an ES

(b) The output fuzzy frequency of PhD1 and PhD2 obtained by means of the Cartesian product

Figure 8. Consideration of a security barrier within the fuzzy semi quantitative approach.

(a) The fuzzy outputs of events \( E \) and \( \bar{E} \)

(b) The output fuzzy frequency of \( E \) and \( \bar{E} \) obtained by means of the Cartesian product
4 HOW THE FUZZY APPROACH HANDLES THE EXISTED LIMITS

However, Figure 9 presents how the proposed methodology deals with the limitations presented in the interval semi-quantitative approach. For the discreteness issue, now the events E1 and E2 are now belong to different classes with different degrees (see Figure 9(a)). Figure 9(b) presents how the fuzzy approach solves the probability underestimation problem. Finally, a deviation due to error in the determination of input frequencies will not affect the result nor the decision. The same example taken in Section 2 is depicted in Figure 10. A small change in the possibility degrees is generated due to the deviation where the classes are the same. Here the effectiveness of fuzzy theory in handling uncertainty lies in.

This methodology will be illustrated in the next section and applied to a loss of containment scenario.

5 ILLUSTRATION OF THE PROPOSED APPROACH

In this section, a Bow-Tie analysis for a LOC scenario is utilized to prove the utility and effectiveness of the proposed methodology, shown in Figure 11. The proposed methodology is applied...
to the Bow-Tie to determine the frequencies of dangerous phenomena. The input frequencies were collected from two different data bases ((for Industrial Environment & (INERIS) 2008); (Center for External Safety & the Environment 2009)) and an expert in the field. These frequencies are translated into fuzzy classes (the dashed rectangle beside each event or risk barrier). These fuzzy classes are propagated through the Bow-Tie using the fuzzy rules set in Section 3.4. The output fuzzy frequencies of the ERC and outcomes are written on the Bow-Tie (the red dashed rectangles in Figure 11).

In addition to the proposed approach, quantitative and interval semi-quantitative analyses were also performed for the Bow-Tie of LOC scenario. In order to compare these approaches, the outputs are presented in Table 3. The output frequencies of the quantitative approach are translated into classes for comparison purposes. The quantitative approach is the most precise but it does not consider uncertainty in the analysis. As there is no consideration of uncertainty, this may lead to risk underestimation in some cases. However, the fuzzy approach presents more accuracy than the interval approach where the output fuzzy classes cover the class obtained using the quantitative approach. Fuzzy approach is more conservative than the quantitative approach as uncertainty is considered (fuzzy numbers are used instead of crisp values). Again the exact result from the quantitative approach lies within the result obtained by the proposed approach, which makes the later more conservative.

Table 3. Estimated results obtained using quantitative, interval semi-quantitative and fuzzy semi quantitative approaches.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Quantitative approach</th>
<th>Interval semi-quantitative approach</th>
<th>Fuzzy semi quantitative approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spreading solvents in a large quantity</td>
<td>$2.1 \times 10^{-4}$ : F2</td>
<td>F3</td>
<td>F3</td>
</tr>
<tr>
<td>Explosion UVCE</td>
<td>$2.1 \times 10^{-4}$ : F3</td>
<td>F4</td>
<td>1.0 in F4</td>
</tr>
<tr>
<td>Formation of toxic gas</td>
<td>$2.1 \times 10^{-4}$ : F3</td>
<td>F3</td>
<td>0.55 in F3</td>
</tr>
<tr>
<td>Pool fire</td>
<td>$2.1 \times 10^{-4}$ : F3</td>
<td>F4</td>
<td>1.0 in F4</td>
</tr>
</tbody>
</table>

6 CONCLUSION

Probability analysis of dangerous phenomena has become a necessary step in risk analysis. Qualitative or quantitative probability analysis can be performed depending on the type of data available. This data is derived from different sources (historical accident data or expert judgments in terms of numerical values or linguistic variables, respectively). Quantitative information for a quantitative analysis is not always provided. Qualitative analysis is subjective and may lead to loss of quantitative information if it exists. In addition, the accuracy of the analysis based on these approaches still a
major issue since uncertainty is not taken into consideration.

That is why this paper proposes a fuzzy-semi quantitative approach relying on the available information from historical data or experts if the former is not available. Fuzzy theory is introduced to handle uncertainty due to imprecision and vagueness in defining the frequency scale. Fuzzy rules are set to propagate the fuzzy input classes through the Bow-Tie analysis. BT analysis is used to model a risk scenario as it gives an exhaustive visual summary of risk by displaying the risk events and the control measures and as a tool for likelihood evaluation.

This methodology is applied to a Bow-Tie case study for a LOC scenario. A comparison with the quantitative and the interval semi-quantitative approaches is discussed. The results show that the proposed methodology provides more simplicity and accuracy in the quantification, in addition to the consideration of uncertainty.

In the future, this work will be extended by using multiple sources of data in probability analysis. Different data bases or experts may provide different probabilities regarding the same parameter. Thus, rating and aggregating the data from different sources will lead to a more robust probability quantification approach for Bow-Tie analysis.

ACKNOWLEDGMENTS

This work is based on research supported and funded by the French National Institute for Industrial Environment and Risks (INERIS).

REFERENCES


