Received: 12 December 2022 | Revised: 21 June 2023 | Accepted: 24 June 2023

DOI: 10.1002/ffo2.165

ORIGINAL ARTICLE

WILEY

A probabilistic cross-impact methodology for explorative scenario analysis

Juho Roponen | Ahti Salo

Department of Mathematics and Systems Analysis, Aalto-university, Espoo, Finland

Correspondence
Juho Roponen, Department of Mathematics and Systems Analysis, Aalto-university, 00076 Aalto, Finland.
Email: juho.roponen@aalto.fi

Funding information
MATINE - The Scientific Advisory Board for Defence of the Finnish Ministry of Defence

Abstract

As one of the approaches to scenario analysis, cross-impact methods provide a structured approach to building scenarios as combinations of outcomes for selected uncertainty factors. Although they vary in their details, cross-impact methods are similar in that they synthesize expert judgments about probabilistic or causal dependencies between pairs of uncertainty factors and seek to focus attention on scenarios that can be deemed consistent. Still, most cross-impact methods do not associate probabilities with scenarios, which limits the possibilities of integrating them in risk and decision analysis. Motivated by this recognition, we develop a cross-impact method that derives a joint probability distribution over all possible scenarios from probabilistically interpreted cross-impact statements. More specifically, our method (i) admits a broad range of probabilistic statements about the realizations of uncertainty factors, (ii) supports the process of eliciting such statements, (iii) synthesizes these judgments by solving a series of optimization models from which the corresponding scenario probabilities are derived. The resulting scenario probabilities can be used to construct Bayesian networks, which expands the range of analyses that can be carried out. We illustrate our method with a real case study on the impacts of three-dimensional (3D)-printing on the Finnish Defense Forces. The scenarios, their probabilities, and the associated Bayesian network resulting from this case study helped explore alternative futures and gave insights into how the Defence Forces could benefit from 3D-printing.

KEYWORDS
cross-impact analysis, probability estimation, scenario analysis

1 | INTRODUCTION

Over the past decades, scenario analysis has established itself as one of the most widely employed approaches to support long-term planning and strategic management (Chermack, 2022; Scholz & Tietje, 2002). The range of scenario methods spans from purely qualitative (Bowman, 2016; Schwartz, 2012) to purely quantitative (Pereira et al., 2010; Siljander & Ekholm, 2018), with a rich array of methods that combine aspects of both (Godet, 1986; Kemp-Benedict, 2004; Kosow & Gaßner, 2008), whereby this diversity reflects differences in the requirements of different application contexts. The properties of the appropriate method depend on questions such as: (1) What is the context of strategic planning and what are the issues at stake? (2) For what purposes are the scenarios created and to whom are they presented? (3) What are the time horizons involved? (4) What data (qualitative or quantitative) can be obtained to support the formulation of scenarios? Consideration of questions such as these reveals that the requirements on scenarios that are designed to cultivate sound managerial thinking (Lehr et al., 2017) are likely to differ markedly from those that are needed to support risk analyses of safety-critical systems such as nuclear waste repositories (Tosoni et al., 2018).

A persistent challenge faced by scenario modelers is the inherent trade-off between what Kemp-Benedict (2004) calls "complexity" and "complicatedness." That is, the more details are included in the narrative or the description of uncertainty factors,[1] the more detailed (and possibly also more captivating) the scenario is likely to become; but at the same time, the number of scenarios that would be needed to span the full range of uncertainties would grow dramatically (Carlsen et al., 2016). In the face of this inherent trade-off, most methods of qualitative scenario analysis advocate the formulation of relatively few scenarios that are diverse enough to facilitate, for instance, the formulation of robust decision strategies that perform satisfactorily across all scenarios regardless of what the future may bring (Bradfield et al., 2005; Wright & Cairns, 2011). In the exploration of possible future, one of the advantages of quantitative methods such as cross-impact analysis (CIA) is that they are capable of retaining all possible scenarios in the analysis instead of focusing only on a few.

In this paper, we adhere to the cross-impact interpretation introduced in Salo et al. (2021) where cross-impacts are measured in terms of the relative change in the probability of a given event when another event is known to occur. More specifically, we build on the literature which embraces probability theory as a theoretically coherent framework within which cross-impact statements are interpreted. While the seminal contributions on CIA (see, e.g., Dalkey, 1971; Gordon & Hayward, 1968) employ uncertainties with binary outcomes only (which either occur or do not), we follow Salo et al. (2021) and view uncertainty factors as random variables which take on one out of several possible outcomes. Thus, each scenario corresponds to a selection of one of the possible outcomes for every uncertainty factor. That is, because uncertainty factors are not limited to binary outcomes, our proposed approach is flexible enough to admit familiar concepts such as key factors and driving forces that are common in scenarios built using intuitive logics (see, e.g., Phadnis et al., 2014).

For the purpose of deriving scenario probabilities, we consider *all* scenarios that can be formed as combinations of outcomes for uncertainty factors. To some extent, this allows us to address three of the seven concerns with probabilistic approaches as outlined by Derbyshire (2017): determinism, openness, and additivity (the remaining four being nonstationarity, crucial decisions, accurate aggregation, and innate subjectivity). First, the concern of determinism is addressed in that no outcome is considered to be represented by a single scenario only: rather, any outcome can be part of many scenarios which may have different probabilities. Second, the concern with openness refers to the need to elaborate all possible outcomes before probabilities can be assigned. While this is the case in our approach as well, the computation of probability estimates for individual outcomes of uncertainty factors helps verify that the stated outcomes are collectively exhaustive in the sense that they cover all possibilities. Moreover, it is possible to introduce further outcomes to the analysis by (i) introducing new uncertainty factors or (ii) dividing some outcomes of previously introduced uncertainty factors into their constituent parts. This helps address the concern with additivity as the initial scenarios can be transformed into more detailed scenarios while still retaining the validity of previously elicited probability judgments. The other four concerns cannot be readily resolved through the development of computational approaches such as ours. However, they can be arguably addressed through sound methods for expert judgment elicitation (see, e.g., Bolger & Rowe, 2015; Bolger & Wright, 2011; Meyer & Booker, 2001).

Although we use the same interpretation of cross-impacts and scenarios as Salo et al. (2021), the main contribution of the present paper is quite different. Where Salo et al. (2021) present robust optimization methods for risk analysis, our main contribution in this paper is the development of a probabilistic cross-impact method which (i) adopts a well-founded interpretation of cross-impact statements and (ii) derives probability estimates for all possible scenarios in a computationally efficient manner during the elicitation process. The resulting distribution of scenario probabilities can be used for several purposes. For example, it can be employed to assess the relative importance of different scenarios or to produce integrated analyses that involve statistical data sets as well. It can also be employed as an integral part of decision support methods like stochastic optimization (Schneider & Kirkpatrick, 2007), robust optimization (Ben-Tal et al., 2009), and decision analysis (Edwards et al., 2007). We also demonstrate how the scenario probabilities can be used to construct Bayesian networks for problem visualization and what-if analyses.

