Refining a probabilistic cross-impact methodology for scenario analysis

Tuomas Haapasalo

School of Science

Bachelor's thesis
Espoo 15.01.2025

Supervisor

Prof. Ahti Salo

Advisor

M.Sc.(Tech.) Jussi Leppinen

**Aalto University
School of Science**

Copyright © 2025 Tuomas Haapasalo

The document can be stored and made available to the public on the open internet pages of Aalto University.
All other rights are reserved.

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the bachelor's thesis

Author Tuomas Haapasalo

Title Refining a probabilistic cross-impact methodology for scenario analysis

Degree programme Bachelor's Programme in Science and Technology

Major Mathematics and Systems Sciences **Code of major** SCI3029

Teacher in charge Prof. Ahti Salo

Advisor M.Sc.(Tech.) Jussi Leppinen

Date 15.01.2025 **Number of pages** 33+1 **Language** English

Abstract

Scenario analysis is an important tool for supporting managerial decision-making. Preparing for the unexpected is increasingly relevant, as demonstrated by the unforeseen recent developments such as the COVID-19 pandemic, the conflict in Ukraine, and the Gaza war. Examining alternative futures may reveal hidden vulnerabilities or strengths. Scenario thinking breaks narrow-mindedness and helps overcome cognitive biases by bringing surprising scenarios into consideration.

The field of scenario analysis methods is broad and diverse. By and large, there are qualitative and quantitative methods, as well as combinations thereof. Another major distinction can be made between methods that utilize probabilities and those that do not. At the core of these methods are the ways of scenario construction. Examples of systematic methods that account for mutual dependencies of uncertainty factors are cross-impact analysis methodologies, in general, choosing a suitable method depends on the intended use and context. In this thesis, we examine whether a probabilistic cross-impact analysis method can be enhanced by changing the interpretation of the cross-impact term with the odds-ratio.

The applicability of the new asymmetric cross-impact term was evaluated by examining the effects of its asymmetry on the functionality of the method. The original model was modified to support the new interpretation, whereafter the distributions produced by both methods were compared using statistical and visual tests. The most important observations included the apparent similarity of the joint probability distributions for the most probable scenarios and significant differences for the remaining scenarios. Computationally, neither method was faster than the other. The conclusion is that both interpretations are viable.

Keywords scenario analysis, cross-impact analysis, risk analysis, probabilistic cross-impact analysis

Aalto University
School of Science

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Tekniikan kandidaatintyön tiivistelmä

Tekijä Tuomas Haapasalo	
Työn nimi Todennäköisyyksiin ja ristivaikutuksiin pohjautuvan skenaarioanalyysimenetelmän jatkojalostaminen	
Koulutusohjelma Teknistieteellinen kandidaattiohjelma	
Pääaine Matematiikka ja systeemitieteet	**Pääaineen koodi** SCI3029
Vastuuopettaja Prof. Ahti Salo	
Työn ohjaaja DI Jussi Leppinen	
Päivämäärä 15.01.2025 **Sivumäärä** 33+1	**Kieli** Englanti

Tiivistelmä

Skenaarioanalyysi on tärkeä johdon päätöksenteon tukityökalu. Viime aikojen ennaltaarvaamattomat käänteet, kuten esimerkiksi koronapandemia, Ukrainan konflikti ja Gazan sota ovat osoittaneet, että varautuminen odottamattomaan on yhä ajankohtaisempaa. Vaihtoehtoisten kehityskulkujen tarkasteleminen saattaa paljastaa analysoitavan systeemin, kuten esimerkiksi liiketoimintastrategian tai -mallin, piileviä haavoittuvaisuuksia tai vahvuuksia. Skenaarioajattelu rikkoo putkinäköä ja avartaa kognitiivisia harhoja tuomalla yllättäviäkin skenaarioita tarkasteltavaksi.

Skenaarioanalyysimenetelmien kenttä on laaja ja monimuotoinen. Karkeasti jaotellen menetelmät voidaan jakaa kvalitatiivisiin ja kvantitatiivisiin sekä näiden yhdistelmiin, jotka edelleen haarautuvat moniin alaluokkiin. Toinen merkittävä jaottelu voidaan tehdä todennäköisyyksiä hyödyntäviin ja sivuuttaviin menetelmiin. Keskiössä on skenaarioiden rakentumisen logiikka, joka voi olla systemaattista tai intuitioon nojaavaa. Esimerkkinä systemaattisista menetelmistä toimii ristivaikutuksiin pohjautuvat menetelmät, jotka huomioivat epävarmuustekijöiden keskinäiset riippuvuussuhteet. Sopivan menetelmän löytäminen riippuu kuitenkin käyttötarkoituksesta ja kontekstista. Tässä tutkielmassa tarkastellaan, voisiko todennäköisyyksiin ja ristivaikutuksiin pohjautuvaa skenaarioanalyysimenetelmää parantaa muuttamalla ristivaikutuskertoimen tulkintaa vetosuhteen (odds-ratio) avulla.

Uuden epäsymmetrisen ristivaikutuskertoimen soveltuvuutta arvioitiin tarkastelemalla, olisiko sitä mahdollista sisällyttää vanhaan menetelmään epäsymmetriasta huolimatta. Alkuperäinen malli muokattiin uuden ristivaikutuskertoimen ympärille. Sen jälkeen menetelmien tuottamia jakaumia verrattiin keskenään tilastollisilla sekä visuaalisilla testeillä. Tärkeimpinä havaintoina nousivat menetelmien yhteisjakaumien samankaltaisuus todennäköisimpien skenaarioiden osalta ja merkittävät erot muiden skenaarioiden todennäköisyyksissä. Laskentanopeuksien välillä ei ollut merkittävää eroa. Johtopäätöksenä molempia menetelmiä voidaan pitää toimivina vaihtoehtoina.

Avainsanat skenaarioanalyysi, ristivaikutusanalyysi, riskianalyysi, todennäköisyyspohjainen ristivaikutusanalyysi

Contents

Abstract 3

Abstract (in Finnish) 4

Contents 5

1 Introduction 6

2 Theoretical background 7

3 Methodology 9
3.1 Scenarios . . . 9
3.2 Odds-ratio and cross-impact multipliers . . . 10
3.3 Cross-impact statements . . . 14
3.4 Iterative calculation of the joint-probability distribution . . . 14
3.5 Conditional independence . . . 19
3.6 Direct scenario probability fitting . . . 23

4 Validation of the revised interpretation 24
4.1 Differences . . . 24
4.2 Statistical tests . . . 25
4.3 Bayesian networks . . . 28

5 Discussion and conclusions 31

1 Introduction

A common challenge faced by organizations and decision-makers is the need to make choices based on assumptions about the future. However, the future is inherently complex and uncertain, making prediction difficult. Despite this, the future developments of the operating environment will significantly impact the organization's actions and success.

To address this challenge, many organizations and decision-makers employ scenario analysis, which has become an essential tool in long-term planning and strategic management (Bunn and Salo, 1993; Chermack, 2022). It is used for identifying plausible future scenarios and assessing their potential impact on the system under consideration (Balaman, 2019). Key indicators and variables, such as uncertainty factors, are first identified. Possible future developments for the variables are then projected, enabling a thorough evaluation of their effects on the analyzed system. This way, decision-makers can better prepare for a range of outcomes and reduce the risks associated with relying solely on assumptions, allowing for more informed decisions despite uncertainty (Bunn and Salo, 1993; Chermack, 2022; Kosow and Gaßner, 2008). To date, various methods of scenario analysis have been developed, each tailored to different use cases (Bunn and Salo, 1993; Kosow and Gaßner, 2008). The choice of method depends on the specific objectives and context of the analysis (Bunn and Salo, 1993; Kosow and Gaßner, 2008).

In this thesis, we adopt the cross-impact interpretation presented in Salo et al. (2022), with the exception of replacing the cross-impact terms with the odds-ratio, and refining the probabilistic cross-impact method presented by Roponen and Salo (2024). This methodology estimates probabilities for all scenarios, regardless of some of the cross-impact estimates being inconsistent. Additionally, it always produces the same results for the joint probability distribution, differentiating itself from simulation approaches.

The aim of this thesis is to explore whether the original method proposed by Roponen and Salo (2024) can be enhanced by modifying the interpretation of the cross-impact term. We anticipate that this modification will lead to improvements in the model's performance and the preservation of the original cross-impact statements within the resulting distributions. This evaluation is conducted by applying multiple statistical tests and visual tests for the resulting joint probability distributions and by comparing the conditional distributions from a case study conducted by Roponen and Salo (2024).

The structure of this thesis is as follows: Section 2 reviews some methods of scenario analysis and the historical development behind them. Section 3 details the refined cross-impact analysis (CIA) approach and the methodology. Section 4 presents the results of the comparison between the two approaches. Section 5 concludes the study.

2 Theoretical background

Scenarios help account for uncertainties in preparing decision-makers for a range of possible future developments. However, they do not provide a comprehensive picture of the future (Bunn and Salo, 1993; Chermack, 2022; Kosow and Gaßner, 2008). The first modern-day scenario analysis methods developed in the 1950s focused on generating feasible and varying descriptions of the future in the form of descriptive narratives (Bunn and Salo, 1993; Millett, 2009). Since then, numerous approaches to constructing scenarios have been developed, ranging from qualitative to purely quantitative methods, and approaches that combine aspects of both (Bunn and Salo, 1993; Kosow and Gaßner, 2008; Roponen and Salo, 2024; Salo et al., 2022).

According to Bunn and Salo (1993), the credibility and quality of scenarios are assessed by the comprehensiveness, consistency, and coherence of the scenarios. A comprehensive scenario represents a broad coverage of the plausible futures relevant to the scenario process. If a scenario is consistent, the outcomes of different uncertainty factors are not in conflict with known information nor with each other. Coherent scenarios are in accordance with the underpinning theories that they are developed with. Furthermore, scenarios can be categorized into three classes based on the outlook they take on scenario construction (Bunn and Salo, 1993; Ducot and Lubben, 1980; Jungermann, 1985; Schnaars, 1987). Scenarios can be exploratory or anticipatory, meaning that the present is used as the starting point and future development is deduced based on the current events and foremost inferences (exploratory), or that the scenarios are first formed after which the events or causes leading to the future state are inferred (anticipatory). This division coincides with the inductive-deductive classification, in which scenarios are built either with a top-down approach (deductive) or a bottom-up approach (inductive).

The methodology presented in this thesis constructs scenarios with the exploratory approach. Descriptive scenarios do not take into account the desirability of the scenarios, whereas normative scenarios do. Surprising scenarios can be considered as peripheral, whereas trend scenarios can be seen as the continuum of known patterns. The quality of the scenarios produced and the perspective on constructing the scenarios can be used to validate the scenario analysis methods the scenarios are generated with, in addition to assessing the appropriate use cases for them.

