On the assessment of uncertainty in risk diagrams

**Reference:**
Goerlandt Floris, Reniers Genserik.- On the assessment of uncertainty in risk diagrams
Safety science - ISSN 0925-7535 - Amsterdam, Elsevier science bv, 84(2016), p. 67-77
Full text (Publisher's DOI): https://doi.org/10.1016/J.SSCI.2015.12.001
To cite this reference: https://hdl.handle.net/10067/1297210151162165141

# On the assessment of uncertainty in risk diagrams

Floris Goerlandt [a,b*], Genserik Reniers [b,c,d]

[a] Aalto University, School of Engineering, Department of Applied Mechanics, Marine Technology, Research Group on Maritime Risk and Safety, P.O. Box 15300, FI-00076 AALTO, Finland

[b] KU Leuven, Campus Brussels, Faculty of Economics and Organizational Sciences, Center for Economics and Corporate Sustainability, Warmoesberg 26, 1000 Brussels, Belgium

[c] TU Delft, Safety and Security Science Group, Faculty of Technology, Policy and Management, Jaffalaan 5, 2628 BX Delft, the Netherlands [d] UAntwerpen, City Campus, Faculty of Applied Economic Sciences, Prinsstraat 13, 2000 Antwerp, Belgium

* Corresponding author:
Tel.: +358 50 343 1186
e-mail address: floris.goerlandt@aalto.fi

## Keywords

Risk matrix, risk diagram, uncertainty, strength of knowledge, risk analysis

## Possible reviewers

- Nijs Jan Duijm, DTU, DK
- Roger Flage, UiS, NOR
- Pieter van Gelder, TU Delft, NL
- Faisal Khan, MUN, CAN
- Xiaobo Qu, Griffith, AUS

## Abstract

Risk matrices and risk diagrams are widely used tools for analyzing, assessing and visualizing risk in many industries, and are used extensively for risk management purposes. Despite their popularity and wide application, they have recently become the object of discussion and research in scientific environments, which can be seen as part of a wider focus on foundational issues in the risk analysis discipline. Identifying several serious limitations and problems with the risk matrix approach, various authors have proposed extensions, modifications and recommendations for their use. One issue which has been raised recently but has attracted relatively limited scientific attention is the consideration of uncertainty in risk diagrams, i.e. how to visually represent and communicate uncertainty. This paper first reviews the available proposals for this question. Subsequently, the

strengths and weaknesses of these proposals are discussed. Finally, some new proposals are made on how to represent uncertainty in risk diagrams in practical applications.

# 1. Introduction

Risk matrices (RMs) are widely used tools for analyzing, assessing and visualizing risk in many industries, and are used extensively for risk-management purposes. The main benefits attributed to RMs are their intuitive appeal and simplicity: they are perceived to be easy to construct, explain and score (Thomas et al., 2014). Belonging to the class of probability-consequence diagrams (PCDS) as described by Ale et al. (2015), they are easier to interpret than FN-curves[1]. Furthermore, RMs are recommended by various international standards and industry guidelines (IMO, 2007; IPIECA/OGP, 2013; ISO, 2010; NHS, 2008).

Notwithstanding its wide application, an increasing body of research has analyzed and discussed the limitations and inconsistencies of the RM approach. Duijm (2015) summarizes critical comments by Franks and Maddison (2006), Cox (2008), Smith et al. (2009), (2010), Ni et al. (2010), Flage and Røed (2012) and Levine (2012). Following issues are discussed: i) the consistency between the risk matrix and quantitative measures and the corresponding appropriateness of decisions based on risk matrices, ii) the subjective classification of consequence and probability, iii) the (linear or logarithmic) definition of risk scores and its relation to the scaling of the categories, iv) the limited resolution of risk matrices, resulting in "risk ties", the aggregation of scenarios and consequences for a single event on different areas of concern, and for multiple hazards originating from a single activity, and vi) problems with the use of corporate-wide risk matrix designs (Duijm, 2015). Similar points are made by Hubbard (2009), Kontovas and Psaraftis (2009), Pickering and Cowley (2010) and Thomas et al. (2014).

In response to these identified problems with RMs, several authors have proposed extensions to the approach. Markowski and Mannan (2008) propose the use of fuzzy sets to account for vagueness in the definition of the linguistic ordinal scales. Ni et al. (2010) propose a methodology based on the Borda count, using the likeliness and consequence ranks as independent scores, as well as other arithmetic extensions. Garvey (2009) and Meyer and Reniers (2013) discuss a method to adjust the categorization of the risk ranking, accounting for the decision-makers risk attitude (consequence- or likeliness averseness). Ruan et al. (2015) propose a method to account for decision-makers risk attitude based on the utility theory. Duijm (2015) provides a number of recommendations, including that the coloring should define risk as a monotonously increasing function of consequences and likeliness, the use of logarithmic scaling and the use of continuous PCDS instead of discrete categories, the benefits of which are also discussed by Ale et al. (2015). Duijm (2015) also identifies challenges to the use of continuous probability-consequence diagrams, one of which concerns how to assess uncertainty in the assigned probability and consequence metrics.

This last issue is the research topic of this paper. In particular, previously proposed methods for representing uncertainty in PCDS are summarized and their merits and shortcomings discussed. Subsequently, proposals are made to represent uncertainty in risk diagrams.

This issue strongly relates to risk communication: graphical displays focus attention and serve a special role in getting the right message across, not in the least because detailed analyses in lengthy reports may not always be fully read by decision makers (Abrahamsen et al., 2014). Hence, it is of

[1] An FN-curve shows the frequency of exceedance (F) of a given number of fatalities (N).

considerable importance to develop and present ideas to assess, visualize and communicate uncertainty in risk diagrams. The relevance of this research is also supported by Fischhoff (1995), who finds that uncertainties are not always appropriately conveyed in risk communication and by Spiegelhalter et al. (2011), who find that there has been rather little progress on the issue of representing uncertainty.

In the remainder of this paper, our focus is exclusively on continuous PCDS, i.e. qualitative risk matrices are beyond the scope. This limitation follows from arguments from Abrahamsen et al. (2014), Ale et al. (2015) and Duijm (2015) that these allow for a more accurate risk picture. Moreover, risk diagrams are here understood as tools for visualizing the risk picture, not as complete risk analysis tools, see Flage and Røed (2012) and Abrahamsen et al. (2014).

The rest of this paper is organized as follows. Section 2 gives a background for the need for assessing uncertainty, and introduces uncertainty-based risk perspectives. In Section 3, some earlier proposals for representing uncertainty in PCDS are outlined. Their elements, strengths and weaknesses are discussed in Section 4. Section 5 presents two new proposals for visualizing uncertainty in PCDS. Section 6 concludes.

# 2. Assessing uncertainty in PCDS: risk perspectives

## 2.1. Background and justification

Risk is often defined through probabilities, either as an expected value of probabilities and consequences (Campbell, 2005), or as the combination of scenarios, probabilities and consequences (Kaplan, 1997). Aven (2012) has made a historic analysis of the risk concept, finding that in many application areas, the predominant definitions are probability-based. This is confirmed in a recent review of definitions in risk analyses concerned with accidental risk in waterways. This study also shows that risk perspectives (systematic methods to describe risk) corresponding to probability-based definitions typically do not consider uncertainties beyond the probabilistic descriptions (Goerlandt and Montewka, 2015a).

The need for considering uncertainties in making scientific claims has been argued for by Douglas (2009) on grounds that scientists have a responsibility to consider the consequences of error. If evidence is poor and if this may lead to foreseeable changes to the conclusions of an inquiry, these uncertainties need to be made explicit[2]. The lack of uncertainty treatment is a relatively common criticism of especially quantitative risk analysis (QRA), e.g. Shrader-Frechette (1993), O'Brien (2000) and Aven (2011), and has been confirmed in e.g. the maritime transportation and offshore oil and gas applications areas (Goerlandt and Montewka, 2015a; Haugen and Vinnem, 2015).

Several authors have argued for perspectives where uncertainty is given a more prominent role than in traditional probability-based perspectives (Aven and Zio, 2011; Flage et al., 2014; Haugen and Vinnem, 2015; Montewka et al., 2014). The rationale of such perspectives is outlined next.

[2] This is a version of the classical "error argument", which is also known as the argument from inductive risk (Douglas, 2000; Rudner, 1953; Steel, 2010). It constitutes one of the primary reasons why science is not (as often thought), value-free: non-epistemic values (values which have no bearing on determining whether a claim is true but stem from a reflective consideration of what is good in a given context) are needed to consider the consequences of error and to identify which uncertainties are relevant to assess (Douglas, 2009).