Furthermore, we describe a case study in which the method was employed to assess the significance of advances in three-dimensional (3D)-printing technology for the Finnish Defence Forces. Additive manufacturing, colloquially referred to as 3D-printing, refers to a wide variety of processes that can be used to construct a 3D object based on a digital file (Kietzmann et al., 2015). The methods by which this is achieved include successively depositing, joining, or solidifying relatively thin material layers. There is plenty of ongoing research and product development on all these methods, which accelerates industry growth (Jiménez et al., 2019). In effect, 3D-printing technology shows a lot of promise for military use with applications ranging from entirely novel production methods to spare part logistics (Booth et al., 2018; Heinen & Hoberg, 2019).

The rest of this paper is structured as follows. Section 2 discusses earlier methods of cross-impact analysis. Section 3 formulates our method and its computations. Section 4 outlines the case study. Section 5 concludes.

2 | METHODS OF CROSS-IMPACT ANALYSIS

The origins of cross-impact methods can be traced to the 1960s when Theodore Gordon and Olaf Helmer developed a game called *Future* for the Kaiser Aluminum and Chemical Company (Gordon, 1994). In this game, uncertain events with a given prior probability were written on cards. A die was then rolled to simulate whether or not that event happened. If it did, the card was flipped over revealing how the probabilities of other events would change as a result. These were the first cross-impacts.

Subsequently, notable contributions to the development and application of cross-impact methods have been made by Gordon (Gordon, 1994; Gordon & Hayward, 1968), Helmer (1977, 1981), and others (Godet, 1976, 1994; Panula-Ontto, 2019). In many of the early methods and their later variants, the probabilities are estimated by considering causal relations between events, even if the temporal occurrence of these events is not necessarily exactly specified. There is the underpinning assumption that the elicitation of conditional probabilities can be instructive in its own right and also potentially cognitively less demanding because, in the elicitation of marginal probabilities, the respondent would need to take an implicit expectation with regard to all the uncertainty factors whose outcomes are not specified for the event whose probability is being elicited. Computationally, many of these cross-impact methods can be viewed as computerized implementations of the original card game, in the sense that the event probabilities define Monte Carlo chains in which the cross-impacts cause changes in event probabilities when a different event is realized.

In principle, the sequences of events in these early Monte Carlo simulation methods could be viewed as scenarios. However, these methods are not well suited for estimating the probabilities of all scenarios, because a very large number of simulation runs would be required to obtain accurate results. This would especially be the case for the scenarios with low probabilities which, by definition, would not appear but in a small fraction of the total number of simulation runs, but could still give rise to high consequences.

Partly in response to this recognition, dedicated scenario probability estimation methods have been developed. One of the first is presented by Dalkey (1971) who computes a feasible set of scenario probabilities for a consistent cross-impact matrix (i.e., a matrix whose elements do not violate the laws of probability theory). Another example is the BASICS tool of Batelle Memorial Institute which computes scenario probabilities using abstracted cross-impact statements instead of strictly probabilistic ones. However, in these methods and even more generally, the specification of cross-impact estimates that are fully consistent with the tenets of probability theory is not easy for any expert (Huss & Honton, 1987).

The third category of cross-impact methods, which we refer to as structural analysis methods, eschew probabilities altogether. These methods seek to identify key scenarios or uncertainty factors based on the strengths of relationships between the factors as quantified on an ordinal scale. As in BASICS, these scales usually do not have a strictly probabilistic interpretation and they can be fully qualitative. Methods in this category include MICMAC (Godet, 1994), Cross-Impact Balances (Weimer-Jehle, 2006), and the consistency analysis method proposed by Seeve and Vilkkumaa (2022), among others. Because these methods do not involve probabilities, they tend to be easy to use and computationally straightforward, making them well-suited for exploratory analyses. However, from a theoretical point of view, without probabilities, they cannot be integrated with probabilistic risk analysis methods or the tenets of expected utility theory.

The method proposed in the present paper is focused on the estimation of scenario probabilities in a setting where some of the cross-impact estimates may be inconsistent. This method is predictable in the sense that the same set of cross-impact estimates always produces the same scenario probabilities, which is in contrast to the presence of some randomness of results obtained by simulation approaches. It also scales up rather well to problems with many uncertainty factors and outcomes, as exemplified by our case study on 3D printing and computational tests in Appendix B.

3 | METHODOLOGICAL DEVELOPMENT

As in Salo et al. (2021), a *scenario* is here defined as a combination of outcomes of uncertainty factors. The uncertainty factors are modeled as discrete random variables $X^i, i = 1, \ldots, N$ which have outcomes $S_i = \{1, \ldots, n_i\}$. Thus, a *scenario* is a vector $\mathbf{s} = (s_1, \ldots, s_N)$, where $s_i \in S_i$ is the outcome for uncertainty factor i. The set of all possible scenarios is the Cartesian product $S \coloneqq S_{1:N} = \times_{i=1}^{N} S_i$. Thus, the number of all possible scenarios is $|S| = \prod_{i=1}^{N} n_i$.

For a subset of uncertainty factors $F \subseteq \{1, \ldots, N\}$, a *partial scenario* is defined as a combination of outcomes for the uncertainty factors which are contained in F. Consequently, the set of partial scenarios for F is $S_F = \times_{i \in F} S_i$. An example of a partial scenario is $\mathbf{s}_{1:i} = (s_1, \ldots, s_i)$ which consists of outcomes for the i first uncertainty factors and thus belonging to the set of partial scenarios $S_{1:i} = \times_{j=1}^{i} S_j$.

When all uncertainty factors belong to F, then, by construction, the set of corresponding partial scenarios coincides with the set of all scenarios. If only the first i uncertainty factors are considered so that $i < N$, then partial scenarios do not cover outcomes for the uncertainty factors $j > i$.