The methods of scenario analysis can be classified into three main schools, which are intuitive logics, probabilistic modified trends school, and La prospective (Bradfield et al., 2005; Bunn and Salo, 1993). In intuitive logics, the scenarios are typically built with a step-by-step workshop-type process, in which the managers' insights and expert logic are used to develop the possible scenarios. By far, the qualitative techniques have been the most popular, likely because of the ease of use and the successful marketing done by the pioneers of scenario development, e.g., Shell, SRI International, and GBN (Millett, 2009; Schnaars, 1987). The probabilistic modified trends school comprises of the Trend Impact Analysis (TIA) and the Cross-Impact Analysis (CIA) methodologies. In TIA, the impact of different trends is assessed by extrapolating relevant trends into the future with time series forecasting techniques and historical data (Bunn and Salo, 1993). The methodology presented in this thesis is based on the CIA method, in which the impact of pairwise relationships of different events is assessed to determine their impact on the future. The La prospective school is considered to be “a blending of the intuitive logics and probabilistic modified trend methodologies” (Bradfield et al., 2005). The choice of the appropriate methodology depends on the context and the objectives of the scenario analysis (Kosow and Gaßner, 2008).

Cross-impacts permit an expert to consider probabilistic or causal relationships between pairs of uncertainty factors, thus allowing to shift the focus on identifying consistent scenarios (Salo et al., 2022). Moreover, the advantage of specific cross-impact analysis methods is that they are comprehensive since all possible scenarios are preserved (Roponen and Salo, 2024). According to Gordon (1994), the first CIA methods were developed in the 1960s. Subsequently, many CIA methods have emerged after significant contributions done by Gordon (Gordon, 1994; Gordon and Hayward, 1968) Helmer (Helmer, 1977, 1981) and others (Godet, 1976, 1994; Panula-Ontto, 2019; Roponen and Salo, 2024). By and large, these methods can be divided into probabilistic and nonprobabilistic approaches. The first probabilistic CIA methods, such as the BASICS tool of Battelle Memorial Institute, and other similar approaches, do not account for inconsistent estimates for the cross-impacts and can thus produce results that violate the rules of probability theory (Huss and Honton, 1987; Roponen and Salo, 2024). On the other hand, the nonprobabilistic approaches, such as the Cross Impact Balances (CIB) discussed in Weimer-Jehle (2006) or the consistency analysis method proposed by Seeve and Vilkkumaa (2022), cannot be used for probabilistic risk analysis nor expected utility theory (Roponen and Salo, 2024). Additionally, Salo et al. (2022) show that some of the nonprobabilistic approaches, such as the CIB method, are not founded on solid frameworks and have limitations on the number of consistent scenarios that can be produced.

3 Methodology

We employ the same definitions and notations for scenarios and partial scenarios as in Roponen and Salo (2024) and Salo et al. (2022). Hereby, we summarize the methodological development in Roponen and Salo (2024) in Section 3.1. In addition, Sections 3.3–3.6 share many similarities with the original paper because we are refining the original method by adjusting the model for the new interpretation of the cross-impact term.

3.1 Scenarios

Scenarios can be defined as combinations of the realizations of different uncertainty factors. The key factors whose outcomes are uncertain and which impact the strategic decisions at hand are called uncertainty factors.

Let X^i, where $i = 1, 2, ..., N$, be a discrete random variable that represents an uncertainty factor. Each uncertainty factor has n_i outcomes $S_i = \{1, ..., n_i\}$. Therefore, a scenario is defined as a vector $\mathbf{s} = (s_1, ..., s_N)$, where $s_i \in S_i$ denotes an outcome of the *i-th* uncertainty factor and the length of the vector N is the number of uncertainty factors. The set S of all possible scenarios is formed as the Cartesian product $S := S_{1:N} = S_1 \times S_2 \times \cdots \times S_N$. The number of possible scenarios is $|S| = \prod_{i=1}^{N} n_i$.

Consequently, the number of possible scenarios grows rapidly when the number of uncertainty factors or the outcomes for the uncertainty factors are increased. As an example, 11 uncertainty factors with three different outcomes for each uncertainty factor produce $3^{11} = 177\,147$ possible scenarios. Therefore, it is not viable to estimate individual probabilities for all possible scenarios without a systematic methodology. Instead, the methods of cross-impact analysis allow us to conclude information about the scenario probabilities by describing probabilistic dependencies between uncertainty factors (Roponen and Salo, 2024).

A partial scenario is defined as $S_F = \bigtimes_{i \in F} S_i$, which is a combination of the outcomes of uncertainty factors contained in a subset $F \subseteq \{1, \ldots, N\}$ of all uncertainty factors. If all uncertainty factors are in F, the set F is the set of all scenarios. A partial scenario comprising of the outcomes for the first i uncertainty factors is defined as $\mathbf{s}_{1:i}$. A partial scenario $\mathbf{s}_{1:i} = (s_1, ..., s_i) \in S_{1:i}$ that has the same outcomes for the first i uncertainty factors as the full scenario is compatible with the full scenario. Therefore, partial scenarios can be defined as the set of scenarios that can be constructed by extending the partial scenarios with the outcomes for the remaining uncertainty factors $j = i + 1, ..., N - 1, N$ such that $E(\mathbf{s}_{1:i}) = \{s' \in S \mid s'_j = s_j, \forall j = 1, ..., i\}$. The sum of the probabilities of scenarios that can be formed by extending the partial scenarios is the probability of a partial scenario $\mathbf{s}_{1:i} \in S_{1:i}, i \leqslant N$. This is expressed as

$$p(\mathbf{s}_{1:i}) = \sum_{\mathbf{s}_{1:N} \in E(\mathbf{s}_{1:i})} p(\mathbf{s}_{1:N}). \tag{1}$$

The marginal probability for the outcome $l \in S_j$ of the *j-th* uncertainty factor is attained similarly as the sum of the probabilities of those scenarios where the uncertainty factor has the same state

$$P(X^j = l) = \sum_{\mathbf{s} \in S_{1:N} | s_j = l} p(\mathbf{s}). \tag{2}$$

A simple example of scenarios and uncertainty factors is given in Table 1, where the possibilities of "green" products are explored. In this fictional example, a company is trying to assess whether the following five years would be favorable for a new "green" product line, which manufactures products in an ecological fashion. Investment in an ecological product line comes with a hefty price tag; consequently, it would be beneficial to employ scenario analysis to gain insight into the state of the market. We assume that the marginal probabilities for the uncertainty factors have been estimated by experts and they indicate how likely each outcome is on its own. In this simplified example, $3^3 = 27$ different scenarios could be constructed.

Table 1: Exploring the possibilities of "green" products using scenario analysis. One possible scenario is indicated by the green outcomes for uncertainty factors.

**Uncertainty factors are described above and possible outcomes below them. The marginal probability of an outcome is described next to the outcome.*

Scenario 1

Geopolitics	Marginal probability	Economic development	Marginal probability	Regulation	Marginal probability
Prolonged conflict between Israel and Palestine leads to blocks in world politics.	0.4	The EU economy has recovered from high inflation and interest rates.	0.3	Moderate regulatory support for green products.	0.3
Israel and Palestine agree on peace, which calms the atmosphere and increases international co-operation.	0.5	Slow economic growth, with lingering effects of high inflation and interest rates in the EU.	0.4	Strong regulatory support for green products, with subsidies and incentives.	0.25
Conflict in the Middle-East spreads to neighboring countries, which suppresses global collaboration and freezes global markets.	0.1	Moderate economic stability, with the EU maintaining steady but unremarkable growth.	0.3	Comprehensive green standards and international agreements promoting sustainable practices.	0.45

3.2 Odds-ratio and cross-impact multipliers

In probability theory, the odds of an event occurring are defined as the ratio of the probability of an event occurring to the probability of it not occurring. Specifically, for an event a, the odds $O(a)$ are defined as

$$O(a) = \frac{P(a)}{1 - P(a)}.$$

Odds provide a measure of the likelihood of a certain outcome relative to its non-occurrence.

We adopt a similar cross-impact approach as Salo et al. (2022). However, we modify the interpretation of the cross-impact between two events. In this thesis, the cross-impact between events a and b is defined as

$$C_{ab} := \frac{O(a \mid b)}{O(a)} = \frac{P(a \mid b)}{1 - P(a \mid b)} \frac{1 - P(a)}{P(a)}, \tag{3}$$

meaning that the cross-impact multiplier C_{ab} expresses the relative ratio of the odds of a occurring when b has occurred to the odds of a occurring independently. We can rearrange Equation (3) so that the conditional probability is

$$P(a \mid b) = \frac{C_{ab} P(a)}{1 - P(a) + C_{ab} P(a)}. \tag{4}$$

Therefore, the cross-impact measures how much the likelihood of the event a changes if b will occur, measuring the relative impact of the event b on a's likelihood. We do not assume the events to occur in a temporal sequence in Equations (3) and (4), meaning that b could occur after a has occurred and it would still affect the likelihood of a occurring.

A major drawback of this approach is that the symmetry in the cross-impact term is, in general, lost:

$$\begin{aligned} C_{ab} &= \frac{P(a \cap b)/P(b)}{1 - P(a \cap b)/P(b)} \frac{1 - P(a)}{P(a)}, \\ C_{ba} &= \frac{P(a \cap b)/P(a)}{1 - P(a \cap b)/P(a)} \frac{1 - P(b)}{P(b)}, \end{aligned}$$

and if $P(a) \neq P(b)$, then $C_{ab} \neq C_{ba}$. However, when a and b are independent, it is possible that $C_{ab} = C_{ba} = 1$.

This complicates the task of estimating the cross-impact statements explained in Section 3.3 since the cross-impact multipliers are not the same in both directions as in Roponen and Salo (2024) and Salo et al. (2022), where only half of the cross-impacts have to be estimated. Assuming that there are N uncertainty factors that have the same number of outcomes n (i.e., $n_i = n, i = 1, ..., N$), the cross-impacts would have to be estimated in both directions resulting in

$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} n_i n_j = N(N-1)n^2$$

estimates, instead of

$$\frac{1}{2} N(N-1)n^2$$

estimates as in Roponen and Salo (2024), making the estimation process arduous.