## 2.2. Uncertainty-based risk perspectives

In uncertainty-based risk perspectives, it is acknowledged that probabilities are tools for describing/measuring risk, but that these tools are not sufficient to assess and communicate all possibly decision-relevant uncertainty. This is most easily understood by distinguishing two broad classes of uncertainty, as distinguished by Levin (2005): outcome and evidence uncertainty. This distinction is applied in this paper[3].

Levin (2005) defines outcome uncertainty as a cognitive attitude (i.e. a state of mind) of an assessor, who, at a given time, simultaneously holds mutually exclusive beliefs about the occurrence or non-occurrence of an event. This type of uncertainty can be measured using probabilities understood as a degree of belief (Singpurwalla, 2006; Watson, 1994). This in turn can be interpreted with reference to an urn standard (Aven and Reniers, 2013; Lindley, 2006): an assessor's the degree of belief about the occurrence of an event is compared with the standard of drawing at random a specific ball from an urn that contains a given number of balls.

In Levin's (2005) taxonomy, evidence uncertainty focuses on the poor or unreliable evidence base for making statements about, e.g. the occurrence or not of an event. Assumptions may be poor, models may be crude and data may be inaccurate or unreliable. Under such conditions, the results of an analysis can still support decision making, but the uncertainties can change the types of decisions made. E.g. in case of the presence of important uncertainties, decision makers may justifiably opt for additional protective measures. It is this type of uncertainty which is often not considered in applications, see Shrader-Frechette (1993), Aven (2011) and Goerlandt and Montewka (2015a). Nevertheless, the argument of inductive risk (see Section 2.1 and note 1), provides a strong reason for assessing evidential uncertainty as part of the risk perspective.

Several variations of uncertainty-based risk perspectives have been proposed. Flage and Aven (2009) focus on outcome and evidence uncertainty, which is qualitatively assessed, and include sensitivity in the risk description. Aven (2013) widens this risk perspective with black swans, i.e. surprises beyond the current knowledge, focuses on the strength of knowledge rather than evidential uncertainty[4] and uses an assumption deviation risk scheme as part of the risk description. Montewka et al. (2014) propose a risk perspective where knowledge and understanding are treated separately. Goerlandt et al. (2014) extend the perspective taken by Flage and Aven (2009) by also accounting for evidential biases. The uncertainty-based perspective as applied throughout this paper is summarized in Appendix A.

[3] Other taxonomies for uncertainty classification exist, e.g. distinguishing aleatory and epistemic uncertainty (Kiureghian and Ditlevsen, 2009), or endoxastic and metadoxastic uncertainty (Murphy et al., 2011).

[4] The authors agree with the view expressed in Abrahamsen et al. (2014) that the term 'strength-of-knowledge' may be less ambiguous than the focus on evidence uncertainty.

# 3. Earlier proposals for visualizing uncertainty in risk diagrams

Despite increasing awareness of the need to assess uncertainty in risk descriptions, there are relatively few practical proposals for how to visualize uncertainty in risk diagrams. The proposals found in the literature are briefly outlined below.

## 3.1. PCDS with family of "risk curves"

A rather well-known method to express uncertainty in risk diagrams is to create a family of "risk curves". In such approaches, alternative methods are applied to a risk model to quantitatively bound the space of probability-consequence combinations.

An early example of the idea of the family of "risk curves" is presented by Kaplan and Garrick (1981), see Figure 1. Starting from a probability-based risk perspective, which focuses on a frequentist probability[5] $P_f^*$ of the occurrence of an event A and related consequences C, an assessor describes his uncertainty about the "correct" value of the frequentist probability $P_f^*$ using a subjective probability[6] $P_s$. These subjective probabilities are subsequently propagated through the model-based analysis, either analytically or using Monte Carlo simulation. The objective is to obtain a probability distribution for the loss in case of consequence occurrence, such that for each risk curve represents, on the vertical axis, a fractile of the probability distribution of the frequency of exceedance of a given loss level. This approach of a family of "risk curves" corresponds to the highest level of uncertainty treatment in the work of Paté-Cornell (1996).

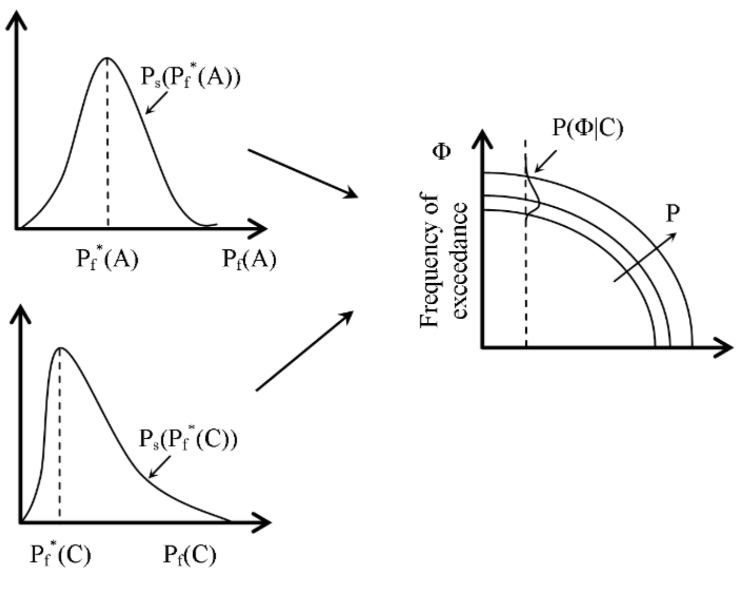


**Figure 1: Family of "risk curves" method for visualizing uncertainty in PCD, based on Kaplan and Garrick (1981)**

[5] A frequentist probability $P_f$ is defined as the fraction of time a specific outcome occurs in an in principle infinite number of repeated tests. Strictly speaking, a distinction needs to be made between $P_f$ as a concept and its measurement $P_f^*$, which is derived from empirical data, a thought-constructed "repeated experiment" or a repeated evaluation of an engineering or statistical model (Aven and Reniers, 2013; Watson, 1994).

[6] A subjective probability $P_s$ is defined as an assessor's degree of belief (e.g. about the occurrence, or not, of a given event). This probability is based on a given background knowledge, and can be interpreted with reference to an uncertainty standard (Aven and Reniers, 2013; Lindley, 2006).

Variations on this basic approach by Kaplan and Garrick (1981) have been proposed, notably through other proposed mathematical tools for propagating uncertainty through the model. These for instance include evidence theory, possibility theory and fuzzy set theory, see (Hayes, 2011; Helton and Johnson, 2011; Zio and Pedroni, 2013). These mathematical variations represent uncertainty through a set of quantitative curves in a very similar way as in Figure 1, and are for reasons of brevity not further described here.

## 3.2. PCDS with uncertainty boxes

Duijm (2015) proposes to use uncertainty boxes in probability-consequence diagrams, see Figure 2. It is not specified what type of probability these diagrams are based on (i.e. frequentist $P_f^*$ or subjective $P_s$). The idea is simple: a box represent an area in which the risk estimate is calculated/believed to be located. The vertical and horizontal lines indicate the expected value for the probability-consequence combination, whereas the outer limits of the boxes can correspond to prediction intervals[7,8].

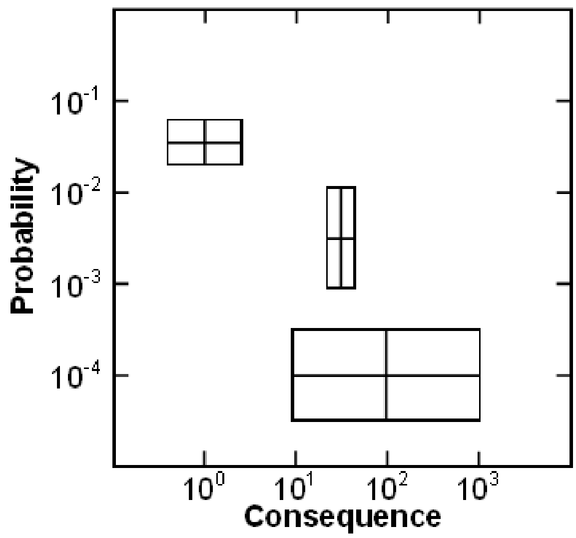


**Figure 2: Uncertainty boxes in PCD,**
**based on Duijm (2015)**

## 3.3. Bubble diagrams

Bubble diagrams have been presented in slightly modified formats by Abrahamsen and Aven (2011) and Amundrud and Aven (2012). Here, each risk event is presented by a bubble, which provides information concerning probability, consequence and evidential uncertainty. The former two are indicated by the position of the bubble in the diagram, whereas the latter is represented using the size of the bubble. Figure 3 shows a graphical representation of bubble diagrams.