For the purposes of probabilistic analysis, however, any partial scenario $\mathbf{s}_{1:i} = (s_1, \ldots, s_i) \in S_{1:i}$ is compatible with all those (full) scenarios in which the outcomes of the first i uncertainty factors are the same as in the partial scenario $\mathbf{s}_{1:i}$. Thus, any partial scenario $\mathbf{s}_{1:i}$ can be viewed as the collection of those scenarios that can be obtained by extending this partial scenario with outcomes for the uncertainty factors $j = i + 1, \ldots, N$ so that $E(\mathbf{s}_{1:i}) = \left\{ s' \in S | s'_j = s_j, \ \forall j = 1, \ldots, i \right\}$. Furthermore, the probability of the partial scenario $\mathbf{s}_{1:i} \in S_{1:i}, i \leq N$ can be defined as the sum of the probabilities of those scenarios which can be obtained by extending it to full scenarios so that

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

Much in the same way, the marginal probability of the outcome $l \in S_j$ for the jth uncertainty factor is the sum of probabilities for all those scenarios in which this uncertainty factor takes on this outcome, that is,

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

In referring to partial or full scenarios (which correspond to the cases $i < N$ and $i = N$, respectively), we may drop the subscript referring to the number of uncertainty factors. In this case, if $\mathbf{s} \in S_{1:i}$ and $j \in \{1, \ldots, i\}$, then s_j refers to the outcome of the jth uncertainty factor in $\mathbf{s}$.

The above definitions are general in that the uncertainty factors can represent events with binary outcomes (something happens/does not happen) as well as multi-state outcomes (the realization of the ith uncertainty factor is one of n_i possible outcomes). The setup is broad enough to accommodate real-valued random variables, given that the measurement scale for recording possible outcomes can be typically discretized into a set of disjoint and mutually exhaustive intervals. For example, the rise in global temperatures during the 100-year period from 2000 to 2100 can be categorized as low (<3°C), medium (3°C‐5°C), or high (>5°C). Because time is a real-valued variable of this kind, it is also possible to include uncertainty factors whose realization indicates when a given event will occur.

A simplified example of what uncertainty factors and scenarios can look like can be seen in Table 1. The fictive island nation of Soisalo, located in some ocean, currently produces most of its electricity with imported natural gas, but the government is looking to transition to carbon-free electricity. As this transition is only in its early planning stages, an outside observer has estimated the outcomes relating to nuclear power, renewables (wind and solar), and electricity storage as listed in Table 1. Highlighted in blue is the scenario in which Soisalo ends up investing mostly in nuclear power and very little in renewables, but the total number of scenarios that could be constructed from these three uncertainty factors is $3 \times 4 \times 3 = 36$. Any realistic examination of electricity production at a national level would have more than three uncertainty factors, including perhaps other alternatives for energy production or more detailed subdivisions of the alternatives considered here. Yet the size of this illustrative example is kept small on purpose to make it easier to follow when we revisit it later.

TABLE 1 Uncertainty factors, their outcomes, and corresponding marginal probabilities describing the electricity production in the fictive island nation of Soisalo in 2060.

Uncertainty factor	Outcome	Probability
1. Nuclear power	<1 TWh	0.5
	1–5 TWh	0.3
	>5 TWh	0.2
2. Renewables	**<1 TWh**	0.3
	1–4 TWh	0.2
	4–8 TWh	0.3
	>8 TWh	0.2
3. Energy storage	**<10 GWh**	0.5
	10–400 GWh	0.4
	>400 GWh	0.1

Note: The outcomes of nuclear power and renewables describe the yearly production and the outcomes of energy storage describe storage capacity. Highlighted with blue are three outcomes that together form one possible scenario.

In what follows, we develop our method for estimating scenario probabilities in four parts. Section 3.1 presents the basic definitions used for cross-impact multipliers. Section 3.2 describes the basic method for computing the scenario probabilities. Section 3.3 explains how conditional independence information can be harnessed to improve the speed and accuracy of the estimation process. Section 3.4 shows how to build Bayesian networks using conditional independence information and computed probabilities. Finally, Section 3.5 discusses the limitations of the method and its computational properties.

3.1 | Cross-impact multipliers

One rationale for the cross-impact analysis is that the number of scenarios is often so large that it is practically impossible to elicit scenario probabilities directly. For example, from 11 uncertainty factors with three possible outcomes each, it is possible to define a total of $3^{11} = 177,147$ distinct scenarios. In this setup, instead of characterizing the scenario probabilities directly, methods of probabilistic cross-impact analysis characterize probabilistic dependencies between uncertainty factors through cross-impacts and use such characterizations to infer information about the scenario probabilities.

In this paper, we employ the cross-impact interpretation in Salo et al. (2021), which we call the cross-impact multiplier approach in which the cross-impact between events a and b is defined as

$$C_{ab} := \frac{P(a|b)}{P(a)}, \tag{3}$$

meaning that

$$P(a|b) = C_{ab} P(a). \tag{4}$$

Thus, the cross-impact multiplier specifies how many times more likely the occurrence of the event a becomes when the event b is known to occur. Here, it is worth pointing out that the expressions in (4) do not refer to the temporal sequence in which the events would occur. For instance, it could be the case that the event b will occur after the event a within the time horizon of interest. Still, within such a time horizon, the probability of the event a may be higher when it is known that the event b, too, will come about.

Normally, the cross-impact multipliers are placed into a matrix. Table 2 shows the cross-impact multipliers for the energy production example for Soisalo, as estimated by our assumed outside observer. There is empirical evidence suggesting that humans are more adept

TABLE 2 A cross-impact matrix containing the cross-impact multipliers for the dependencies between the uncertainty factors in Table 1.

			Nuclear power			Renewables				Energy storage		
			< 1 TWh	1–5 TWh	> 5 TWh	< 1 TWh	1–4 TWh	4–8 TWh	> 8 TWh	< 10 GWh	10–400 GWh	> 400 GWh
			0.5	0.3	0.2	0.3	0.2	0.3	0.2	0.5	0.4	0.1
Nuclear power	< 1 TWh	0.5				$\frac{1}{4}$	1	$\frac{5}{4}$	$\frac{7}{4}$	$\frac{3}{4}$	$\frac{5}{4}$	$\frac{5}{4}$
	1–5 TWh	0.3				$\frac{3}{2}$	$\frac{3}{2}$	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{3}{2}$	$\frac{3}{4}$	$\frac{1}{10}$
	> 5 TWh	0.2				$\frac{3}{2}$	$\frac{3}{2}$	1	$\frac{1}{10}$	$\frac{7}{4}$	$\frac{1}{4}$	$\frac{1}{10}$
Renewables	< 1 TWh	0.3								$\frac{7}{4}$	$\frac{1}{4}$	$\frac{1}{10}$
	1–4 TWh	0.2								$\frac{5}{4}$	1	$\frac{1}{4}$
	4–8 TWh	0.3								$\frac{3}{4}$	$\frac{3}{2}$	$\frac{1}{2}$
	> 8 TWh	0.2								$\frac{1}{10}$	1	4
Energy storage	<10 GWh	0.5										
	10–400 GWh	0.4										
	> 400 GWh	0.1										

at estimating relative magnitudes than providing numerical values (Gallistel & Gelman, 1992). Thus, one motivation for employing the above relative change in probability is that statements about the above ratio (3) may be easier for the experts to provide than estimates about the conditional probabilities as such.