However, if we assume that the cross-impact statements in Section 3.3 refer to the consistency of the outcome pairs of the uncertainty factors without accounting the causal relation or temporal sequence of the pairs, the direction of the cross-impact should be irrelevant. Hence, the consistency estimate would, in reality, reflect how likely these outcomes would appear in the same world, and only a one-time pairwise estimate of the consistency of the outcome pair would be needed. Therefore, only the upper diagonal cross-impacts would be estimated, resulting in the same number of cross-impact statements as in Roponen and Salo (2024). This is illustrated in Table 2, where the cross-impacts of the example case of exploring the possibilities of "green" products have been estimated.

Table 2: The example of exploring the possibilities of "green" products continued. The cross-impact statements have been estimated for all uncertainty factors and only for the upper half of the matrix. The cross-impact estimates ranging from -3 to 3 have been converted into cross-impact multipliers using Equation (7). The multipliers are inside the brackets.

			Geopolitics			Economic development			Regulation		
			Prolonged conflict between Israel and Palestine leads to blocks in world politics.	Israel and Palestine agree on peace, which calms the atmosphere and increases international co-operation.	Conflict in the Middle-East spreads to neighboring countries, which suppresses global collaboration and freezes global markets.	The EU economy has recovered from high inflation and interest rates.	Slow economic growth, with lingering effects of high inflation and interest rates in the EU.	Moderate economic stability, with the EU maintaining steady but unremarkable growth.	Moderate regulatory support for green products.	Strong regulatory support for green products, with subsidies and incentives.	Comprehensive green standards and international agreements promoting sustainable practices.
			0.40	0.50	0.10	0.30	0.40	0.30	0.30	0.25	0.45
Geopolitics	Prolonged conflict between Israel and Palestine leads to blocks in world politics.	0.40				-2 (1/2)	2 (2)	-1 (2/3)	1 (5/2)	0 (1)	-1 (2/3)
	Israel and Palestine agree on peace, which calms the atmosphere and increases international co-operation.	0.50				2 (2)	0 (1)	2 (2)	1 (5/2)	2 (2)	3 (3)
	Conflict in the Middle-East spreads to neighboring countries, which suppresses global collaboration and freezes global markets.	0.10				-2 (1/2)	1 (5/2)	-2 (1/2)	-1 (2/3)	-1 (2/3)	-2 (1/2)
Economic development	The EU economy has recovered from high inflation and interest rates.	0.30							1 (5/2)	2 (2)	2 (2)
	Slow economic growth, with lingering effects of high inflation and interest rates in the EU.	0.40							1 (5/2)	1 (5/2)	1 (5/2)
	Moderate economic stability, with the EU maintaining steady but unremarkable growth.	0.30							2 (2)	3 (3)	2 (2)
Regulation	Moderate regulatory support for green products.	0.30									
	Strong regulatory support for green products, with subsidies and incentives.	0.25									
	Comprehensive green standards and international agreements promoting sustainable practices.	0.45									

Furthermore, we interpret the cross-impact multiplier between outcome i of uncertainty factor k and outcome j of uncertainty factor l using the abbreviated notations for the marginal and conditional probabilities of any $k \in S_i$ and $l \in S_j$ by stating that $p_k^i := P(X^i = k)$, $p_l^j := P(X^j = l)$ and $p_{k|l}^{i|j} = P(X^i = k|X^j = l)$ as

$$C_{kl}^{ij} = \frac{p_{k|l}^{i|j}}{1 - p_{k|l}^{i|j}} \frac{1 - p_k^i}{p_k^i} \Leftrightarrow p_{k|l}^{i|j} = \frac{C_{kl}^{ij} p_k^i}{1 - p_k^i + C_{kl}^{ij} p_k^i}.$$

The Kolmogorov definition of conditional probability states that if A and B are dependent, the conditional probability is $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. Moreover, the probability of both events occurring is defined as $P(a \cap b) = P(b \cap a)$. Thus, $p_{kl}^{ij} = p_{k|l}^{i|j} p_l^j$ and $p_{lk}^{ji} = p_{l|k}^{j|i} p_k^i$. However, if we interpret the cross-impact multiplier to be symmetric (i.e. $C_{kl}^{ij} = C_{lk}^{ji} = C$), the Kolmogorov definition of conditional probability would be satisfied only if the marginal probabilities p_k^i and p_l^j would be equal since $p_{kl}^{ij} = p_{lk}^{ji} \Leftrightarrow \frac{C p_k^i p_l^j}{1 - p_k^i + C p_k^i} = \frac{C p_l^j p_k^i}{1 - p_l^j + C p_l^j} \Leftrightarrow p_k^i = p_l^j$. This is not a favorable condition for the marginal probabilities. If we use the condition $p_{kl}^{ij} = p_{lk}^{ji}$ as the basis without stating anything about the symmetry of the cross-impacts (i.e., $C_{kl}^{ij} \stackrel{?}{=} C_{lk}^{ji}$), the requirement for mathematical consistency would be

$$p_{lk}^{ji} = p_{kl}^{ij} \Leftrightarrow \frac{C_{lk}^{ji} p_l^j}{1 - p_l^j + C_{lk}^{ji} p_l^j} p_k^i = \frac{C_{kl}^{ij} p_k^i}{1 - p_k^i + C_{kl}^{ij} p_k^i} p_l^j,$$

which can be simplified into

$$C_{lk}^{ji} = \frac{C_{kl}^{ij}(1 - p_l^j)}{1 - p_k^i + C_{kl}^{ij}(p_k^i - p_l^j)}, \tag{5}$$

meaning that for the rules of probability to be logically consistent (i.e., $p_{kl}^{ij} = p_{lk}^{ji}$), the lower diagonal of the cross-impact matrix is to be inferred from the upper diagonal cross-impact statements using (5). This condition does not necessarily require the marginal probabilities to be equal but further justifies the estimation of only the upper diagonal of the cross-impact matrix, as it provides all the necessary information for the calculations. Thus, the joint probability $p_{kl}^{ij} = p_{lk}^{ji}$ can be calculated using

$$p_{kl}^{ij} = \frac{C_{kl}^{ij} p_k^i p_l^j}{1 - p_k^i + C_{kl}^{ij} p_k^i}. \tag{6}$$

To calculate the joint probabilities using (6), we only need to estimate the upper half of the cross-impact matrix.

It is worth mentioning that the joint probability could also be calculated by switching the roles of p_{kl}^{ij} and p_{lk}^{ji} in the presented reasoning, and estimating only the lower diagonal of the cross-impact statements. However, since we can infer the upper diagonal cross-impact multipliers by inverting Equation (5), the joint probabilities remain consistent regardless of which set of cross-impact statements we estimate. Therefore, we can choose to estimate the upper diagonal entries for convenience.

3.3 Cross-impact statements

As in Roponen and Salo (2024), we assume that the cross-impact statements are elicited with a -3 to 3 scale that can be converted into fractional cross-impact multipliers by rounding the results of the following conversion

$$C_{kl}^{ij} = \sqrt{2}^{V_{kl}^{ij}}, \tag{7}$$

where C_{kl}^{ij} is the cross-impact multiplier and V_{kl}^{ij} is the statement between the outcome pair of the two uncertainty factors at hand. The statements can thus be expressed by numerical values $\frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1, \frac{3}{2}, 2$ and 3. This can be seen in Table 2, where the example of exploring the possibilities of "green" products is continued.

Roponen and Salo (2024) show that not all probabilistic dependencies between two uncertainty factors can be motivated by causality between them. Therefore, we also employ the interpretation that the cross-impacts express other types of probabilistic dependencies in addition to causality. As an example, the cross-impacts of the outcomes for Economic development and Geopolitics in Table 2 are not seen as causal relations in either direction.

They also demonstrate that estimating cross-impacts that follow the laws of probability theory may be challenging for experts. They give an example where the sum of the joint probabilities of the outcomes for some uncertainty factors is greater than one. In addition, the marginal probabilities ensuing the joint probability distribution can differ from the original estimations and from probabilities obtained from the joint distributions of different uncertainty factor pairs. However, we also interpret the cross-impact estimates as insightful enough for them to be the basis for the construction of a consistent and accurate probability distribution, regardless of some of the estimates not being mutually consistent.

3.4 Iterative calculation of the joint-probability distribution

In this section, we follow the notation and methodological development regarding conditional probability updating, exercised in Roponen and Salo (2024) with the exception of modifying the methods to suit the new interpretation of the cross-impact multiplier addressed in Section 3.2. Since we are refining the original method in Roponen and Salo (2024), a large part of the methodological development addressed here and in the Section 3.5 overlaps with the original paper.

We also represent estimates that are used as input variables with a hat symbol. The underlying probability distribution $P(\cdot)$ is represented by $p(\cdot)$. Approximations about scenario probabilities are referred to as $\hat{p}(\cdot)$, where the scenarios being considered are specified by the argument. As an example, $\hat{p}(s_i)$ is the estimation of the marginal probability $p(s_i)$ for some outcome $s_i \in S_i$. Probabilities derived from estimates are represented with $q(\cdot)$.

If the estimates about marginal probabilities $\hat{p}_k^i$, $\hat{p}_l^j$ are correct (i.e. $\hat{p}_k^i = p_k^i = P(X^i = k)$, $\hat{p}_l^j = p_l^j = P(X^j = l)$ and $\hat{C}_{kl}^{ij} = C_{kl}^{ij}$), these estimates can be used to determine the marginal probabilities trough summing all scenarios where $X^i = k$ and

$X^j = l$, since the probability $p_{kl}^{ij} = P(X^i = k, X^j = l)$ is equal to $(\hat{C}_{kl}^{ij}\hat{p}_k^i\hat{p}_l^j)/(1-\hat{p}_k^i + \hat{C}_{kl}^{ij}\hat{p}_k^i)$ as expressed by Equation (6). This introduces a constraint on the scenario probabilities

$$\sum_{\substack{\mathbf{s}\in S_{1:N} \\ s_i=k, s_j=l}} p(\mathbf{s}) = \frac{\hat{C}_{kl}^{ij}\hat{p}_k^i\hat{p}_l^j}{1-\hat{p}_k^i + \hat{C}_{kl}^{ij}\hat{p}_k^i}, \tag{8}$$

where the sum is taken over scenarios in which the outcomes for uncertainty factors i and j correspond to the outcomes on the right side of the equation. The sum of the probabilities of these scenarios must equal the probability indicated by the right side of the equation.