[7] In the work by Duijm (2015), it is not entirely clear what the limits exactly represent.
[8] A prediction interval for the consequences C is an interval such that C will be in the interval with a certain probability, typically 90% or 95%.

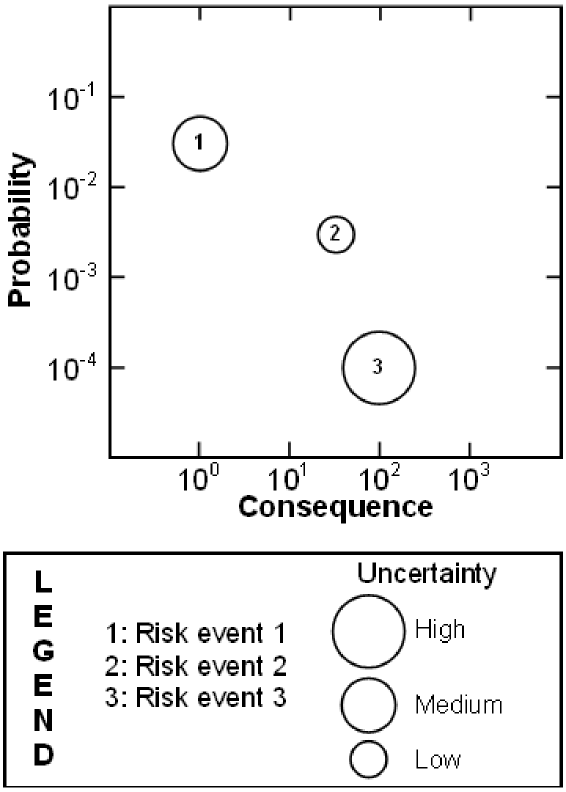


**Figure 3: Bubble diagram, based on Amundrud and Aven (2012)**

In this type of diagram, probabilities are subjective (knowledge-based) probabilities $P_s$, and the expected consequences are used as a measure to describe the possible consequences in case the event occurs, i.e. E[C|A]. The (evidential) uncertainties are categorized using an uncertainty rating scheme first introduced by Flage and Aven (2009), see Table 1.

**Table 1: Uncertainty rating classification scheme, based on Flage and Aven (2009)**

| Rating | Conditions |
|---|---|
| Low uncertainty | All of the following conditions are met:<br>- The assumptions made are seen as very reasonable<br>- Much reliable data are available<br>- There is broad agreement/consensus among experts<br>- The phenomena involved are well understood; models used are known to give predictions with the required accuracy |
| High uncertainty | One or more of the following conditions are met:<br>- The assumptions made represent strong simplifications<br>- Data are not available, or are unreliable<br>- There is lack of agreement/consensus among experts<br>- The phenomena involved are not well understood; models are non-existent or known/believed to give poor predictions |
| Medium uncertainty | Conditions between those characterizing low and high uncertainty |

Note: A modified version of this scheme is applied in Amundrud and Aven (2012), where the conditions for low and high uncertainty both use the phrasing "*one or more* of the following conditions are met" [emphasis added]. In Goerlandt and Montewka (2015b), the conditions for low and high uncertainty both use the phrasing "*all* of the following conditions are met" [emphasis added].

In Abrahamsen et al. (2014), an additional condition is applied for categorizing the uncertainty as 'high' in bubble diagrams. This the case for situations where there is a potential for significant deviation between the expected consequences E[C|A] as assessed based on the available evidence and the consequences as they may actually occur.

### 3.4. PCDS with prediction intervals and strength-of-evidence assessments

As an alternative to focusing on the evidential uncertainty in bubble diagrams, Aven (2013) and Abrahamsen et al. (2014) propose probability-consequence diagrams with prediction intervals and strength-of-evidence assessments (PCD-PISEA). Such diagrams visualize risk through the following three dimensions): i) the assigned probability for the event occurrence, ii) a 90% (or any other meaningful percentage) prediction interval for the consequence given the occurrence of the events and iii) a measure of the strength-of-evidence on which the probability and consequence assignments are based. In such a PCD-PISEA, probabilities are subjective (knowledge-based), while the strength-of-evidence assessments are used to overcome the inability of the probabilities and prediction intervals to communicate the strength of the evidence on which these measurements are based.

Aven (2013) proposes a 3-dimensional PCS-PISEA visualization as shown in Figure 4. It shows the probabilities P and prediction intervals for the consequences C on the horizontal axes, and the strength-of-evidence SE on the vertical axis using different lengths of the vertical bars. The rating for the strength-of-evidence assessment applies the same criteria as for the uncertainty rating for the bubble diagrams of Section 2.2.3, see Table 1. Here, low and high uncertainty correspond to strong and weak knowledge, respectively.

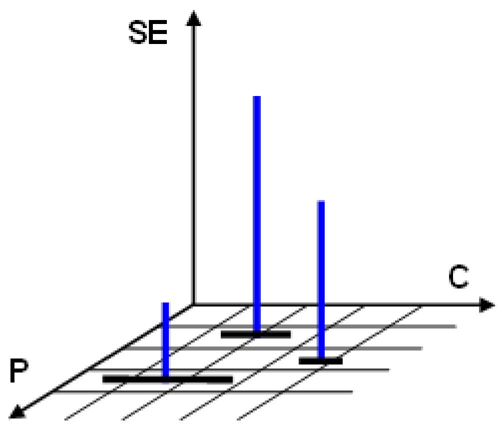


**Figure 4: 3-dimensional PCD-PISEA, based on Aven (2013)**

An alternative, 2-dimensional PCS-PISEA is proposed by Abrahamsen et al. (2014), see Figure 5. Here, the different risk events are depicted using subjective probabilities and predictions intervals, where the strength of evidence is indicated by bubbles of different sizes. The same criteria for categorizing the strength of evidence are applied as in the 3-dimensional PCD-PISEA above.

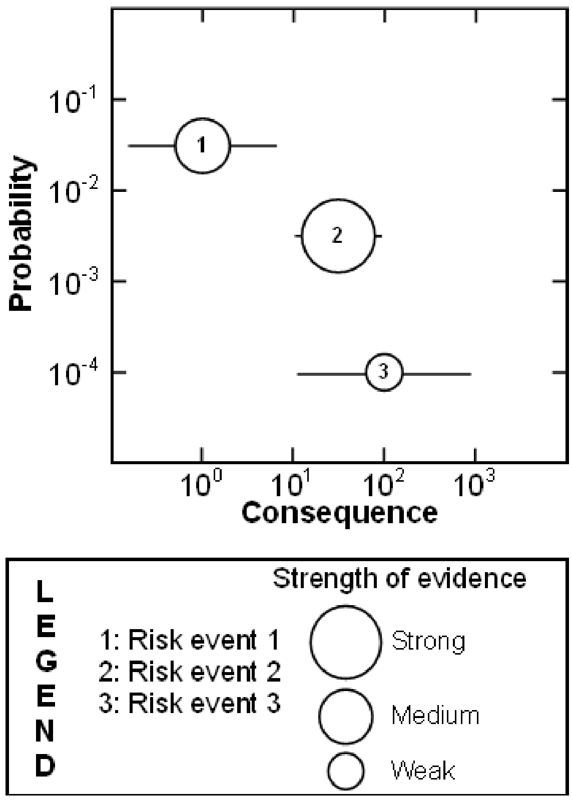


**Figure 5: 2-dimensional PCD-PISEA,**
**based on Abrahamsen et al. (2014)**

# 4. Discussion on the existing proposals

In this section, the proposals for visualizing uncertainty in risk diagrams are discussed, focusing on the types of features included in the various risk diagrams and their strengths and weaknesses. The obtained insights from this discussion are subsequently used in Section 5 to present a number of new proposals for visualizing uncertainty in PCDS.

## 4.1. Elements of PCDS with uncertainty treatment

When inspecting the proposals of Section 3.1. for visualizing uncertainty in PCDS, following elements are considered:

- markers of risk events, showing their probability and consequence dimension
- indicators of the uncertainty / strength-of-evidence for assessing the risk events
- indicators of the potential for large deviations / surprises between the assessed risk events and the risk events as they may actually occur

### 4.1.1. Markers for probability and consequences of risk events

Various markers for the probability and consequences of risk events are used: expected values E[C|A] (Section 3.2, 3.3 and 3.4) , 90%-prediction intervals (Section 3.4) and probability distributions over the consequence dimension (Section 3.1).