An appealing property of the cross-impact multiplier is that it is symmetric

$$C_{ab} = \frac{P(a|b)}{P(a)} = \frac{P(a \wedge b)}{P(a)P(b)} = \frac{P(b|a)}{P(b)} = C_{ba}. \tag{5}$$

This reduces the number of cross-impacts to be estimated in half because these multipliers are the same in either direction. This can also be seen Table 2 in which the lower-left half is shown in gray. These gray cells would be equal to those in the top-right half and thus they need not be considered separately.

Unlike some earlier cross-impact methods, such as the seminal work by Helmer (1981) and Gordon (1994), our interpretation of cross-impacts is not limited to causal relations but reflects also other types of probabilistic dependencies. Specifically, while the presence of a causal relationship does give rise to probabilistic dependence, all probabilistic dependencies between pairs of uncertainty factors cannot be attributed to direct causality between the two uncertainty factors. This would be the case, for instance, when there is a shared underlying cause for two distinct events which are not causally related to each other. To illustrate, consider a situation where two different kinds of alarm systems have been installed for fire detection, one for heat detection and the other for smoke detection. Then neither one of the alarms would cause the other to go off, yet the two alarms would be related to each other in the sense that there would be a positive correlation between them. Similarly, in Table 2, the cross-impacts between outcomes relating to nuclear power and energy storage are not be interpreted as direct causal relations one way or another.

However, enforcing the strict mathematical interpretation of cross-impact estimates can be challenging, because the experts often struggle to provide cross-impact multipliers that satisfy the laws of probability theory. For example, the seemingly reasonable multipliers in Table 2 violate these laws, as shown by Table 3 which contains the joint probabilities of outcomes relating to nuclear power and renewable energy. The sum of the joint probabilities is 1.014 even if every probability distribution should sum up to exactly 1. Furthermore, the marginal probabilities implied by this joint distribution deviate from the original estimates and from probabilities calculated from joint distributions of other uncertainty factor pairs.

There are two possible approaches to addressing this issue. One is to facilitate the elicitation of mathematically consistent cross-impact multipliers by proceeding incrementally and by using software tools to support the elicitation process at each stage of the analysis (Salo et al., 2021). In this paper, we take the other approach, realizing that the elicited estimates may not be mutually consistent, but that they are still informative so it is instructive to use them for deriving a probability distribution that, by construction, is mathematically consistent and also fits the elicited estimates as closely as possible.

TABLE 3 The joint probabilities of outcomes relating to nuclear power and renewable energy are shown at the intersections of rows and columns.

			Renewables				
			< 1 TWh	1–4 TWh	4–8 TWh	> 8 TWh	
			0.3	0.2	0.3	0.2	
Nuclear power	< 1 TWh	0.5	0.0375	0.100	0.1875	0.175	0.500
	1–5 TWh	0.3	0.135	0.090	0.060	0.015	0.300
	> 5 TWh	0.2	0.090	0.060	0.060	0.004	0.214
			0.2625	0.2500	0.3075	0.194	

Note: Next to the outcome is the original marginal probability estimate. The marginal probabilities calculated from the joint distribution are shown outside the matrix.

In the methodological development that follows, we employ the notation $p(\cdot)$ to refer to the underlying probability distribution $P(\cdot)$ over scenarios. Estimates about scenario probabilities are indicated through $\hat{p}(\cdot)$, where the argument specifies which scenarios are being considered. More generally, all variables with a hat represent estimates which are used as input variables. Thus, for example, for a given outcome $s_i \in S_i$, the $\hat{p}(s_i)$ is the elicited estimate about the marginal probability $p(s_i)$. The probabilities which are derived from the estimates through computations are indicated by $q(\cdot)$. These distinctions are useful in that it is possible, for example, the explore conditions under which the computed probabilities are guaranteed to converge to the true underlying probabilities.

For any $k \in S_i$ and $l \in S_j$, we introduce the abbreviated notations $p^i_k := P(X^i = k)$ and $p^j_l := P(X^j = l)$ for the marginal probabilities and $p^{i|j}_{k|l} = P(X^i = k | X^j = l)$ for the conditional probability. From (4), we get

$$p^{i|j}_{k|l} = C^{ij}_{kl} p^i_k \Leftrightarrow p^{ij}_{kl} = C^{ij}_{kl} p^i_k p^j_l \tag{6}$$

where $p^{ij}_{kl} = P(X^i = k, X^j = l)$ and C^{ij}_{kl} is the cross-impact multiplier for the outcome pair in which the outcome for the uncertainty factor i is k and that for the uncertainty factor j is l.

3.2 | Conditional probability updating

The probabilistic approach proposed by Salo et al. (2021) invites the respondent to specify lower and upper bounds for the cross-impact multiplier (4) and then converts these bounds on the scenario probabilities $p(\mathbf{s}_{1:N})$, $\forall\, \mathbf{s}_{1:N} \in S_{1:N}$. Together with estimates of the expected consequences in each of the scenarios, lower and upper bounds for the expected disutility are then derived to provide an aggregate measure of risk. In particular, it is consequently possible to verify whether or not the risk level of the systems is acceptable.

A limitation of this approach is that it presumes that the cross-impact statements elicited from the respondents remain fully consistent (i.e., for any given set of statements that have been elicited from the respondent, there exists at least one assignment of probabilities to all scenarios such that the constraints which correspond to these statements are satisfied). To guide the respondent in providing such statements, however, it is necessary to solve an optimization problem with quadratic constraints which will give rise to computational challenges when the number of uncertainty factors is large. More generally, this approach is not suitable for synthesizing a set of possibly inconsistent cross-impact statements to determine a single probability distribution.

Against this backdrop, one of the main contributions of this paper lies in developing a computationally efficient method that (i) admits cross-impacts statements, including inconsistent ones, as well as many other forms of statements that correspond to constraints on scenario probabilities and (ii) synthesizes such statements into a single probability distribution over scenarios in such a way that the resulting distribution represents the best fit to the statements. The estimates from which the scenario probabilities are derived consist of marginal probabilities $\hat{p}^i_k, \hat{p}^j_l$ for all uncertainty factors and their outcomes and cross-impact multipliers $\hat{C}^{ij}_{kl}$ for selected pairs of uncertainty factors and their outcomes. Specifically, the probability distribution over scenarios can be derived even in the absence of information about some pairs of cross-impact multipliers.