The problem where some estimates for the cross-impacts are missing is solved by assuming that there exists a binary relation $R_{ij} : S_i \times S_j$ such that $(s_i, s_j) \in R_{ij}$ if and only if the cross-impact multiplier (4) has been estimated (Roponen and Salo, 2024). Now, the term $\hat{C}_{kl}^{ij}$ will be included for every pair of outcomes that satisfy the relation $R(s_i, s_j)$. This implies that the most accurate probability distribution, based on these estimates, can be found by solving the minimization problem

$$\min_{p(s)} \sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{(k,l)\in R_{ij}} \left[\left(\sum_{\substack{\mathbf{s}\in S_{1:N} \\ s_i=k, s_j=l}} p(\mathbf{s}) \right) - \frac{\hat{C}_{kl}^{ij}\hat{p}_k^i\hat{p}_l^j}{1-\hat{p}_k^i + \hat{C}_{kl}^{ij}\hat{p}_k^i} \right]^2. \tag{9}$$

As Roponen and Salo (2024) mention, from a computational perspective, a significant issue with problem (9) is that the optimization process must consider all possible scenarios. This becomes problematic when the number of scenarios is large, which occurs if there are numerous uncertainty factors or if these factors have multiple possible outcomes (recall that the total number of scenarios is given by $\prod_{i=1}^{N} n_i$, where n_i represent the number of outcomes for the *i-th* uncertainty factor). This becomes evident when trying to fit the scenario probability distribution directly, which is addressed in Section 3.6.

However, they also note that the probability of any scenario $s \in S_{1:N}$ can be expressed by conditioning the *i-th* uncertainty factor's outcome on the partial scenario containing outcomes for all uncertainty factors prior to the *i-th* uncertainty factor. Specifically, this can be written as

$$p(\mathbf{s}) = p(s_i \mid \mathbf{s}_{1:i-1})p(\mathbf{s}_{1:i-1}). \tag{10}$$

Thus, the problem of estimating scenario probabilities becomes much smaller if the probabilities for these partial scenarios are known. Roponen and Salo (2024) state that for a given order of uncertainty factors, relationship (10) can be expressed as

$$p(\mathbf{s}_{1:N}) = p(s_1)p(s_2 \mid s_1)\cdots p(s_{N-1} \mid \mathbf{s}_{1:N-2})p(s_N \mid \mathbf{s}_{1:N-1}).$$

This iterative approach presented by them, indicates that scenario probabilities can be progressively determined as follows: 1. Start with the marginal probabilities of the first uncertainty factor, $p(s_1)$. 2. Calculate the conditional probabilities $p(s_2 \mid s_1)$

to optimally fit the cross-impact multipliers for the first two uncertainty factors. 3. Utilize these conditional probabilities to compute the probabilities for partial scenarios that include the first two uncertainty factors.

As demonstrated by Roponen and Salo (2024), this method can then be extended step-by-step to include additional uncertainty factors. The iterative procedure is as follows:

1. Calculate the conditional probabilities for the next uncertainty factor based on the previously computed partial scenario probabilities and the estimates of marginal probabilities and cross-impact multipliers.

2. Update the set of partial scenarios to incorporate the new uncertainty factor. The number of these partial scenarios is the product of (i) the number of partial scenarios from the previous iteration and (ii) the number of possible outcomes for the new uncertainty factor.

3. Use the derived conditional probability distributions to compute the joint probability distribution for the updated set of partial scenarios.

As in Roponen and Salo (2024), the iteration procedure can be performed by calculating probabilities for all partial scenarios $\mathbf{s}_{1:i}$, $i = 1, \ldots, N$ and using the conditional probabilities $q(s_i \mid \mathbf{s}_{i:i-1})$ so that the the iteration is initialized by setting $q(k) \leftarrow \hat{p}_k^1$ for some $k \in S_1 = \{1, \ldots, n_1\}$. The conditional probabilities following every step for the iteration can be calculated using

$$\min_{q(k|\mathbf{s}_{1:i-1})} \sum_{j=1}^{i-1} \sum_{(k,l)\in R_{ij}} \left[\left(\sum_{\{\mathbf{s}\in S_{1:i-1}|s_j=l\}} q(k \mid \mathbf{s}) q(\mathbf{s}) \right) - \frac{\hat{C}_{kl}^{ij} \hat{p}_k^i \hat{p}_l^j}{1 - \hat{p}_k^i + \hat{C}_{kl}^{ij} \hat{p}_k^i} \right]^2 \tag{11}$$

$$\sum_{\mathbf{s}\in S_{1:i-1}} q(k \mid \mathbf{s}) q(\mathbf{s}) = \hat{p}_k^i, \quad \forall k \in \{1, 2, \ldots, n_i\} \tag{12}$$

$$\sum_{k=1}^{n_i} q(k \mid \mathbf{s}_{1:i-1}) = 1, \quad \forall \mathbf{s}_{1:i-1} \in S_{1:i-1} \tag{13}$$

$$q(k \mid \mathbf{s}_{1:i-1}) \geqslant 0. \quad \forall k \in \{1, 2, \ldots, n_i\}, \mathbf{s}_{1:i-1} \in S_{1:i-1}. \tag{14}$$

In the third term of the objective function, the summation is performed over partial scenarios where the *j-th* uncertainty factor matches the specified outcome in the relation R_{ij}. The final two constraints ensure the proper definition of the conditional probability distribution. The derived probabilities for the following partial scenarios (formed by adding the states of the *i-th* uncertainty factor $k \in S_i$ to the preceding partial scenarios $\mathbf{s}_{1:i-1}$) can be defined as $q(\mathbf{s}_{1:i-1}, k_i) \leftarrow q(\mathbf{s}_{1:i-1}) q(k_i \mid \mathbf{s}_{1:i-1})$. Constraint (12) guarantees that the marginal probability aligns with the estimated marginal probability $\hat{p}_k^i$ for the outcome $s_i = k$, ensuring the computed probabilities precisely match the estimated marginal probabilities. The evolution of the iteration process is demonstrated by Table 3, where the conditional probabilities of each uncertainty factor's outcomes are calculated.

Table 3: Example of the iterations when calculating the conditional probabilities for the "green" product example.

1st iteration		-
1. Geopolitics	$p(s^{11})$	0.40
	$p(s^{12})$	0.50
	$p(s^{13})$	0.10

2nd iteration		1. Geopolitics		
		s^{11}	s^{12}	s^{13}
2. Economic development	$p(s^{21} \mid s^{11})$	0.21	0.37	0.29
	$p(s^{22} \mid s^{12})$	0.55	0.27	0.49
	$p(s^{23} \mid s^{13})$	0.24	0.36	0.22

3rd iteration		1. Geopolitics								
		s^{11}			s^{12}			s^{13}		
		2. Economic development								
		s^{21}	s^{22}	s^{23}	s^{21}	s^{22}	s^{23}	s^{21}	s^{22}	s^{23}
s_{D3}		(s^{11}, s^{21})	(s^{11}, s^{22})	(s^{11}, s^{23})	(s^{12}, s^{21})	(s^{12}, s^{22})	(s^{12}, s^{23})	(s^{13}, s^{21})	(s^{13}, s^{22})	(s^{13}, s^{23})
3. Regulation	$p(s^{31} \mid s_{D3})$	0.47	0.41	0.38	0.32	0.21	0.23	0.07	0.06	0.08
	$p(s^{32} \mid s_{D3})$	0.24	0.21	0.33	0.17	0.22	0.27	0.43	0.41	0.45
	$p(s^{33} \mid s_{D3})$	0.29	0.38	0.29	0.51	0.57	0.5	0.5	0.53	0.47

Even though this approach appears unequivocal, Roponen and Salo (2024) show that there are significant aspects to consider regarding how the cross-impacts ought to be used, how the resulting distribution is formed, and how to handle inconsistent cross-impact statements. First, cross-impacts alone are insufficient to fully characterize multivariate distributions, as the number of conditional probabilities $q(k_i \mid \mathbf{s}_{1:i-1})$ significantly exceeds the number of cross-impact multipliers in the objective function. This discrepancy is especially pronounced when only a subset of cross-impact multipliers has been elicited. When estimates for all cross-impact multipliers are obtained, the objective function has $\sum_{i=1}^{N-1} n_i n_j$ terms. The number of constraints (12) is n_i, while Equation (13) assigns $\prod_{j=1}^{i-1} n_j$ constraints. The total number of parameters, $\prod_{j=1}^{i} n_j$, matches the number of partial scenarios of length i. Consequently, the algebraic equation system (12)–(14) is underdetermined, allowing for multiple optimal solutions.

This leads to the second consideration, how the resulting distribution depends on the method used to solve the optimization problem. The results presented here and in the original paper were obtained using Matlab's built-in interior point method.

Third, a convenient property of this optimization formulation is its ability to handle inconsistent cross-impact statements (i.e. some of the estimates $\hat{C}_{kl}^{ij}, \hat{p}_k^i, \hat{p}_l^j$

deviate from the cross-impact multipliers and marginal probabilities ensuing the calculated probabilities $q(\mathbf{s}), \mathbf{s} \in S_{1:N}$). The corresponding cross-impact multipliers $\mathring{C}^{ij}_{kl}$ are defined as

$$\mathring{C}^{ij}_{kl} = \frac{(1-\hat{p}^i_k)}{\hat{p}^i_k \hat{p}^j_l - \hat{p}^i_k \sum_{\substack{\{\mathbf{s}\in S| \\ s_i=k,\, s_j=l\}}} q(\mathbf{s})} \sum_{\{\mathbf{s}\in S | s_i=k,\, s_j=l\}} q(\mathbf{s}).$$

This relationship is valuable as it helps identify and revise those cross-impact estimates that most significantly differ from the implied cross-impacts, either in absolute terms or based on the joint event probabilities $P(X_k = i, X_l = j)$ in the objective function (11). Since marginal probabilities are matched, the cross-impact terms that maximize

$$\underset{\substack{i,j\in\{1,\ldots N\} \\ (k,l)\in R_{ij}}}{\arg\max} \left| \hat{C}^{ij}_{kl} - \mathring{C}^{ij}_{kl} \right|$$

deviate the most from the implied cross-impact multiplier derived from scenario probabilities $q(s)$. Conversely, the solution

$$\underset{\substack{i,j\in\{1,\ldots N\} \\ (k,l)\in R_{ij}}}{\arg\max} \left| \frac{(\hat{C}^{ij}_{kl} - \mathring{C}^{ij}_{kl})\hat{p}^i_k \hat{p}^j_l}{1-\hat{p}^i_k + (\hat{C}^{ij}_{kl} - \mathring{C}^{ij}_{kl})\hat{p}^i_k} \right|$$

gives the cross-impact multiplier(s) with the greatest discrepancy between the estimated and computed event probabilities. This analysis helps identify and revise inconsistent cross-impact multipliers.