Expected values are convenient metrics to summarize the information about the severity of the consequences. Its meaningfulness as a measure can be derived from the law or large numbers, i.e. that the average of a number of random quantities can be accurately approximated by the expected value when the number of quantities is high, see e.g. (Faber, 2009). However, expected values can be very misleading if the underlying distribution is known or believed to be heavily skewed, i.e. when it is highly asymmetrical. The limitations of expected values have been stressed also elsewhere, e.g. Abrahamsen and Aven (2011) and Fenton and Neil (2012).

Prediction intervals provide more flexibility to communicate the range in which the consequence is expected to occur, and can, in combination with other measures of the distribution provide more insight in the possible consequences in case the risk event occurs.

The use of different risk curves, which represent probability distributions over the consequence range, is conceptually similar to prediction intervals. However, an important drawback in their application is that it is not easy to explain how to interpret these. From Section 3.1, it is found that the family of "risk curves" is derived from assigning subjective probabilities $P_s$ to model parameters $P_f^*$, see also Figure 1. First, it is difficult to provide a meaningful interpretation to the parameters $P_f^*$, see e.g. Aven and Reniers (2013). Second, the measurement $P_s(P_f^*)$ is a second-order probability. This concept has been intensely debated in the literature, with many researchers finding it a meaningless construct (Apostolakis, 1990; Mosleh and Bier, 1996). Finally, in practical applications, only a selection of parameters $P_f^*$ will be made for assessing a second-order probability. This obscures the meaning of the curves as these do not capture all uncertainty, but only part of it (Aven, 2010).

In the bubble diagrams and the PCD-PISEA, subjective probabilities $P_s$ are used for the probability-dimension. These have the benefit that these can be easily interpreted, e.g. based on an uncertainty standard, see Section 2.2.

### 4.1.2. Indicators for uncertainty / strength-of-evidence

For assessing the evidential uncertainty, two approaches are available, see Section 3.3 and 3.4. The evidence uncertainty assessment uses a qualitative scale (Table 1) to categorize the uncertainty about the evidence on which the probability and consequence assessments are made. The strength-of-evidence assessment is its logical counterpart, categorizing how good the evidence is for these assessments. Both are feasible concepts to communicate some key aspects of the evidence.

However, the authors agree with Aven (2013) that the strength-of-evidence is likely easier to understand in practical applications. The label "uncertainty" could cause confusion what it is one is uncertain about. As argued e.g. by Aven and Reniers (2013), the subjective probability $P_s$ as a measure of one's uncertainty about the occurrence of an event A is not uncertain for the assessor. However, various probability assignments with a same numerical value can be based on a very different evidential support, which can be important for a decision maker to appreciate. See also Flage et al. (2014) for further discussions on subjective probability and its underlying background knowledge in a risk analysis and communication context.

One issue which has not been addressed in earlier work on uncertainty treatment in risk diagrams is the potentially different strength of the evidence for assessing the probability and the consequences of risk events. The classification scheme of Table 1 combines all evidence into one uncertainty/strength-of-evidence rating. In some contexts, it may be advisable to separate these, e.g. when the consequence severity can be accurately calculated using advanced engineering models, but when the occurrence probability is highly uncertain[9]. Following the conditions of Table 1, this would result in medium uncertainty, but this is not an accurate reflection of the status of the underlying evidence.

Another issue with the classification scheme of Table 1 is the ambiguity in the definitions of the rating scheme as proposed in Flage and Aven (2009) and Amundrud and Aven (2012). In particular, there is linguistic ambiguity[10] in relation to the delineation of the conditions under which each category is to be applied. This is due to fact that the combinations of the phrases "all of the following conditions…" and "one or more of the following conditions…" for low and high uncertainty and the phrase "conditions between those characterizing low and high…" in Table 1 lead to not mutually exclusive categories. This is illustrated in Table 2, where the various combinations of the evidential support (assumptions, data, expert judgment and models) are varied in turn in relation to their level of support (good or poor evidence). For each combination, the uncertainty classification is determined based on the guidelines in different phrasings by Flage and Aven (2009), Amundrud and Aven (2012) and Goerlandt and Montewka (2015b). It is seen that in the former two, ambiguities arise as soon as one evidence category is rated as 'poor' or 'good'. This undesirable ambiguity is removed by the phrasing applied in the latter phrasing, but then all but two combinations lead to a 'medium' rating, which diminishes the usefulness of the assessment. Finally, in all variations of the classification scheme, it is not clear how to deal with cases where not all evidential categories are available as a basis for the probability or consequence assessment[11].

**Table 2: Uncertainty rating classification scheme, based on Flage and Aven (2009)**

| Case | A | D | EJ | M | $U_1$ | $U_2$ | $U_3$ | Case | A | D | EJ | M | $U_1$ | $U_2$ | $U_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | G | G | G | G | L | L | L | 9 | P | G | G | G | M/H | L/M/H | M |
| 2 | G | G | G | P | M/H | L/M/H | M | 10 | P | G | G | P | M/H | L/M/H | M |
| 3 | G | G | P | G | M/H | L/M/H | M | 11 | P | G | P | G | M/H | L/M/H | M |
| 4 | G | G | P | P | M/H | L/M/H | M | 12 | P | G | P | P | M/H | L/M/H | M |
| 5 | G | P | G | G | M/H | L/M/H | M | 13 | P | P | G | G | M/H | L/M/H | M |
| 6 | G | P | G | P | M/H | L/M/H | M | 14 | P | P | G | P | M/H | L/M/H | M |
| 7 | G | P | P | G | M/H | L/M/H | M | 15 | P | P | P | G | M/H | L/M/H | M |
| 8 | G | P | P | P | M/H | L/M/H | M | 16 | P | P | P | P | H | H | H |

Note: A: assumptions | D: data | EJ: expert judgment | G: good evidence | M: model | P: poor evidence | $U_1$: uncertainty rating as in Flage and Aven (2009) | $U_2$: uncertainty rating as in Amundrud and Aven (2012) | $U_3$: uncertainty rating as in Goerlandt and Montewka (2015b), interpretations and notes as in Table 1

[9] One example of such a case is waterway risk analysis: the accident probability and the conditions under which accidents occur are highly uncertain, but advanced time-domain simulation models exist for accurately determining the consequence (Goerlandt and Kujala, 2014; Ståhlberg et al., 2013).

[10] Linguistic ambiguity concerns a statement which can be interpreted in two or more possible ways. It is one of the ambiguity types identified in risk assessments, see Johansen and Rausand (2015).

[11] One could argue that the ambiguity in the phrasing for the evidence rating is a minor issue. The authors however disagree with such a view. Considering the intense debates about the nature and interpretation of probabilities as tools to measure risk (Aven and Reniers, 2013; Watson, 1994), it should be evident that all elements of a risk description should have a clear interpretation, not only the probabilities. The need for clear interpretation of qualitative measurement schemes is also stressed by e.g. Trochim and Donnely (2008).

### 4.1.3. Indicators for surprises / potential for large deviations

Abrahamsen et al. (2014) account for the potential for large deviations between the risk as assessed based on the available evidence and the consequences as they may actually occur. Thus, this corresponds to the perspective of Appendix A to account for the potential for surprises in risk descriptions. Abrahamsen et al. (2014) consider these cases through assigning a “high” uncertainty in bubble diagrams. The other proposals of Section 3 do not consider this element.

## 4.2. Strengths and weaknesses of earlier proposals

Table 3 summarizes a number of strengths and weaknesses of the proposals outlined in Section 3. Reference is made to the above sections where the relevant issues are discussed in more detail.