To motivate the approach, assume that estimates about the marginal probabilities $\hat{p}^i_k, \hat{p}^j_l$ as well as the cross-impact multiplier $\hat{C}^{ij}_{kl}$ have been elicited. If these estimates are correct in the sense that $\hat{p}^i_k = p^i_k$, $\hat{p}^j_l = p^j_l$ and $\hat{C}^{ij}_{kl} = C^{ij}_{kl}$, the probability $p^{ij}_{kl} = P(X^i = k, X^j = l)$ is equal to $\hat{C}^{ij}_{kl}\hat{p}^i_k\hat{p}^j_l$. Equivalently, however, this same probability can be expressed as the sum of probabilities for all those scenarios such that $X^i = k$ and $X^j = l$. Thus, we have the following constraint on scenario probabilities

$$\sum_{\substack{\mathbf{s}\in S_{1:N}\\ s_i=k, s_j=l}} p(\mathbf{s}) = \hat{C}^{ij}_{kl}\hat{p}^i_k\hat{p}^j_l, \tag{7}$$

where the summation on the left side is taken over those scenarios whose outcomes for the ith and jth uncertainty factors match those on the right side of the equality.

To consider the situation where several (but not necessarily all) cross-impact multipliers and all marginal probabilities have been specified, assume 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 estimate $\hat{C}^{ij}_{kl}$ for the cross-impact multiplier (4) is available. In this case, the above term would appear for all pairs of outcomes such that $R_{ij}(s_i, s_j)$, suggesting that the probability distribution that best matches these estimates can be obtained by solving the minimization problem

$$\min_{p(s)} \sum_{i=2}^{N}\sum_{j=1}^{i-1}\sum_{(k,l)\in R_{ij}} \left[\left(\sum_{\substack{\mathbf{s}\in S_{1:N}\\ s_i=k, s_j=l}} p(\mathbf{s})\right) - \hat{C}^{ij}_{kl}\hat{p}^i_k\hat{p}^j_l\right]^2 \tag{8}$$

Computationally, however, a concern with the problem (8) is that the optimization would need to be carried out over *all* scenarios. This will be challenging if the number of scenarios is large, either because there are many uncertainty factors or if these factors have several possible outcomes (recall that the total number of scenarios is $\prod_{i=1}^{N} n_i$, where n_i is the number of possible outcomes for the ith uncertainty factor).

However, the probability of any scenario $s \in S_{1:N}$ can be written as by conditioning the realization of the ith uncertainty factor on the partial scenario defined by $i - 1$ preceding uncertainty factors, that is,

$$p(\mathbf{s}) = p(s_i|\mathbf{s}_{1:N\setminus i})p(\mathbf{s}_{1:N\setminus i}) \tag{9}$$

where $\mathbf{s}_{1:N\setminus i}$ is the partial scenario that contains the outcomes for all uncertainty factors except the ith one. In particular, if the terms $p(\mathbf{s}_{1:N\setminus i})$ representing probabilities for the partial scenarios excluding the ith uncertainty factor are known, then the estimation of the scenario probabilities becomes a significantly smaller problem in that it is necessary to only consider cross-impact multipliers $\hat{C}_{kl}^{ij}$ that relate to the uncertainty factor i to estimate the conditional probabilities $p(s_i|\mathbf{s}_{1:N\setminus i})$.

Hence, for a given ordering of the uncertainty factors, the relationship (9) can be exploited to build

$$\begin{aligned} p(s) = p(s_N|\mathbf{s}_{1:N-1})p(\mathbf{s}_{1:N-1}) &= p(s_N|\mathbf{s}_{1:N-1})p(s_{N-1}|\mathbf{s}_{1:N-2}) \\ &\quad p(\mathbf{s}_{1:N-2}) = \cdots \\ = p(s_N|\mathbf{s}_{1:N-1})p(s_{N-1}|\mathbf{s}_{1:N-2})\dots p(s_2|s_1)p(s_1), \end{aligned} \tag{10}$$

This relationship leads to the recognition that the scenario probabilities can be derived iteratively by (i) starting from the marginal probabilities for the first uncertainty factor $p(s_1)$, (ii) computing the conditional probabilities $p(s_2|s_1)$ which represent the best fit to the cross-impact multipliers for the outcomes of the two first uncertainty factors, and (iii) using these conditional probabilities to estimate the probabilities for partial scenarios which comprise these two uncertainty factors. After this step, the iteration can proceed to the third uncertainty factor and so on until all uncertainty factors have been reached.

Thus, the procedure can be described as follows:

1. Use previously computed probabilities for partial scenarios and estimates about marginal probabilities and cross-impact multipliers to compute the conditional probabilities for the next uncertainty factor whose outcomes are conditioned on the previously analyzed partial scenarios.
2. Generate the updated set of partial scenarios which includes this new uncertainty factor. The number of these partial scenarios is equal to the product of (i) the number of partial scenarios in the previous iteration and (ii) the number of outcomes for the new uncertainty factor.
3. Use the computed conditional probability distributions to compute the joint probability distribution for the updated set of partial scenarios.

More formally, the iteration can be carried out by computing probabilities for all partial scenarios $\mathbf{s}_{1:i}$, $i = 1, \dots, N$ and with the help of conditional probabilities $q(s_i|\mathbf{s}_{i:i-1})$ such that the iteration is initialized by setting $q(k) \leftarrow \hat{p}_k^1$ for any $k \in S_1 = \{1, \dots, n_1\}$. At each step of the ensuing iteration, the conditional probabilities can be computed from

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

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

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

$$q(k|\mathbf{s}_{1:i-1}) \geq 0.\ \forall\ k \in \{1, 2, \dots, n_i\},\ \ \mathbf{s}_{1:i-1} \in S_{1:i-1} \tag{14}$$

In the third summation of the objective function, the sum is taken over those partial scenarios in which the state of the jth uncertainty factor is equal to the outcome specified by the term in the relation R_{ij}. The last two constraints ensure that the conditional probability distribution is well-defined. The computed probabilities for the next partial scenarios (which are constructed by appending the states of the ith uncertainty factor $k \in S_i$ to the previous partial scenarios $\mathbf{s}_{1:i-1}$) can be defined by $q((\mathbf{s}_{1:i-1}, k)) \leftarrow q(\mathbf{s}_{1:i-1})q(k|\mathbf{s}_{1:i-1})$. Thus, constraint (12) ensures that the marginal probability is the same as the estimated marginal probability $\hat{p}_k^i$ of the outcome $s_i = k$, ensuring that the computed probabilities match the estimated marginal probabilities exactly.