Moreover, other probability distributions than $q(\mathbf{s})$, which match the estimated cross-impact multipliers and marginal probabilities $\hat{C}^{ij}_{kl}, \hat{p}^i_k, \hat{p}^j_l$ just as effectively, can be explored with the implied cross-impacts through the following optimization problem

$$\min_{\mathring{q}(\mathbf{s})} f(\mathring{q})$$

$$\sum_{\substack{\mathbf{s}\in S_{1:N} \\ s_i=k, s_j=l}} \mathring{q}(\mathbf{s}) = \frac{\mathring{C}^{ij}_{kl}\hat{p}^i_k \hat{p}^j_l}{1-\hat{p}^i_k + \mathring{C}^{ij}_{kl}\hat{p}^i_k}, \forall i \in \{2,\ldots,N\}, j \in \{1,\ldots,i-1\}, (k,l) \in R_{ij}$$

$$\sum_{\substack{\mathbf{s}\in S_{1:N} \\ s_i=k}} \mathring{q}(\mathbf{s}) = \hat{p}^i_k, \forall i \in \{1,\ldots,N\}, k \in \{1,\ldots,n_i\}$$

$$\mathring{q}(\mathbf{s}_{1:N}) \geq 0, \forall \mathbf{s}_{1:N} \in S_{1:N}$$

As Roponen and Salo (2024) demonstrate, here, $f : \mathbb{R}^{|S|} \to \mathbb{R}$ is a function chosen to identify a probability distribution $\mathring{q}$ with specific desired properties, such as maximizing a scenario's probability or approaching a uniform distribution.

3.5 Conditional independence

As Roponen and Salo (2024) note in the original paper, there are inherent limitations to using cross-impact multipliers to predict all possible scenario probability distributions when dealing with numerous uncertainty factors. For instance, assume N uncertainty factors each have n outcomes. The number of different cross-impact multipliers that can be elicited is proportional to the square of the number of uncertainty factors and their outcomes since $\sum_{i=1}^{N}\sum_{j=i+1}^{N} n_i n_j = N(N-1)n^2/2$ is the number of estimated cross-impact statements. Nevertheless, the number of scenarios grows exponentially with the number of uncertainty factors (i.e. $\prod_{i=1}^{N} n_i = n^N$). This means that estimates of marginal probabilities and cross-impact multipliers do not fully describe all possible scenario probabilities, as the number of constraints from cross-impact multipliers will be far fewer than the number of scenarios.

In this context, Roponen and Salo (2024) mention two noteworthy observations on the optimization problem (11)–(14). First, the number of estimates for cross-impact multipliers increases at each step of the iterative algorithm because the outcomes of the new uncertainty factor n_i are compared with previously considered factors. If all estimates are provided, there are $n_i \sum_{j=1}^{i-1} n_j/2$ terms in the objective function while conditional probabilities $q(k|\mathbf{s}_{1:i-1})$ have already been fixed, resulting in the growth of the optimization problems with each step. Second, the outcomes of early uncertainty factors in the sequence influence a larger number of optimization problems and thus have a greater impact on the final scenario probabilities. This means that the sequence should be structured so that early uncertainty factors are unaffected by later ones.

To address the complexity, they propose a solution, which we also employ, to limit the number of uncertainty factors by focusing on relevant dependencies. When estimating the conditional probability of an outcome for a particular uncertainty factor, any factor whose outcome does not affect this probability is deemed irrelevant. An uncertainty factor is irrelevant if and only if its removal does not change the conditional probability of the outcome for another uncertainty factor. More formally, let i be the uncertainty factor whose *k-th* outcome's conditional probability we are estimating. Uncertainty factor a is irrelevant for uncertainty factor i in partial scenario set $S_{1:i}$ if and only if $p(k \mid \mathbf{s}_{1:i-1}) = p(k \mid \mathbf{s}_{1:i-1\setminus a}), \forall k = 1, \ldots, n_i, \forall \mathbf{s}_{1:i-1} \in S_{1:i-1}$, when $\mathbf{s}_{1:i-1\setminus a}$ is the same partial scenario as $\mathbf{s}_{1:i-1}$ but with the uncertainty factor a removed. In other words, uncertainty factor a does not impact uncertainty factor i within the partial scenario set $S_{1:i-1}$ if and only if the random variables X_a and X_i remain conditionally independent across every partial scenario excluding a.

As illustrated by Roponen and Salo (2024), a directed acyclic graph (DAG) elucidates dependencies between different uncertainty factors. In Figure 1, the dependence structure between uncertainty factors weather, umbrella usage, and wetness of a person outside is visualized. The nodes represent uncertainty factors. A connecting arrow implies relevance between the uncertainty factors, while the lack of an arrow indicates irrelevance. The DAG and the dependency structure describe parent-to-child relationships within the nodes. The arrows point from parent-nodes to children-nodes. Based on layman's intuition, the use of an umbrella and the wetness of a person outside will not affect the weather, hence they cannot be the parents of weather. A person who has an umbrella will likely use the umbrella if it rains. Thus, weather is the parent of umbrella usage. Rain will make a person outside wet if she has no umbrella or has but does not use it. On the other hand, the wetness of a person will not affect how much an umbrella is used. Therefore, umbrella usage is the parent of the wetness of a person outside. Weather influences both remaining uncertainty factors although it is only the parent of umbrella usage.

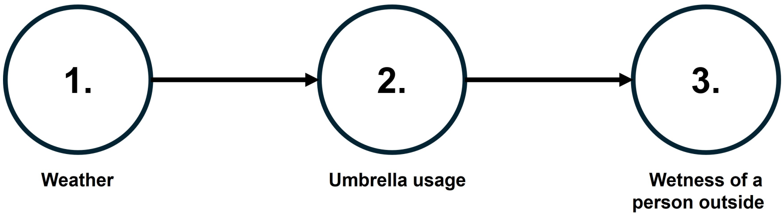

Figure 1: Dependence structure of uncertainty factors in an illustrative DAG example.

Moreover, uncertainty factors can be marginally independent. An uncertainty factor can depend on two parent nodes that are marginally independent. This is illustrated in Figure 2.

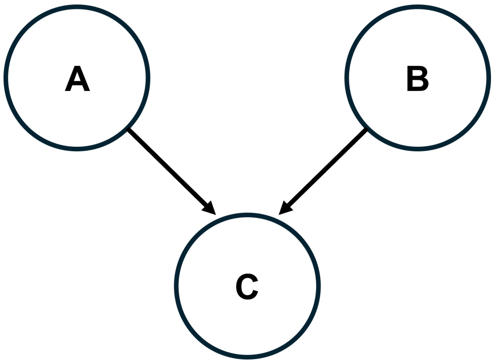

Figure 2: Uncertainty factor C depends on both uncertainty factors A and B. However, A and B do not depend on each other nor on C. Thus, C is the child of uncertainty factors A and B.

Conditional independence can be incorporated into the process of eliciting cross-impact statements in two ways. First, one can build a DAG demonstrated in Figure 1, to express the dependence structure of the uncertainty factors and, subsequently, estimate the cross-impacts only for those uncertainty factors that are dependent. Second, one can directly leave those cross-impacts empty in the cross-impact table, which “do not provide any meaningful information about each other” (Roponen and Salo, 2024). This is illustrated in Table 4.

Table 4: In the "green" product example, uncertainty factors Regulation and Economic development have been denoted irrelevant to each other for the sake of an example.

			Geopolitics			Economic development			Regulation		
			Prolonged conflict between Israel and Palestine leads to blocks in world politics.	Israel and Palestine agree on peace, which calms the atmosphere and increases international co-operation.	Conflict in the Middle-East spreads to neighboring countries, which suppresses global collaboration and freezes global markets.	The EU economy has recovered from high inflation and interest rates.	Slow economic growth, with lingering effects of high inflation and interest rates in the EU.	Moderate economic stability, with the EU maintaining steady but unremarkable growth.	Moderate regulatory support for green products.	Strong regulatory support for green products, with subsidies and incentives.	Comprehensive green standards and international agreements promoting sustainable practices.
			0.40	**0.50**	**0.10**	**0.30**	**0.40**	**0.30**	**0.30**	**0.25**	**0.45**
Geopolitics	Prolonged conflict between Israel and Palestine leads to blocks in world politics.	**0.40**				**-2 (1/2)**	**2 (2)**	**-1 (2/3)**			
	Israel and Palestine agree on peace, which calms the atmosphere and increases international co-operation.	**0.50**				**2 (2)**	**0 (1)**	**2 (2)**			
	Conflict in the Middle-East spreads to neighboring countries, which suppresses global collaboration and freezes global markets.	**0.10**				**-2 (1/2)**	**1 (5/2)**	**-2 (1/2)**			
Economic development	The EU economy has recovered from high inflation and interest rates.	**0.30**							**1 (5/2)**	**2 (2)**	**2 (2)**
	Slow economic growth, with lingering effects of high inflation and interest rates in the EU.	**0.40**							**1 (5/2)**	**1 (5/2)**	**1 (5/2)**
	Moderate economic stability, with the EU maintaining steady but unremarkable growth.	**0.30**							**2 (2)**	**3 (3)**	**2 (2)**
Regulation	Moderate regulatory support for green products.	**0.30**									
	Strong regulatory support for green products, with subsidies and incentives.	**0.25**									
	Comprehensive green standards and international agreements promoting sustainable practices.	**0.45**									

Following Roponen and Salo (2024) in adjusting the model for conditional independence, we also exploit joint irrelevancy of the uncertainty factors (i.e., when two or more uncertainty factors lack relevance to each other, any scenario derived from combining their states becomes insignificant), which is proven by Roponen and Salo (2024). Formally, consider the set of relevant uncertainty factors for a given uncertainty factor i to be D_i, with the associated partial scenario set denoted as $S_{D_i} = \times_{j \in D_i} \{s_1^j, \ldots, s_{n_j}^j\}$. The probability distribution over the partial scenarios in S_{D_i} can be derived from the probability vector p of the set $S_{1:i-1}$ by aggregating over all partial scenarios within $S_{1:i-1}$ that extend the partial scenario $\mathbf{s}_{D_i}$ so that

$$p(\mathbf{s}_{D_i}) = \sum_{\mathbf{s} \in E(\mathbf{s}_{D_i})} p(\mathbf{s}) = \sum_{\substack{\mathbf{s}_{1:i-1} \in S_{1:i-1} \\ E(\mathbf{s}_{1:i-1}) \subseteq E(\mathbf{s}_{D_i})}} p(\mathbf{s}_{1:i-1}).$$

This implies that the probability distribution p over the partial scenario set S_{D_i} is effectively the marginal distribution for the uncertainty factors within D_i. The constraint $E(\mathbf{s}_{1:i-1}) \subseteq E(\mathbf{s}_{D_i})$ implies that any scenario extending $\mathbf{s}_{1:i-1}$ must also extend $\mathbf{s}_{D_i}$, indicating that partial scenarios $\mathbf{s}_{1:i-1}$ and $\mathbf{s}_{D_i}$ yield identical outcomes for all factors in D_i.