**Table 3: Strengths and weaknesses of earlier proposals for visualizing uncertainty in risk diagrams**

| Strength | Weakness |
|---|---|
| **PCD with family of “risk curves” (Kaplan and Garrick, 1981; Paté-Cornell, 1996)** | |
| • Simple visual representation. | • $P_f^*$, the focus of the risk analysis, is difficult to interpret (Section 4.1.1).<br>• The use of second-order probabilities $P_s(P_f^*)$ is disputed, with many thinkers finding it a meaningless construct (Section 4.1.1).<br>• Risk curves give impression that all uncertainty can be quantified, which is misleading. (Section 4.1.1.)<br>• The evidential uncertainties beyond the quantification, e.g. due to poor assumptions or unreliable data, are not reflected in the risk picture.<br>• The potential for deviations between the presented risk quantifications and the possible actual outcomes are not communicated. |
| **PCD with uncertainty boxes (Duijm, 2015)** | |
| • Simple visual representation.<br>• Communicates both expected values and uncertainty bounds for probabilities and consequences. | • Not clear exactly which probabilities ($P_f^*$ or $P_s$) are used as a basis of the PCD (Section 3.2).<br>• Not clear exactly what the limits of the boxes mean (Section 3.2).<br>• The evidential uncertainties beyond the quantification, e.g. due to poor assumptions or unreliable data, are not reflected in the risk picture.<br>• The potential for deviations between the presented risk quantifications and the possible actual outcomes are not communicated. |
| **Bubble diagram (Abrahamsen and Aven, 2011; Amundrud and Aven, 2012)** | |
| • Simple visual representation<br>• Focus on subjective (knowledge-based) probabilities $P_s$ allows for an easy interpretation (Appendix A)<br>• The evidential uncertainties beyond the quantification, e.g. due to poor assumptions or unreliable data, are reflected in the risk picture in a simple way.<br>• The potential for deviations between the presented risk quantifications and the possible actual outcomes are incorporated in a simple way. | • Communicates only expected values of probabilities and consequences, while these may provide poor insight in the range of possible values (Section 4.1.1).<br>• Uncertainty categorization system of Table 1 can lead to ambiguity and unclear grounds for selecting an uncertainty rating (Section 4.1.2).<br>• Uncertainty categorization system of Table 1 does not differentiate between evidential uncertainties related to the probabilities and consequences (Section 4.1.2.)<br>• Uncertainties related to the evidence and uncertainties related to the potential for surprises are mixed, which can conceal disputes about the locus of uncertainty. |

| **PCD with prediction intervals and strength-of-evidence assessment (PCD-PISEA) (Abrahamsen et al., 2014; Aven, 2013)** | |
|---|---|
| • Simple visual representation (2-dimenstional version)<br>• Focus on subjective (knowledge-based) probabilities $P_s$ allows for an easy interpretation (Appendix A)<br>• Communicates both expected values and uncertainty bounds (prediction interval) for the consequences.<br>• The evidential uncertainties beyond the quantification, e.g. due to poor assumptions or unreliable data, are reflected in the risk picture in a simple way. | • The 3-dimensional representation can easily become visually cluttered if axis labels and markers, risk event identification numbers are shown in addition to the prediction intervals and strength-of-evidence bars.<br>• Only communicates prediction interval for consequences, not for probabilities.<br>• Strength-of-evidence categorization system of Table 1 can lead to ambiguity and unclear grounds for selecting a rating (Section 4.1.2).<br>• Strength-of-evidence categorization system of Table 1 does not differentiate between evidential uncertainties related to the probabilities and consequences (Section 4.1.2.)<br>• The potential for deviations between the presented risk quantifications and the possible actual outcomes are not communicated (Section 4.1.3). |

Overall, it is found that the PCDS with family of “risk curves” has most drawbacks, whereas uncertainty boxes, bubble diagrams and PCDS-PISEA have attractive features, but also some weaknesses. The most favorable tool, which best captures the uncertainty-based risk perspectives as outlined in Appendix A, is the PCD-PISEA. However, some inconsistencies in the definition of the categories for the strength-of-evidence assessment and the lack of consideration for the potential for surprises leave room for improvements to these proposals. This is considered next.

# 5. Proposals for representing uncertainty in risk diagrams

Starting from an uncertainty-based risk perspective, justified in Section 2.2. and outlined further in Appendix A, following aspects are found necessary in the design of PCDS. First, clarity is needed about the type of probability applied in the diagrams. Here, use is made of subjective probabilities as these can be easily interpreted, see Appendix A. Second, consequences can be measured using nominal, ordinal or ratio scale numbers, and should be defined according to a suitable scale, see Trochim and Donnely (2008) for the issue of scaling. It is found advisable to use separate diagrams for different consequence groups (environmental, financial, injuries/fatalities), see Flage and Røed (2012) and Abrahamsen et al. (2014). In the proposals in Section 5.1 and 5.2, the consequence dimension applies an arbitrary scale. Third, expected consequences can be used to locate the risk event on the diagram, but it is advisable to apply additional markers (boxes, prediction intervals or similar) to communicate that these expected values do not appreciate the space of possible outcomes very well. Fourth, while they provide a similar message, a strength-of-evidence assessment is preferred over an uncertainty assessment, primarily because it is clearer what the focus of this assessment is. The evidence assessment can be made separately for the probability and consequence dimension, should allow for an unambiguous categorization and give consideration to the fact that not all evidence types are necessarily used for a given risk event. Finally, consideration can be given to the potential for surprises, i.e. to the possibility of a significant deviation between the assessed probability and consequence and the consequences as they really may occur. This can be considered through the assessment of ‘assumption deviation risk’ as described by Aven (2013), see Appendix B.

In addition to the above, further recommendations for using PCDS can be found in Duijm (2015) and Ale et al. (2015), e.g. the benefits of using continuous (i.e. ordinal or ratio) scales and the additional attractive features of logarithmic scales. It is reminded that PCDS are here intended to be used for visualizing as the result of an analysis, not as complete risk analysis tools, see Flage and Røed (2012) and Abrahamsen et al. (2014).

### 5.1. Proposal 1: PCDS with Tukey box plots and strength-of-evidence assessments (PCDS-USEA$_1$)

A first proposal for visualizing uncertainty in probability-consequence diagrams is shown in Figure 6. It applies Tukey box plots for communicating the uncertainty about the consequences and/or about the event occurrence, combined a qualitative strength-of-evidence assessment.

Tukey box plots are exploratory graphics for showing the distribution of a variable. They are quick to interpret, easy to understand and provide insight in some main characteristics of a distribution, including the minimum and maximum (excluding outliers), the median and lower and upper quartile, see McGill et al. (1978) for further details. The use of box plots is especially useful if the results of the quantitative analysis are derived by a probabilistic model such as Bayesian Networks, such that distributions over the outcomes are inherent in the modeling approach. In the left pane of Figure 6, the box plots are one-dimensional over the consequences, whereas in the right pane, the box plot information is 2-dimensional. As such, the Tukey box plots are similar to the uncertainty boxes (Section 3.2) and the prediction intervals (Section 3.4), but provide more detailed insight in the distributions.