Table 4 shows how the iteration proceeds in the Soisalo example. The first iteration is not conditioned on previously defined uncertainty factors and thus involves the marginal probability distribution only. In the second iteration, conditional probabilities based on alternative outcomes for the first uncertainty factor are calculated. In the third iteration, conditional probabilities based on the consideration of outcomes for the first and second uncertainty factors are derived. As seen from Table 4, the number of conditional probabilities calculated grows at every step of the iteration. This may cause problems, not only because of the amount of computation required but also because cross-impacts are not well-suited for describing highly multivariate probability distributions. Section 3.3 discusses how exploiting conditional independence can mitigate both these issues.

The reason why cross-impacts are not sufficient for the characterization of multivariate distributions fully is that the number of conditional probabilities $q(k|\mathbf{s}_{1:i-1})$ becomes significantly higher than the number of cross-impact multipliers which appear in the objective function. This is the case especially when only a fraction of all cross-impact multipliers have been elicited. If estimates about all cross-impact multipliers have been elicited, there are $n_i \sum_{j=1}^{i-1} n_j$ terms in the objective function (11). Equation (12) gives rise to n_i constraints and Equation (13) has $\prod_{j=1}^{i-1} n_j$ constraints. The number of parameters

TABLE 4 The calculated conditional probabilities for every uncertainty factor in the Soisalo example.

1st iteration		-
1. Nuclear power	$p(s_1^1)$	0.50
	$p(s_2^1)$	0.30
	$p(s_3^1)$	0.20

2nd iteration		1. Nuclear power		
		s_1^1	s_2^1	s_3^1
2. Renewables	$p(s_1^2 \mid s^1)$	0.10	0.50	0.50
	$p(s_2^2 \mid s^1)$	0.17	0.25	0.21
	$p(s_3^2 \mid s^1)$	0.37	0.20	0.28
	$p(s_4^2 \mid s^1)$	0.36	0.06	0.02

3rd iteration		1. Nuclear power											
		s_1^1				s_2^1				s_3^1			
		2. Renewables											
		s_1^2	s_2^2	s_3^2	s_4^2	s_1^2	s_2^2	s_3^2	s_4^2	s_1^2	s_2^2	s_3^2	s_4^2
	$\mathbf{s}_{D_3}$	(s_1^1, s_1^2)	(s_1^1, s_2^2)	(s_1^1, s_3^2)	(s_1^1, s_4^2)	(s_2^1, s_1^2)	(s_2^1, s_2^2)	(s_2^1, s_3^2)	(s_2^1, s_4^2)	(s_3^1, s_1^2)	(s_3^1, s_2^2)	(s_3^1, s_3^2)	(s_3^1, s_4^2)
3. Storage	$p(s_1^3 \mid \mathbf{s}_{D_3})$	0.81	0.62	0.29	0.07	0.88	0.49	0.36	0.20	0.88	0.70	0.57	0.02
	$p(s_2^3 \mid \mathbf{s}_{D_3})$	0.19	0.38	0.71	0.51	0.12	0.51	0.64	0.29	0.10	0.30	0.25	0.02
	$p(s_3^3 \mid \mathbf{s}_{D_3})$	0.00	0.00	0.00	0.42	0.00	0.00	0.00	0.51	0.02	0.00	0.18	0.96

is $\prod_{j=1}^{i} n_j$, the same as the number of partial scenarios of length i, and thus the algebraic equation system (12)–(14) is underdetermined, and thus, multiple optimal solutions may exist.

In general, the specific distribution generated by the optimization algorithm depends on the method used to solve the problem as well as the specific starting point used. We used Matlab's built-in solver's interior point method with all conditional probabilities $q(k|\mathbf{s}) = 1$ as a starting point. This produces results close to the uniform distribution, which seems appropriate in the absence of explicitly stated information about dependencies.

However, the nonuniqueness of the implied probability distribution is often not the only problem in using cross-impacts. An attractive property of the above optimization formulation is that it is capable of handling situations where the cross-impact statements are not consistent. In this case, at least some of the estimates $\hat{C}_{kl}^{ij}, \hat{p}_k^i, \hat{p}_l^j$ differ from the cross-impact multipliers and marginal probabilities implied by the computed probabilities $q(\mathbf{s}), \mathbf{s} \in S_{1:N}$. The implied cross-impacts $\mathring{C}_{kl}^{ij}$ can be obtained from the computed probabilities as

$$\mathring{C}_{kl}^{ij} = \frac{1}{\hat{p}_k^i \hat{p}_l^j} \sum_{\{\mathbf{s} \in S \mid s_i = k, s_j = l\}} q(\mathbf{s}). \tag{15}$$

This recognition is useful in that it can be harnessed to support the identification and possible revision of those cross-impact estimates which differ most from the implied cross-impacts, either in absolute terms or in terms of the probabilities for the joint event $X^i = k, X^j = l$ that appears in the objective function (11). Because the marginal probabilities are matched exactly, the cross-impact terms for which the following term is maximized

$$\underset{\substack{i,j \in \{1,\dots N\} \\ (k,l) \in R_{ij}}}{\operatorname{argmax}} \left| \hat{C}_{kl}^{ij} - \mathring{C}_{kl}^{ij} \right| \tag{16}$$

is the one that deviates most from the implied cross-impact multiplier based on the derived scenario probabilities $q(s)$. On the other hand, the solution

$$\underset{\substack{i,j \in \{1,\dots N\} \\ (k,l) \in R_{ij}}}{\operatorname{argmax}} \left| \left(\hat{C}_{kl}^{ij} - \mathring{C}_{kl}^{ij} \right) \hat{p}_k^i \hat{p}_l^j \right| \tag{17}$$

helps identify the cross-impact multiplier for which there is the greatest discrepancy between the estimated probability of the event $X^i = k, X^j = l$ and that of the computed probabilities. This analysis can thus be employed to identify and, if need be, revise inconsistent cross-impact multipliers.