Given that factors not in D_i are irrelevant for calculating the conditional probabilities of factor i based on the partial scenarios $\mathbf{s}_{1:i-1}$, the following relationship holds

$$p(s_k^i \mid \mathbf{s}_{1:i-1}) = p(s_k^i \mid \mathbf{s}_{D_i}), \quad \text{if} \quad E(\mathbf{s}_{1:i-1}) \subseteq E(\mathbf{s}_{D_i}). \forall \mathbf{s}_{1:i-1} \in S_{1:i-1}, \mathbf{s}_{D_i} \in S_{D_i}. \tag{15}$$

If the cross-impact statements for pairs i, j are estimated so that $j < i$ and $j \notin D_i$, the optimization problem (11)–(14) can be solved in S_{D_i} instead of $S_{1:i-1}$ such that

$$\min_{q(k|\mathbf{s}_{D_i})} \sum_{j \in D_i} \sum_{(k,l) \in R_{ij}} \left[\left(\sum_{\{\mathbf{s} \in S_{D_i} | s_j = l\}} q(k \mid \mathbf{s}) q(\mathbf{s}) \right) - \frac{\hat{C}_{kl}^{ij} \hat{p}_k^i \hat{p}_l^j}{1 - \hat{p}_k^i + \hat{C}_{kl}^{ij} \hat{p}_k^i} \right]^2 \tag{16}$$

$$\sum_{\mathbf{s} \in S_{D_i}} q(k \mid \mathbf{s}) q(\mathbf{s}) = \hat{p}_k^i, \forall k \in \{1, 2, \ldots, n_i\} \tag{17}$$

$$\sum_{k=1}^{n_i} q(k \mid \mathbf{s}_{D_i}) = 1, \forall \mathbf{s}_{D_i} \in S_{D_i} \tag{18}$$

$$q(k \mid \mathbf{s}_{D_i}) \geqslant 0, \forall k \in \{1, 2, \ldots, n_i\}, \forall \mathbf{s}_{D_i} \in S_{D_i}. \tag{19}$$

Subsequently, probabilities for all partial scenarios in $S_{1:i-1}$ can be attained by using (15). Consequently, the number of scenarios at each iteration depends only on the relevant uncertainty factors D_i instead of all uncertainty factors $\{1, \ldots, i-1\}$.

According to Roponen and Salo (2024), incorporating conditional independence into the methodology introduces auxiliary limitations to the ordering of the uncertainty factors in the iterative process, due to the conditional independence relying on the uncertainty factors that are already in $S_{1:i-1}$. Furthermore, they state that the marginal independence of uncertainty factors, illustrated in Figure 2, has to be taken into account in the order of the calculation. Uncertainty factors that have subsequent uncertainty factors depending on them must be incorporated before the ensuing uncertainty factors.

3.6 Direct scenario probability fitting

The joint probability distribution can be obtained by solving a single optimization problem

$$\min_{q(\mathbf{s})} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{(k,l)\in R_{ij}} \left[\left(\sum_{\substack{\mathbf{s}\in S_{1:N} \\ s_i=k, s_j=l}} q(\mathbf{s}) \right) - \frac{\hat{C}^{ij}_{kl} \hat{p}^i_k \hat{p}^j_l}{1 - \hat{p}^i_k + \hat{C}^{ij}_{kl} \hat{p}^i_k} \right]^2 \tag{20}$$

$$\sum_{\substack{\mathbf{s}\in S_{1:N} \\ s_i=k}} q(\mathbf{s}) = \hat{p}^i_k, \forall i \in \{1, \dots, N\}, k \in \{1, \dots, n_i\} \tag{21}$$

$$0 \leqslant q(\mathbf{s}) \leqslant 1, \forall \mathbf{s} \in S_{1:N}. \tag{22}$$

However, the results of this approach deviate only marginally from the iterative approach proposed earlier. For multiple uncertainty factors, approach (20) fails since the computational capacity of a regular laptop is insufficient due to the optimization problem growing exponentially when new uncertainty factors are introduced. The iterative approach handles larger problems since the size of the optimization problem depends only on the size of the partial scenario set containing the relevant uncertainty factors (Roponen and Salo, 2024). The required computational effort grows linearly if additional uncertainty factors are introduced, and the average number of relevant factors per each new uncertainty factor remains unchanged (Roponen and Salo, 2024). The direct fit approach is adept for smaller problems, while the iterative method suits larger ones.

4 Validation of the revised interpretation

To evaluate whether the odds-ratio-interpretation of the cross-impact term proves to be superior or inferior to the previous approach, we perform a series of statistical tests for the resulting scenario probability distributions across multiple optimization runs. We begin by discussing about the most significant differences between the approaches.

4.1 Differences

In Salo et al. (2022) and Roponen and Salo (2024) the cross-impact between events a and b is defined as $C_{ab} := \frac{P(a|b)}{P(a)}$. In this paper, we use the definition given in (3). The most significant difference between these approaches is eminent when comparing the probability updating functions with different cross-impact multipliers, as seen in Figure 3.

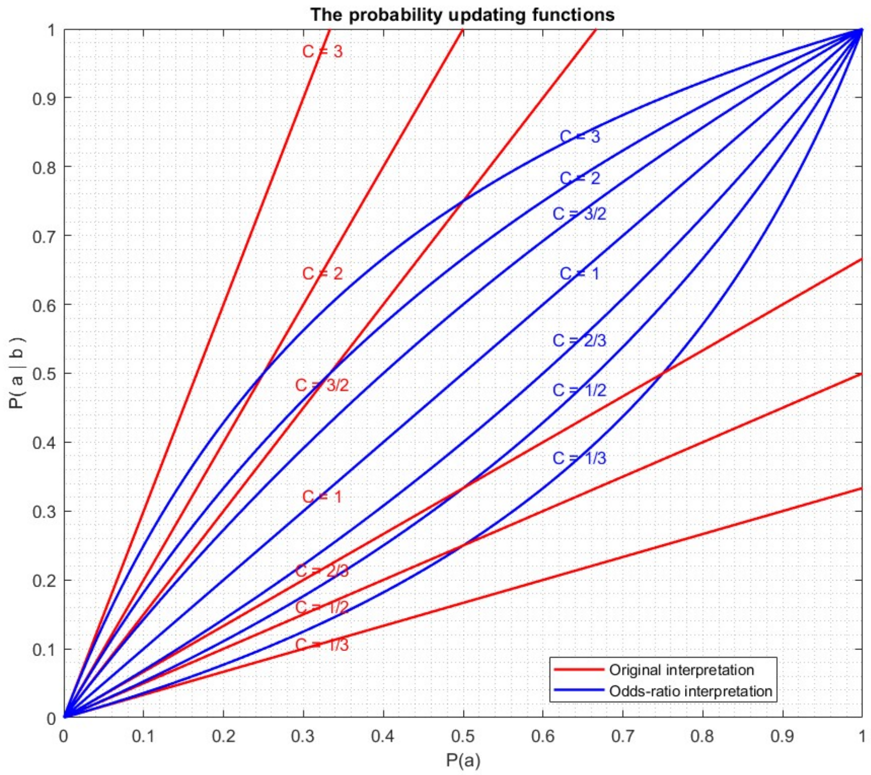

Figure 3: The probability updating functions of both cross-impact interpretations. The straight lines represent probability updates corresponding to the cross-impact interpretation in Salo et al. (2022) and Roponen and Salo (2024). The curved lines represent the probability updates of the odds-ratio interpretation.

Figure 3 demonstrates that the linear mapping $P(a \mid b) = C_{ab}P(a)$ of the cross-impact term has a major caveat, the updated probabilities $P(a \mid b)$ can exceed 100%

if they are not limited. Take for example, a cross-impact statement of +3 between events a and b, which is converted into a cross-impact multiplier of 3 by using (7), and a probability of $P(a) = 50\%$ for event a. The linear probability update would yield a resulting probability of $P(a \mid b) = 150\%$ for the event of a and b both occurring, which is undoubtedly inconsistent with the rules of probability theory. This is solved by setting the optimization framework so that inconsistent probabilities are adjusted accordingly. Clearly, the interpretation in (3) does not have this issue. On the other hand, an attractive property of the original interpretation is that it is symmetric $C_{ab} = \frac{P(a|b)}{P(a)} = \frac{P(a \cap b)}{P(a)P(b)} = \frac{P(b|a)}{P(b)} = C_{ba}$ (Roponen and Salo, 2024). The odds-ratio interpretation does not preserve symmetry. However, we have shown that it could be used as it would preserve symmetry, this was discussed more thoroughly in Section 3.2.

4.2 Statistical tests

We ran 100 optimization runs for both approaches and analyzed the resulting joint probability distributions on each individual run. Subsequently, we repeated the entire process to analyze the most probable 10% of scenarios within each joint probability distribution from each run. We used 11 uncertainty factors, each having three possible outcomes (i.e., 177 147 scenarios) and a maximum of 7 parent-child relationships for each uncertainty factor. The cross-impact terms, marginal probabilities, and the parent-child-relationships were randomly generated on each individual run. The joint probability vectors were sorted in a descending order according to the probabilities of the original approach, such that the most probable scenarios were on the left and least probable on the right. With a 12th Gen Intel(R) Core(TM) i7-1255U 1.70 GHz processor and 15.7 GB of RAM, the statistical tests took roughly 20 minutes to complete. The results are presented in Table 5 and 6 and the statistical tests are explained below the tables.

Table 5: First run. The mean values of the test statistics calculated over every iteration.

Percentage of scenarios	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
JS Divergence	0.017	0.034	0.050	0.067	0.083	0.099	0.115	0.130	0.144	0.241
TV Distance	0.043	0.082	0.121	0.159	0.197	0.233	0.270	0.305	0.341	0.504
KS test p-value	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
KS test rejections	95%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 6: Second run. The mean values of the test statistics calculated over every iteration for the first 10% of the scenarios.