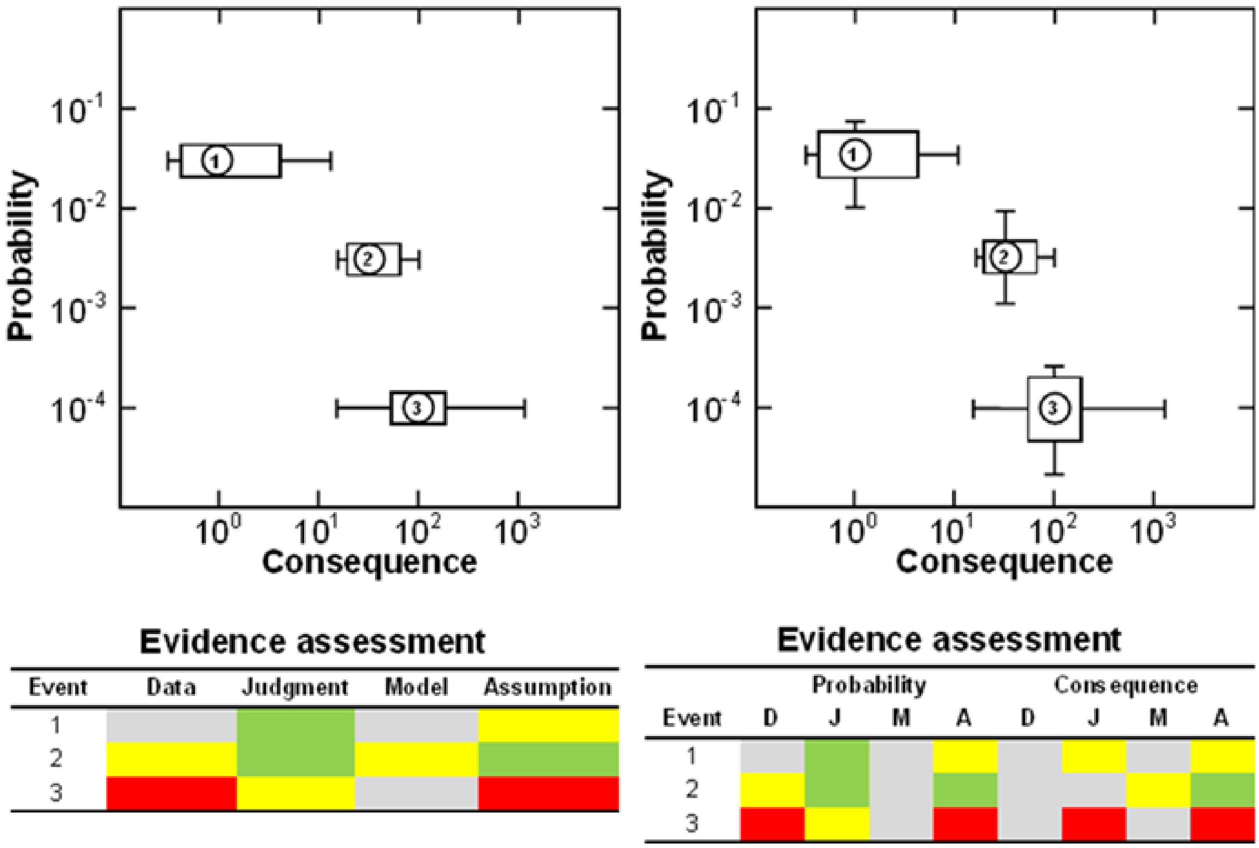


**Figure 6: PCD with Tuckey Box Plots and strength-of-evidence assessment**

The qualitative strength-of-evidence assessment is performed using a 3-level categorization of each applicable evidence category, namely data, expert judgments, models and assumptions. A simple traffic light symbolism is applied to rate each evidence category separately, whereas non-applicable evidence categories are marked in gray. In the left pane of Figure 6, the evidence for probability and consequence is combined, whereas in the right pane these are kept separate. Separating these evidence types serves three purposes. First, the ambiguities in defining a categorization which combines all three, as discussed in Section 4.1.2., are avoided. Second, it is clarified which evidential types are available for the judgment of the various risk events, a feature which is obscured in the PCDS-PISEA of Section 3.4. Third, it allows the decision maker, rather than the analyst, to judge the overall evidential strength for each risk event. This can be important, because it is known that certain people choose to rely more readily on data and models than on judgments, and vice versa (Glendon et al., 2006).

The selection of the appropriate rating for the evidence category is done using following procedure. For each risk event, each evidence type is assessed using a set of evidential qualities. These qualities are described in Table 4 for the data and model evidence types, based on suggestions by Goerlandt and Montewka (2015b). Table 5 shows the qualities of the judgments and assumptions evidence types, based on Flage et al. (2014) and Boone et al. (2010).

**Table 4: Evidential characteristics and criteria for strength-of-evidence rating for data and model evidence types**

| Evidence type | Strong evidential characteristics | Weak evidential characteristic |
|---|---|---|
| **Data** | | |
| Quality | Low number of errors<br>High accuracy of recording<br>High reliability of data source | High number of errors<br>Low accuracy of recording<br>Low reliability of data source |
| Amount | Much relevant data available | Little data available |
| **Models** | | |
| Empirical validation | Many different experimental tests performed<br>Existing experimental tests agree well with model output | No or little experimental confirmation available<br>Existing experimental tests show large discrepancy with model output |
| Theoretical viability | Model expected to lead to good predictions | Model expected to lead to poor predictions |

**Table 5: Evidential characteristics and criteria for strength-of-evidence rating for judgment and assumption evidence types**

| Evidence type | Strong | Medium | Weak |
|---|---|---|---|
| **Judgments** | Broad intersubjectivity: more than 75% of peers support the judgment | Moderate intersubjectivity: between 25% and 75% of peers support the judgment | Predominantly subjective: less than 25% of peers support the judgment |
| **Assumptions** | | | |
| Agreement among peers | Many (more than 75%) would have made the same assumption | Several (between 25% and 75%) would have made the same assumption | Few (less than 25%) would have made the same assumption |
| Influence on results | The assumption has only local influence | The assumption has wider influence in the analysis | The assumption greatly determines the results of the analysis |

Using the characteristics of the data (quality and amount) and models (empirical validation and theoretical viability), a strength of evidence rating for these evidence types is determined as follows. First, it is decided based on the descriptions of strong and weak evidential characteristics, which categories apply for the quality, amount, empirical validation and theoretical viability. Then, the strength of evidence is determined as the average rating of the relevant characteristics of the data and models evidence categories, respectively.

A similar approach is taken for the judgments and assumptions. Here, the evidence characteristics are rated as strong, medium or weak. For the assumptions, the average of the ratings is taken as the strength of evidence. Combinations involving ‘medium’ and another rating (‘strong’ or ‘weak’), get the rating of the latter as the rating of the strength of evidence.

## 5.2. Proposal 2: PCDS with uncertainty intervals, strength-of-evidence assessments and assessments of assumption deviation risks (PCDS-USEA$_2$)

A second proposal for visualizing uncertainty in probability-consequence diagrams is shown in Figure 7. It applies uncertainty intervals, strength-of-evidence assessments and assessments of assumption deviation risks. These are visualized using bubbles, which are positioned on the location of the expected value of probability and consequence. These bubbles are segmented, where each segment is colored using a simple traffic light analogy. In the left pane of Figure 7, the evidence for probability and consequence is combined, whereas in the right pane these are kept separate in two concentric segments. At the center of the bubble, a colored star-shaped icon is used to convey information concerning the assumption deviation risk assessment.

Together, these graphical elements provide a clear insight in the uncertainty of event occurrence and consequence, the strength of the evidence base for assessing these uncertainties, and the potential deviations compared to the assessed uncertainties.

Uncertainty intervals have already been introduced in Section 3.4. The strength-of-evidence assessments for data, models and judgments are performed using the classification schemes proposed in Section 5.1. The assumption deviation risk methodology (Aven, 2013) is outlined in Appendix B. It focuses on the magnitude of the deviations which could occur to the results of the risk analysis, due to the assumptions made in the analysis. In the context of this paper, it is applied as follows. First, all main assumptions on which the assessment of the quantification of the risk event is based, are identified. Subsequently, these assumptions are rated using the assumption deviation risk method. Finally, the maximum rating of all assumptions underlying the measurement of a specific risk event is taken, as this can cause most deviation to the basic risk picture. This rating is visually represented as in Figure 7.