Furthermore, the implied cross-impacts can be used to explore which probability distributions other than $q(\mathbf{s})$ would match the given expert judgments $\hat{C}_{kl}^{ij}, \hat{p}_k^i, \hat{p}_l^j$ equally well. If the cross-impact terms $\mathring{C}_{kl}^{ij}$ that are implied by the computed scenario probabilities are assigned back to (8) instead of $\hat{C}_{kl}^{ij}$, the solution $p(\mathbf{s}) = q(\mathbf{s})$ will make all the sum terms equal to 0, but often $q(\mathbf{s})$ is not unique in this regard. To explore other equally feasible distributions, an optimization problem can be formulated

$$\min_{\mathring{q}(\mathbf{s})} f(\mathring{q}) \tag{18}$$

$$\sum_{\substack{\mathbf{s} \in S_{1:N} \\ s_i = k, s_j = l}} \mathring{q}(\mathbf{s}) = \mathring{C}_{kl}^{ij} \hat{p}_k^i \hat{p}_l^j, \ \forall\ i \in \{2, \dots, N\}, j \in \{1, \dots, i-1\}, (k,l) \in R_{ij} \tag{19}$$

$$\sum_{\substack{\mathbf{s} \in S_{1:N} \\ s_i = k}} \mathring{q}(\mathbf{s}) = \hat{p}_k^i, \forall i \in \{1, ..., N\}, k \in \{1, ..., n_i\} \tag{20}$$

$$\mathring{q}(\mathbf{s}_{1:N}) \geq 0, \forall \mathbf{s}_{1:N} \in S_{1:N} \tag{21}$$

where $f : \mathbb{R}^{|S|} \to \mathbb{R}$ is chosen to find a scenario probability distribution $\mathring{q}$ with specific properties. To give a few examples, $f(\mathring{q}) = -\mathring{q}(\mathbf{s}^*)$ will maximize the probability of a specific scenario s^* and $f(\mathring{q}) = \sum_{\mathbf{s} \in S} \left(\frac{1}{|S|} - \mathring{q}(\mathbf{s})\right)^2$ will find the distribution closest to the uniform distribution when $|S|$ is the total number of possible scenarios. Finding expected utility maximizing or minimizing scenario probability distributions is also possible if the utilities of all the scenarios are known. Because all the constraints are linear, the optimization problem can be solved with commonly used optimization tools as long as $f(\mathring{q})$ is convex. These cross-impact constraints could also be used in conjunction with the methods presented in Salo et al. (2021) to analyze system risk or with some different robust optimization methods.

3.3 | Conditional independence

When there are many uncertainty factors, there are inherent limitations in using cross-impact multipliers to estimate all possible scenario probability distributions. To illustrate this, assume that the N uncertainty factors have equally many outcomes n (i.e., $n_i = n, i = 1, ..., N$). Then the number of different cross-impact multipliers that can be elicited is $\sum_{i=1}^{N} \sum_{j=i+1}^{N} n_i n_j = N(N-1)n^2/2$. This is proportional to the square of the number of uncertainty factors and their outcomes. Still, the number of scenarios $\prod_{i=1}^{N} n_i = n^N$ grows exponentially with the number of uncertainty factors. This implies that estimates about marginal probabilities and cross-impact multipliers will not suffice to fully characterize all possible scenario probabilities, because the number of constraints implied by the cross-impact multipliers will be much lower than the number of scenarios.

Against this backdrop, two observations on the optimization problem in (11)–(14) are in order. First, the number of estimates about cross-impact multipliers increases in every step of the iterative algorithm because the n_i outcomes of the new uncertainty factor are compared with the previously considered uncertainty factors. That is, if all these estimates have been provided, there are $n_i \sum_{j=1}^{i-1} n_j/2$ terms in the objective function (11) while the conditional probabilities $q(k|\mathbf{s}_{1:i-1})$ have already been fixed. As a result, the optimization problems grow in size at every step of the process.

Second, the outcomes of those uncertainty factors which appear earlier on in the sequence of uncertainty factors appear in a larger number of optimization problems. As a result, they likely exert more influence on the final scenario probabilities. More specifically, the probabilities for the partial scenarios defined by the uncertainty factors in the early part of the sequence will not be impacted by the cross-impact terms in the latter part of the sequence. In qualitative terms, this implies that the sequence should be developed so that the uncertainty factors in the early part of the sequence should not be impacted by the later uncertainty factors.

A way to solve this problem of increased complexity is to limit the number of uncertainty factors in every iterative step by focusing only on the relevant dependencies. Specifically, when estimating the conditional probability of the outcome k for the ith uncertainty factor $p(k|\mathbf{s}_{1:i-1})$, any uncertainty factor whose outcome does not affect this conditional probability is irrelevant. That is, an uncertainty factor a is *irrelevant* for uncertainty factor i in partial scenario set $S_{1:i}$ if and only if $p(k|\mathbf{s}_{1:i-1}) = p(k|\mathbf{s}_{1:i-1\setminus a})$, $\forall\ k = 1, ..., n_i$, $\forall\ \mathbf{s}_{1:i-1} \in S_{1:i-1}$, when $\mathbf{s}_{1:i-1\setminus a}$ is the same partial scenario as $\mathbf{s}_{1:i-1}$ but with uncertainty factor a removed. Equivalently, uncertainty factor a is *irrelevant* for i in partial scenario set $S_{1:i}$, if and only if, random variables X_a and X_i are conditionally independent in every partial scenario $\mathbf{s}_{1:i-1\setminus a} \in S_{1:i-1\setminus a}$.

The conditional dependencies between uncertainty factors can be visualized with a directed acyclic graph. Figure 1 illustrates this with the uncertainty factors from the Soisalo energy production example. The uncertainty factors are represented by nodes drawn as circles. The edges connecting the nodes, drawn as arrows, indicate that uncertainty factors are relevant to each other, whereas the lack of a connecting edge implies *irrelevance*. This is how conditional independence between variables is represented in graphical models such as Bayesian networks (Pearl & Paz, 2022). Figure 1 shows that uncertainty factors representing nuclear power and energy storage have been deemed irrelevant to one another, provided that the outcome for renewables is known. The rationale for this is that energy storage is not a prerequisite for using nuclear power. However, as the share of renewable energy production increases, there is an increased need for load-balancing: because the generation of solar and wind power may be highly variable, energy storage is one way to meet that need. The demand for electricity in Soisalo is finite, so it is unlikely that both nuclear and renewables are built in large quantities.

Incorporating conditional independence in the expert judgment elicitation process can be done in two ways. The first is to start by constructing a directed acyclic graph that depicts the dependence structure of the uncertainty factors and then collecting cross-impact information on the directly connected uncertainty factors. The second way is to give the experts the option to state that some pairs of uncertainty factors do not provide any meaningful information about each other (possibly conditional on information about one or more other factors). This irrelevance assertion would then be recorded in the cross-impact matrix instead of the cross-impact estimate.