Percentage of scenarios	**1%**	**2%**	**3%**	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
JS Divergence	0.004	0.005	0.006	0.008	0.010	0.011	0.013	0.014	0.016	0.018
TV Distance	0.009	0.012	0.015	0.020	0.023	0.027	0.031	0.035	0.039	0.042
KS test p-value	0.274	0.196	0.127	0.075	0.049	0.031	0.022	0.015	0.008	0.006
KS test rejections	0%	11%	30%	51%	69%	84%	89%	94%	98%	99%

We used the Kolmogorov-Smirnov (KS) test (Kolmogorov, 1933; Smirnov, 1939) to determine the similarity of the two empirical cumulative distribution functions (ECDF) calculated from the joint probability distributions. Despite both approaches producing similar ECDFs for the most probable 10% of the scenarios, we observed highly significant differences with small p-values for the remaining scenarios. With a statistical significance of $\alpha = 5\%$, the null hypothesis (i.e. the ECDFs come from the same probability distributions) was rejected in every iteration when considering the remaining 90% of all scenarios (see Table 5). On the other hand, the null hypothesis was rejected on average in 51% of the cases when factoring in 4% of the scenarios (see Table 6). These observations are also visible in Figure 4 and 5 where the ECDFs are similar for the first 5–10 scenarios whereafter they differ significantly. Recall that roughly 10% of the probability mass is obtained by summing approximately 5–10 of the most probable scenarios out of the 177 147 scenarios, since the joint probability vectors are sorted in a descending order (see Figure 5). These results imply that the probability distributions are quite similar for the most probable scenarios and significantly different for the rest of the scenarios. Moreover, the probability mass seems to be more evenly distributed across all scenarios in the new interpretation. This observation is especially present in Figure 4. One could argue that from a managerial perspective, only the most probable scenarios are relevant for consideration. However, the discussion regarding the correct practice for scenario selection, e.g. preferring most probable scenarios over other scenarios or considering all scenarios relevant, is omitted here.

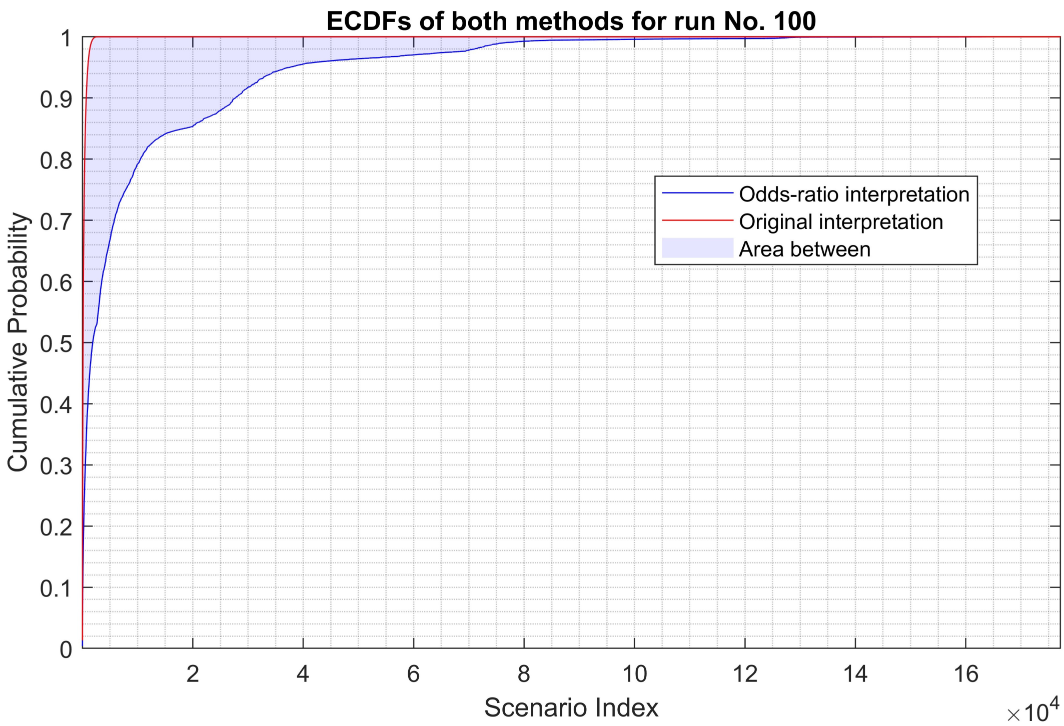

Figure 4: The empirical distribution functions of the 100−*th* optimization run. In most cases, the ECDFs seemed visually very similar to this example.

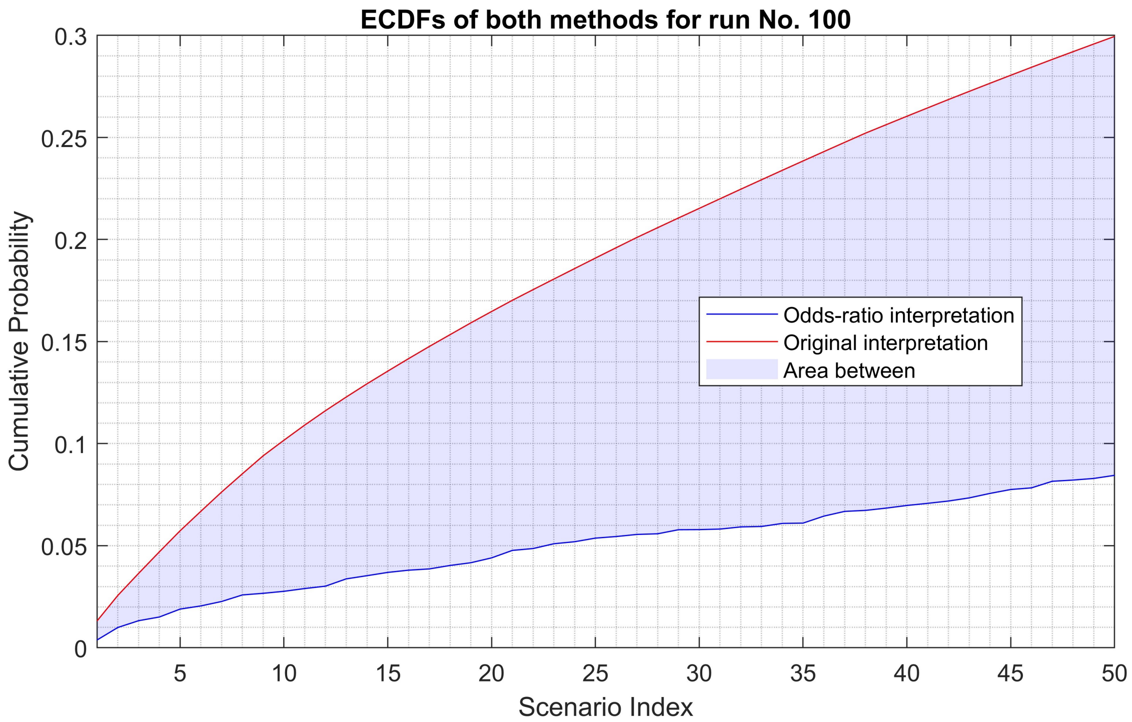

Figure 5: A closer look at the most probable scenarios and the development of the ECDFs. The original method accumulates probability mass much quicker than the new approach. Note that the scenarios are sorted according to the probabilities of the original method's scenarios.

To limit the reliance on individual tests, we also calculated some complementary test statistics such as the Jensen-Shannon (JS) Divergence test and Total Variation Distance (TV) as seen in Tables 5 and 6. The JS divergence is a symmetric measure of the similarity between two probability distributions. Across the 100 optimization runs, the mean JS divergence was especially small for the first 20% of the scenarios and relatively small for the remaining scenarios, indicating small differences for the most probable scenarios and moderate differences for the least probable scenarios. Finally, we employed the TV metric, which measures the greatest absolute difference between the probabilities of corresponding events within two distributions. The mean TV was small for the first 10% of the scenarios and relatively small for the remaining scenarios excluding the least probable 20% of the scenarios. Overall, the least probable 80% of the scenarios are significantly different across all statistical measures while the differences for the most probable 4% were not large enough to warrant significant concern in practical applications. In addition to the statistical tests, we evaluated the computational efficiency of the two approaches. The mean execution time across 1000 optimization runs was 0.9741 for the revised odds-ratio-based method, compared to 1.0 for the original method, indicating negligible differences in computational speed. Both interpretations were relatively similar in terms of their runtime, suggesting no real advantage in terms of performance for either approach.

4.3 Bayesian networks

We demonstrate the possible applications of our approach using the 3D printing case study by Roponen and Salo (2024), which was originally conducted for the Finnish Defence Forces (FDF). This case study allows us to highlight differences in the conditional probabilities for each uncertainty factor's outcome (see Table 3) compared to the original approach. A detailed description of the case study itself is omitted here for brevity.

Leveraging the conditional probability distributions and the conditional independence structures, we also constructed a Bayesian network using the GeNIe Modeler software (BayesFusion, LLC, 2024), as shown in Figure 6. This network structure provides a visual representation of the relationships and dependencies among the uncertainty factors.

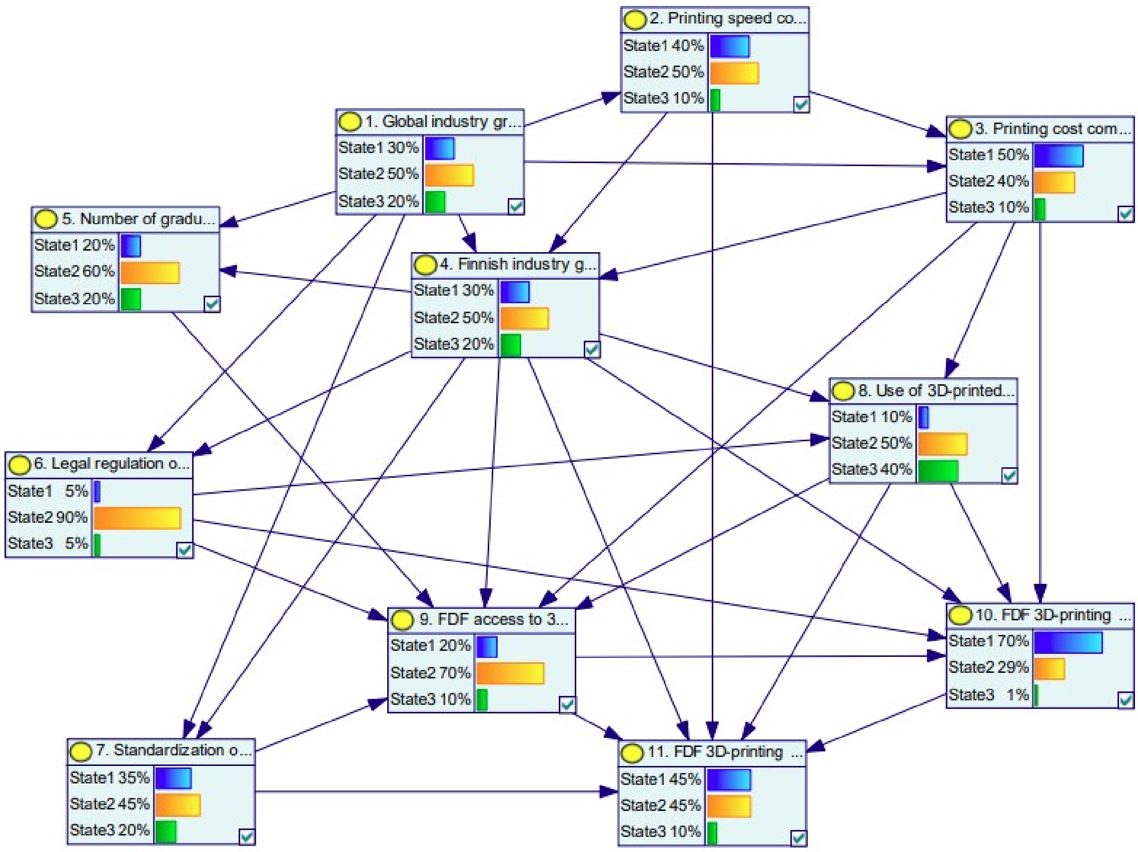

Figure 6: Overview of the generated Bayesian network in GeNIe Modeler software (BayesFusion, LLC, 2024).