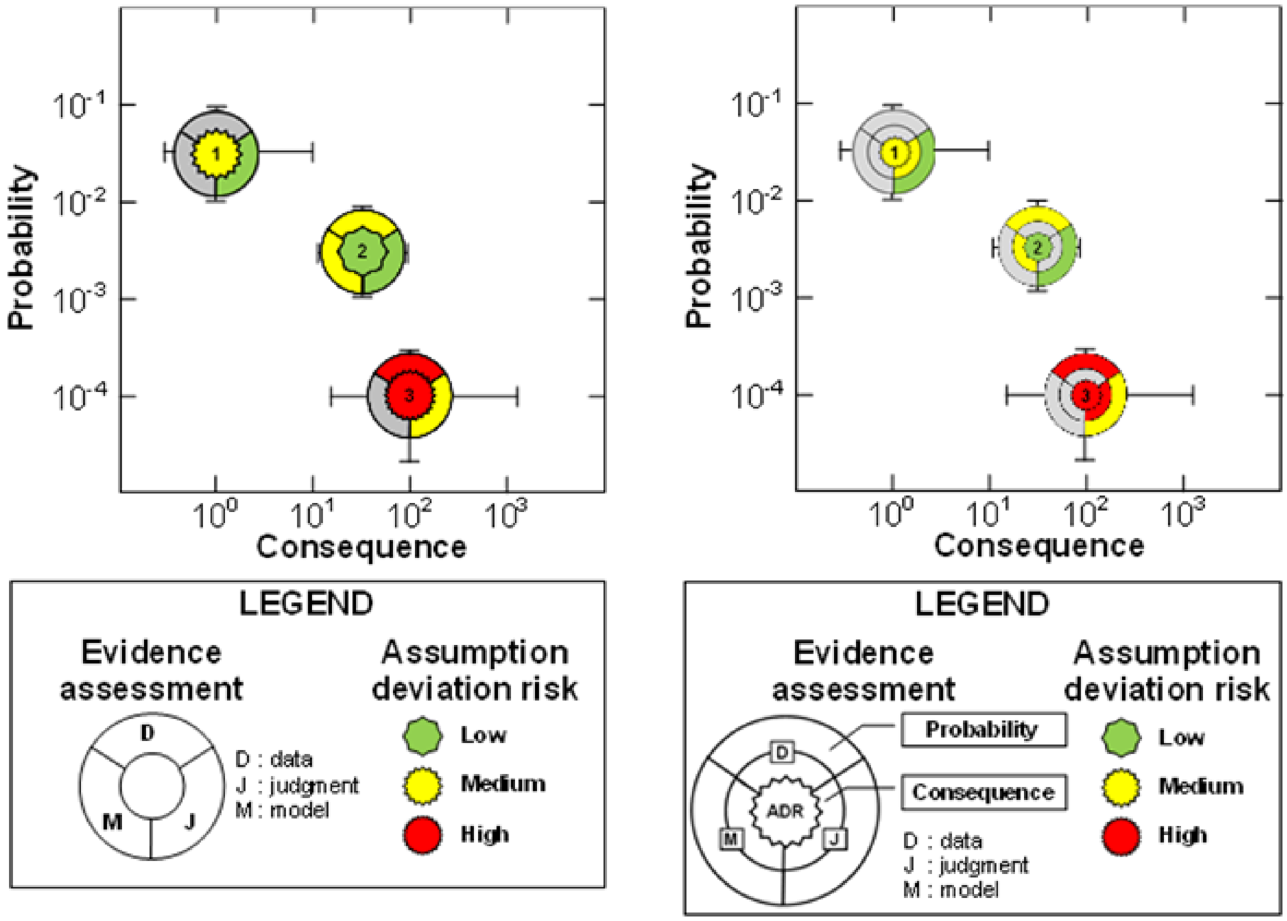


**Figure 7: PCD with strength of evidence and assumption deviation risk assessment**

# 6. Conclusions

In this paper, the assessment and visualization of uncertainty in PCDS has been considered. After justifying why uncertainty needs to be considered in such diagrams, a relatively recently proposed uncertainty-based risk perspective is adopted as the basis for considering the issue of uncertainty visualization.

A review of existing proposals for representing uncertainty in PCDS is made, including the family of "risk curves" approach, uncertainty boxes, bubble diagrams and PCDS with prediction intervals and strength-of-evidence assessments (PCD-PISEA). A discussion on the elements found in these proposals has revealed a number of strengths and weaknesses of these. Overall, the PCS-PISEA approach was found most favorable. However, due to some inconsistencies in the strength-of-evidence rating and the lack of inclusion of the potential for surprises, some modifications to this approach have been suggested.

The new approaches focus directly on the strength-of-evidence, which is treated separately for different evidence types. This is done to alleviate the inconsistencies in earlier proposals, and especially to provide more direct insight in the types of evidence supporting the quantitative uncertainty assessments of probabilities and consequences. In another proposal, an assumption deviation risk assessment is visualized along with a segmented strength-of-evidence assessment.

# Abbreviations

A: event | ADR: assumption deviation risk | C: consequence | E[C|A]: expected consequences in case an event occurs | FN-curve: curve showing the frequency of exceedance (F) of a given number of fatalities (N) | K: knowledge on which the analysis is based | PCD: probability-consequence diagram | PCD-USEA$_1$: probability-consequence diagram with Tukey box plots and strength-of-evidence assessments | PCD-USEA$_2$: probability-consequence diagram with uncertainty intervals, strength-of-evidence assessments and assessments of assumption deviation risks | PCD-PISEA: probability-consequence diagram with prediction intervals and strength-of-evidence assessments | $P_f$: frequentist probability | Ps: subjective (knowledge-based) probability | Q: measure of uncertainty | QRA: quantitative risk analysis | RM: risk matrix | SE: strength-of-evidence | U: uncertainty

# Acknowledgements

The research presented in this paper has been conducted in the context of the “Strategic and Operational Risk Management for Wintertime Maritime Transportation System” (STORMWINDS) project. This has received funding from BONUS, the joint Baltic Sea research and development programme (Art 185), funded jointly from the European Union’s Seventh Programme for research, technological development and demonstration and from the Academy of Finland. The financial support is acknowledged.

# References

Abrahamsen, E.B., Amundrud, O., Aven, T., Gelyani, A.M., 2014. Safety oriented bubble diagrams vs. risk plots based on prediction intervals and strength-of-knowledge assessments. Which one to use as an alternative to risk matrices? Int. J. Bus. Contin. Risk Manag. 5, 197–209.

Abrahamsen, E.B., Aven, T., 2011. Safety oriented bubble diagrams in project risk management. Int. J. Perform. Eng. 7, 91–96.

Ale, B., Burnap, P., Slater, D., 2015. On the origin of PCDS - (probability consequence diagrams). Saf. Sci. 72, 229–239.

Amundrud, O., Aven, T., 2012. A practical guide on how to present and visualize the result of risk and vulnerability analyses in a societal safety and security context. Presented at the 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference, Helsinki, Finland.

Apostolakis, G., 1990. The concept of probability in safety assessments of technological systems. Science 250, 1359–1364.

Aven, T., 2013. Practical implications of the new risk perspectives. Reliab. Eng. Syst. Saf. 115, 136–145. doi:10.1016/j.ress.2013.02.020

Aven, T., 2012. The risk concept—historical and recent development trends. Reliab. Eng. Syst. Saf. 99, 33–44. doi:10.1016/j.ress.2011.11.006

Aven, T., 2011. Selective critique of risk assessments with recommendations for improving methodology and practise. Reliab. Eng. Syst. Saf. 96, 509–514. doi:10.1016/j.ress.2010.12.021

Aven, T., 2010. On how to define, understand and describe risk. Reliab. Eng. Syst. Saf. 95, 623–631. doi:10.1016/j.ress.2010.01.011

Aven, T., Reniers, G., 2013. How to define and interpret a probability in a risk and safety setting. Saf. Sci. 51, 223–231. doi:10.1016/j.ssci.2012.06.005

Aven, T., Zio, E., 2011. Some considerations on the treatment of uncertainties in risk assessment for practical decision making. Reliab. Eng. Syst. Saf. 96, 64–74. doi:10.1016/j.ress.2010.06.001

Boone, I., Van der Stede, Y., Dewulf, J., Messens, W., Aerts, M., Daube, G., Mintiens, K., 2010. NUSAP: a method to evaluate the quality of assumptions in quantitative microbial risk assessment. J. Risk Res. 13, 337–352.
Campbell, S., 2005. Determining overall risk. J. Risk Res. 8, 569–581.
Cox, L.A., 2008. What's wrong with risk matrices? Risk Anal. 28, 497–512.
Douglas, H., 2009. Science, Policy and the value-free ideal. University of Pittsburgh Press, Pittsburgh, PA.
Douglas, H., 2000. Inductive risk and values in science. Philos. Sci. 67, 559–579.
Duijm, N.J., 2015. Recommendations on the use and design of risk matrices. Saf. Sci. 76, 21–31.
Faber, M.H., 2009. Risk and safety in engineering, Coursebooks in Engineering. ETH Zürich.
Fenton, N., Neil, M., 2012. Risk assessment and decision analysis with Bayesian networks. CRC Press.
Fischhoff, B., 1995. Risk perception and communication unplugged: twenty years of process. Risk Anal. 15, 137–145.
Flage, R., Aven, T., 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis (QRA). Reliab. Risk Anal. Theory Appl. 2, 9–18.
Flage, R., Aven, T., Zio, E., Baraldi, P., 2014. Concerns, challenges, and directions of development for the issue of representing uncertainty in risk assessment. Risk Anal. 34, 1196–1207.
Flage, R., Røed, W., 2012. A reflection on some practices in the use of risk matrices. Presented at the 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference, Helsinki, Finland, pp. 881–891.
Franks, A.P., Maddison, T., 2006. A simplified method for the estimation of individual risk. Process Saf. Environ. Prot. 84, 101–108.
Garvey, P.R., 2009. Analytical methods for risk management. A systems engineering perspective. CRC Press, Boca Raton, FL.
Glendon, A.I., Clarke, S.G., McKenna, E.F., 2006. Human safety and risk management, 2nd ed. Taylor & Francis Group, Boca Raton, FL.
Goerlandt, F., Kujala, P., 2014. On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk. Saf. Sci. 62, 348–365. doi:10.1016/j.ssci.2013.09.010
Goerlandt, F., Montewka, J., 2015a. Maritime transportation risk analysis: review and analysis in light of some foundational issues. Reliab. Eng. Syst. Saf. 138, 115–134.
Goerlandt, F., Montewka, J., 2015b. A framework for risk analysis of maritime transportation sytems: a case study for oil spill from tankers in a ship-ship collision. Saf. Sci. 76, 42–66.
Goerlandt, F., Montewka, J., Kujala, P., 2014. Tools for an Extended Risk Assessment for Ropax Ship-Ship Collision, in: Beer, M., Au, S.-K., Hall, J.W. (Eds.), Vulnerability, Uncertainty, and Risk: Quantification,Mitigation, and Management. American Society of Civil Engineers, pp. 2292–2302.
Haugen, S., Vinnem, J.E., 2015. Perspectives on risk and the unforeseen. Reliab. Eng. Syst. Saf. 137, 1–5.
Hayes, K.R., 2011. Uncertainty and uncertainty analysis methods (No. EP102467). CSIRO.
Helton, J.C., Johnson, J.D., 2011. Quantification of margins and uncertainties: alternative representations of epistemic uncertainty. Reliab. Eng. Syst. Saf. 96, 1034–1052.
Hubbard, D., Evans, D., 2010. Problems with scoring methods and ordinal scales in risk assessment. IBM J. Res. Dev. 54.
Hubbard, Douglas W., 2009. The failure of risk management: why it's broken and how to fix it. Wiley.
IMO, 2007. Formal safety assessment - consolidated text of the guidelines for formal safety assessment (FSA) for use in the IMO rule-making process.
IPIECA/OGP, 2013. Oil spill risk assessment and response planning for offshore installations - final report. IPIECA/OGP, London, UK.
ISO, 2010. IEC/ISO 31010:2009 Risk management - risk assessment techniques. CENELEC, Brussels.
Johansen, I.L., Rausand, M., 2015. Ambiguity in risk assessment. Saf. Sci. 80, 243–251.
Kaplan, S., 1997. The words of risk analysis. Risk Anal. 17, 407–417.
Kaplan, S., Garrick, J.B., 1981. On the quantitative definition of risk. Risk Anal. 1, 11–27.
Kiureghian, A. Der, Ditlevsen, O., 2009. Aleatory or epistemic? Does it measure? Saf. Sci. 31, 105–112.
Kontovas, C.A., Psaraftis, H.N., 2009. Formal safety assessment: a critical review. Mar. Technol. 46, 45–59.
Levine, E.S., 2012. Improving risk matrices: the advantages of logarithmically scales axes. J. Risk Res. 15, 209–222.
Levin, R., 2005. Uncertainty in risk assessment - contents and modes of communication (Doctoral Thesis). Royal Institute of Technology, Sweden.

Lindley, D.V., 2006. Understanding uncertainty, 11th ed. Wiley-Interscience, Hoboken, NJ.
Markowski, A.S., Mannan, M.S., 2008. Fuzzy risk matrix. J. Hazard. Mater. 159, 152–157.
Mayer, T., Reniers, G., 2013. Engineering risk management. De Gruyter Graduate.
McGill, R., Tukey, J.W., Larsen, W.A., 1978. Variations of box plots. Am. Stat. 32, 12–16.
Montewka, J., Goerlandt, F., Kujala, P., 2014. On a systematic perspective on risk for formal safety assessment (FSA). Reliab. Eng. Syst. Saf. 127, 77–85. doi:10.1016/j.ress.2014.03.009
Mosleh, A., Bier, V.M., 1996. Uncertainty about probability: a reconciliation with the subjectivist viewpoint. IEEE Trans. Syst. Man Cybern. - Part Syst. Hum. 26, 303–310.
Murphy, C., Gardoni, P., Harris Jr., C.E., 2011. Classification and moral evaluation of uncertainties in engineering modeling. Sci. Eng. Ethics 17, 553–570.
NHS, 2008. A risk matrix for risk managers. The National Patient Safety Agency, London, UK.
Ni, H., Chen, A., Chen, N., 2010. Some extensions on risk matrix approach. Saf. Sci. 48, 1269–1278.
O'Brien, M., 2000. Making better environmental decisions. The MIT Press, Cambridge.
Paté-Cornell, E.M., 1996. Uncertainties in risk analysis: six levels of treatment. Reliab. Eng. Syst. Saf. 54, 95–111.
Pickering, A., Cowley, S., 2010. Risk matrices: implied accuracy and false assumptions. J. Health Saf. Res. Pract. 2, 9–16.
Ruan, X., Yin, Z., Frangopol, D.M., 2015. Risk matrix integrating risk attitudes based on utility theory. Risk Anal., 1437-1447 35. doi:10.1111/risa.12400
Rudner, R., 1953. The scientist qua scientist makes value judgments. Philos. Sci. 20, 1–6.
Shrader-Frechette, K.S., 1993. Burying uncertainty: risk and the case against geological disposal of nuclear waste. University of California Press.
Singpurwalla, N.D., 2006. Reliability and risk: a Bayesian perspective, Wiley Series in probability and statistics. Wiley, New York, NY.
Smith, E.D., Siefert, W.T., Drain, D., 2009. Risk matrix input data biases. Syst. Eng. 12, 344–360.
Spiegelhalter, D., Pearson, M., Short, I., 2011. Visualizing uncertainty about the future. Science 333, 1393–1400.
Ståhlberg, K., Goerlandt, F., Ehlers, S., Kujala, P., 2013. Impact scenario models for probabilistic risk-based design for ship–ship collision. Mar. Struct. 33, 238–264. doi:10.1016/j.marstruc.2013.06.006
Steel, D., 2010. Epistemic values and the argument from inductive risk. Philos. Sci. 77, 14–34.
Thomas, P., Bratvold, R.B., Bickel, J.E., 2014. The risk of using risk matrices. SPE Econ. Manag. 56–66.
Trochim, W., Donnely, J.P., 2008. The research methods knowledge base, 3rd ed. Atomic Dog Publishing.
Watson, S.R., 1994. On the meaning of probability in probabilistic safety analysis. Reliab. Eng. Syst. Saf. 45, 261–269.
Zio, E., Pedroni, N., 2013. Methods for representing uncertainty - a literature review (No. 2013-03), Cahiers de la Sécurité Industrielle. Fondation pour une Culture de Sécurité Industrielle, Toulouse, France.

# Appendix A

In an uncertainty-based risk perspective, focus is on events A, consequences C and the uncertainties U related to their occurrence. A risk description is made by specifying the relevant events and consequences, and by measuring the uncertainties using an uncertainty measure Q. The most commonly used quantitative tool for describing the uncertainties is probability P, but others exist, e.g. imprecise (interval) probabilities and representations based on theories of evidence, see Zio and Pedroni (2013) for an overview. Other descriptors of uncertainty can e.g. be the possibility of surprises or deviations from the quantified uncertainty measures ($Q_{qu}$), which can be described using qualitative uncertainty measures ($Q_{ql}$). The specification of the events and consequences leads to a set of quantities of interest Z, e.g. the societal costs or the number of fatalities. The entire analysis is conditional to the available knowledge K, which contains uncertainties U as well. This knowledge base can be described in different ways, e.g. by measures expressing the strength of evidence SE.

In the current proposals for visualizing uncertainty in PCDS, following risk perspective is used:

$$R \sim (A, C, Ps \quad , SE \mid K) \tag{Eq. A.1}$$

$$R \sim \left(A, C, P_s, SE, Q_{ql} \,\middle|\, K\right) \tag{Eq. A.2}$$

Where $P_s$ is a subjective (knowledge-based) probability, which is a measure of an assessor's degree of belief (e.g. about the occurrence, or not, of a given event). This probability is a subject-bound measurement based on a given background knowledge, and can be interpreted with reference to an uncertainty standard (Aven and Reniers, 2013; Lindley, 2006). Depending on the strength of the knowledge base, the probability can be understood as entirely subjective to broad intersubjectively objective.

# Appendix B

The assumption deviation risk assessment is introduced in Aven (2013), based on which the below outline is made. The identified assumptions on which the analysis is based are converted to a set of uncertainty factors. These uncertainty factors are given an assumption deviation risk score, which represents the criticality/importance of the assumption. This assessment captures following aspects: i) the deviation of the analysis due to the assumptions, ii) a measure of the uncertainty of this deviation and iii) the knowledge on which this assessment is based.

The magnitude of deviation is classified in low, medium and high. These respectively correspond to situations where maximum plausible changes in base values result in outcome changes of less than an order of magnitude (low), about an order of magnitude (medium) and two or more orders of magnitude (high). The degree of belief about the deviation occurring is assessed in low (negligible), medium ($P_s$=0.01) and high categories ($P_s$=0.5). Finally, the strength-of-knowledge for assessing the magnitude of deviation and the probability of deviation occurrence is rated in low, medium and high.

Based on the rating for the magnitude of deviation and its occurrence probability, a low, medium or high criticality rating is assigned to the assumption. If the strength-of-knowledge for this assumption assessment is low or medium, the basic rating of the assumption criticality is moved up one category, i.e. from low to medium and from medium to high.