We adopted the categorization presented in Roponen and Salo (2024) and organized the uncertainty factors as follows: factors 1–3 represent the global state of the 3D printing industry, factors 4–7 describe the national situation in Finland, and factors 8–11 focus on internal dynamics within the FDF. This structure distinguishes between exogenous factors beyond the direct control of the FDF, those partially influenced through collaboration with the Finnish government and industry, and factors that are endogenous to the FDF. This setup enables a comprehensive exploration of various "what-if" scenarios and their implications (Fenton and Neil, 2001; Roponen and Salo, 2024).

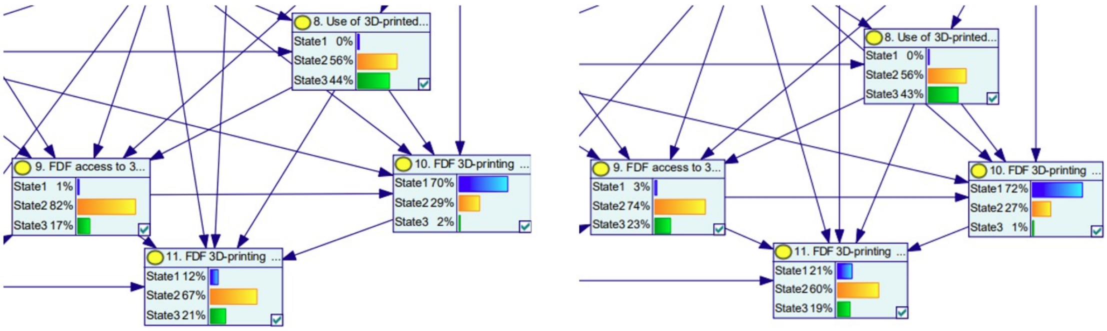

Figure 7: The most likely exogenous scenario. Both interpretations of the cross-impact multipliers, definition of Roponen and Salo (2024) on the left and our definition on the right.

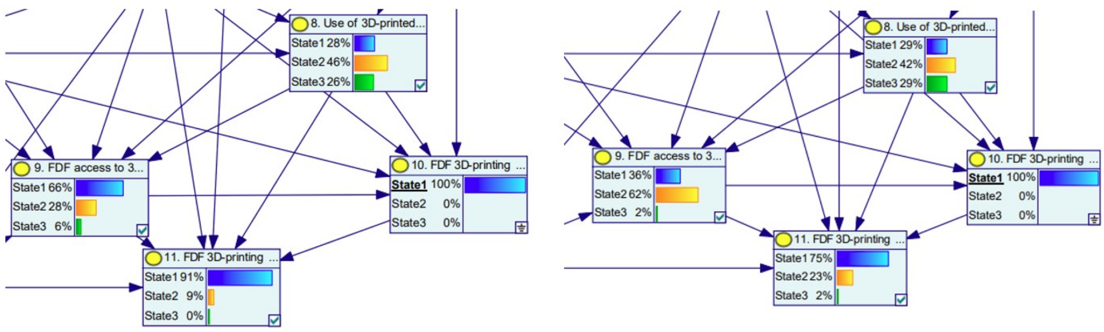

Figure 8: The most pessimistic exogenous scenario. Both interpretations of the cross-impact multipliers, definition of Roponen and Salo (2024) on the left and our definition on the right.

In Figure 7, the uncertainty factors 1–7 are locked into their second state. The comparison between the original and new interpretation shows that both approaches identify specific trends in the probability distributions for key 3D-printing factors within the FDF. In the original methodology (shown on the left), there is a clear concentration of probability for the node "FDF access to 3D-printing model files." However, the new interpretation introduces slightly more variability in this node, indicating a more nuanced distribution of probable outcomes. This suggests that while both methods maintain a similar network structure, the new approach may capture additional variations in dependencies between uncertainty factors. On the other hand, Figure 8 demonstrates the case where uncertainty factors 1–5, 7, and 10 are set to their first outcome, while factor 6 takes its second outcome. This setup represents a constrained growth outlook for the 3D-printing industry. Both approaches show a reduction in the reliance on 3D-printing technology within the FDF, as reflected in the decreased likelihoods for the key nodes. However, the new approach introduces a more balanced distribution across states, particularly for "FDF 3D-printing spare parts in crisis times". This suggests that the new approach offers greater flexibility in modeling uncertain scenarios, leading to a broader range of potential outcomes when external conditions are unfavorable.

5 Discussion and conclusions

In this thesis, we explored whether an alternative cross-impact interpretation would improve a probabilistic cross-impact methodology intended for scenario analysis. For a large part, we followed the methodological development expressed in the original paper by Roponen and Salo (2024) and adjusted the method presented by them for the new interpretation. Subsequently, we performed extensive statistical and visual tests to compare the differences between the two approaches. We compared the resulting conditional probability distribution with Bayesian networks and the resulting joint probability distributions with statistical tests such as the Kolmogorov Smirnov (KS) test and the Jensen-Shannon (JS) Divergence test.

Results showed that the approaches can yield quite different probability distributions on a large scale. On the other hand, the most probable scenarios were similar. The odds-ratio interpretation produced slightly more balanced conditional distributions across different states in the Bayesian network. Moreover, the original interpretation seemed to produce higher probabilities for individual scenarios, whereas the probability mass was more evenly distributed in the odds-ratio interpretation. The differences over the original approach are significant and should be taken into consideration in practical application. These test results indicate that the new interpretation remains stable and valid but differs notably and does not offer significant computational or practical advantages.

Further research could focus on making the methodology more consistent with the original cross-impact statements stated in the elicitation process either by modifying the cross-impact elicitation process or, alternatively, the methodology itself. Additionally, incorporating a larger number of cross-impact relationships into the model could be explored to assess the scalability and robustness of the approach in more complex scenarios. Although our testing was comprehensive, more extensive testing could be conducted for both approaches to identify any conditions under which the odds-ratio interpretation might provide more pronounced advantages or disadvantages.

References

Ş. Y. Balaman. *Decision-making for biomass-based production chains*. Academic Press, 2019.

BayesFusion, LLC. *GeNIe modeler*. *https://www.bayesfusion.com/*, 2024. Retrieved October 21, 2024.

R. Bradfield, G. Wright, G. Burt, G. Cairns, and K. Van Der Heijden. The origins and evolution of scenario techniques in long range business planning. *Futures*, 37 (8):795–812, 2005.

D. W. Bunn and A. A. Salo. Forecasting with scenarios. *European Journal of Operational Research*, 68(3):291–303, 1993.

T. J. Chermack. *Using scenarios: Scenario planning for improving organizations*. Berrett-Koehler Publishers, 2022.

G. Ducot and G.J. Lubben. A typology for scenarios. *Futures*, 12(1):51–57, 1980.

N. Fenton and M. Neil. Making decisions: using Bayesian nets and MCDA. *Knowledge-Based Systems*, 14(7):307–325, 2001.

M. Godet. Scenarios of air transport development to 1990 by SMIC 74—a new cross-impact method. *Technological Forecasting and Social Change*, 9(3):279–288, 1976.

M. Godet. *From anticipation to action: a handbook of strategic prospective*. UNESCO: United Nations Educational, Scientific and Cultural Organisation, 1994.

T. J Gordon. *Cross-impact method*, volume 4. American Council for the United Nations University Washington, DC, USA, 1994.

T. J. Gordon and H. Hayward. Initial experiments with the cross impact matrix method of forecasting. *Futures*, 1(2):100–116, 1968.

O. Helmer. Problems in futures research: Delphi and causal cross-impact analysis. *Futures*, 9(1):17–31, 1977.

O. Helmer. Reassessment of cross-impact analysis. *Futures*, 13(5):389–400, 1981.

W. R. Huss and E. J. Honton. Scenario planning—what style should you use? *Long Range Planning*, 20(4):21–29, 1987.

H. Jungermann. Inferential processes in the construction of scenarios. *Journal of Forecasting*, 4(4):321–327, 1985.

A. Kolmogorov. Sulla determinazione empirica di una legge didistribuzione. *Giorn Dell'inst Ital Degli Att*, 4:89–91, 1933.

H. Kosow and R. Gaßner. *Methods of Future and Scenario Analysis: Overview, Assessment, and Selection Criteria*, volume 39 of *DIE Studies*. Deutsches Institut für Entwicklungspolitik gGmbH, Bonn, 2008.

S. M Millett. Should probabilities be used with scenarios. *Journal of Futures Studies*, 13(4):61–68, 2009.

J. Panula-Ontto. The AXIOM approach for probabilistic and causal modeling with expert elicited inputs. *Technological Forecasting and Social Change*, 138:292–308, 2019.

J. Roponen and A. Salo. A probabilistic cross-impact methodology for explorative scenario analysis. *Futures & Foresight Science*, 6(1):e165, 2024.

A. Salo, E. Tosoni, J. Roponen, and D. W. Bunn. Using cross-impact analysis for probabilistic risk assessment. *Futures & Foresight Science*, 4(2):e2103, 2022.

S. P. Schnaars. How to develop and use scenarios. *Long Range Planning*, 20(1): 105–114, 1987.

T. Seeve and E. Vilkkumaa. Identifying and visualizing a diverse set of plausible scenarios for strategic planning. *European Journal of Operational Research*, 298 (2):596–610, 2022.

N. V. Smirnov. Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin Moscow University*, 2(2):3–16, 1939.

W. Weimer-Jehle. Cross-impact balances: A system-theoretical approach to cross-impact analysis. *Technological Forecasting and Social Change*, 73(4):334–361, 2006.