The updated cross-impact matrix for the Soisalo energy production example is shown in Table 5. The submatrix describing the dependencies between outcomes of nuclear power and energy

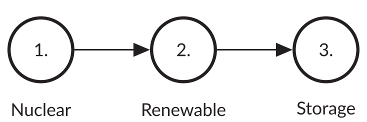


FIGURE 1 *Irrelevance* between uncertainty factors from Table 1 depicted as a graph.

TABLE 5 A cross-impact matrix containing the cross-impact multipliers describing the dependencies between the uncertainty factors from Table 1.

			Nuclear power			Renewables				Energy storage		
			< 1 TWh	1–5 TWh	> 5 TWh	< 1 TWh	1–4 TWh	4–8 TWh	> 8 TWh	< 10 GWh	10–400 GWh	> 400 GWh
			0.5	0.3	0.2	0.3	0.2	0.3	0.2	0.5	0.4	0.1
Nuclear power	< 1 TWh	0.5				$\frac{1}{4}$	1	$\frac{5}{4}$	$\frac{7}{4}$			
	1–5 TWh	0.3				$\frac{3}{2}$	$\frac{3}{2}$	$\frac{2}{3}$	$\frac{1}{4}$			
	> 5 TWh	0.2				$\frac{3}{2}$	$\frac{3}{2}$	1	$\frac{1}{10}$			
Renewables	< 1 TWh	0.3								$\frac{7}{4}$	$\frac{1}{4}$	$\frac{1}{10}$
	1–4 TWh	0.2								$\frac{5}{4}$	1	$\frac{1}{4}$
	4–8 TWh	0.3								$\frac{3}{4}$	$\frac{3}{2}$	$\frac{1}{2}$
	> 8 TWh	0.2								$\frac{1}{10}$	1	4
Energy storage	<10 GWh	0.5										
	10–400 GWh	0.4										
	> 400 GWh	0.1										

Note: Uncertainty factors representing nuclear power and energy storage have been deemed irrelevant to one another, so the associated cross-impact multipliers are not included.

storage is now left empty. (The cells below the diagonal are not considered, because cross-impacts are symmetric.) remaining non-empty submatrices now correspond to connections in the graph in Figure 1. Depending on the context, it is equally valid to either form the graph first and then ask the experts for the cross-impacts only in specific submatrices, or the experts can fill out the entire cross-impact matrix freely and leave those submatrices empty that they deem independent or conditionally independent.

Incorporating conditional independence into the computational model is straightforward because multiple *irrelevant* uncertainty factors are also *jointly irrelevant* (i.e., if the two or more uncertainty factors are *irrelevant* any partial scenario formed as a combination of their states is also irrelevant. See Appendix A for proof). Let us denote the set of relevant (not *irrelevant*) uncertainty factors for uncertainty factor i with D_i and the associated partial scenario set with $S_{D_i} = \mathsf{X}_{j \in D_i}\left\{s_1^j, \ldots, s_{n_j}^j\right\}$. The probability distribution over partial scenarios in S_{D_i} is calculated from the probability vector p of $S_{1:i-1}$ by taking a sum over all the partial scenarios in $S_{1:i-1}$ that can be obtained by extending the partial scenario $\mathbf{s}_{D_i}$

$$p(\mathbf{s}_{D_i}) = \sum_{\mathbf{s} \in E(\mathbf{s}_{D_i})} p(\mathbf{s}) = \sum_{\substack{\mathbf{s}_{1:i-1} \in S_{1:i-1} \\ E(\mathbf{s}_{1:i-1}) \subseteq E(\mathbf{s}_{D_i})}} p(\mathbf{s}_{1:i-1}). \tag{22}$$

Thus, the probability distribution p for partial scenario set S_{D_i} is the marginal distribution for uncertainty factors in D_i. The constraint $E(\mathbf{s}_{1:i-1}) \subseteq E(\mathbf{s}_{D_i})$ means that every scenario in the extension of $\mathbf{s}_{1:i-1}$ can also be found in the extension of $\mathbf{s}_{D_i}$, that is, partial scenarios $\mathbf{s}_{1:i-1}$ and $\mathbf{s}_{D_i}$ have the same outcomes for all uncertainty factors in D_i.

Now, because the uncertainty factors that are not included in D_i are *irrelevant* for estimating the conditional probabilities for the ith uncertainty factor based on the partial scenarios $\mathbf{s}_{1:i-1}$, we have

$$p\left(s_k^i \,\middle|\, \mathbf{s}_{1:i-1}\right) = p\left(s_k^i \,\middle|\, \mathbf{s}_{D_i}\right), \quad \text{if } E(\mathbf{s}_{1:i-1}) \subseteq E(s_{D_i}).\ \forall\ \mathbf{s}_{1:i-1} \in S_{1:i-1}, \mathbf{s}_{D_i} \in S_{D_i} \tag{23}$$

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

$$\min_{q(k|\mathbf{s}_{D_i})} \sum_{j \in D_i} \sum_{(k,l) \in R_{ij}} \left[\left(\sum_{\{\mathbf{s} \in S_{D_i} | s_j = l\}} q(k|\mathbf{s}) q(\mathbf{s})\right) - \hat{C}_{kl}^{ij} \hat{p}_k^i \hat{p}_l^j\right]^2 \tag{24}$$

$$\sum_{\mathbf{s} \in S_{D_i}} q(k|\mathbf{s}) q(\mathbf{s}) = \hat{p}_k^i,\ \forall\ k \in \{1, 2, \ldots, n_i\} \tag{25}$$

$$\sum_{k=1}^{n_i} q(k|\mathbf{s}_{D_i}) = 1,\ \forall\ \mathbf{s}_{D_i} \in S_{D_i} \tag{26}$$

$$q(k|\mathbf{s}_{D_i}) \geq 0,\ \forall\ k \in \{1, 2, \ldots, n_i\},\ \forall\ \mathbf{s}_{D_i} \in S_{D_i} \tag{27}$$

and then (23) can be used to get probabilities for all partial scenarios in $S_{1:i-1}$. As before, $\hat{C}$ and $\hat{p}$ represent expert judgments about cross-impacts and probabilities, respectively, and $q(\cdot)$ represents computed probabilities. Incorporating irrelevance between uncertainty factors like this makes the optimization problem more tractable because the number of scenarios at each iterative step will now depend only on the relevant uncertainty factors D_i instead of all factors $\{1, \ldots, i-1\}$. This also leads to tighter constraints and thus less uncertainty in scenario probabilities without increasing the size of the optimization problems.

Implementing conditional independence, however, adds some additional limitations on the order in which the uncertainty factors are included in the iterative process, because the uncertainty factors upon which the conditional independence relies are already contained in the partial scenario set $S_{1:i-1}$. Similarly, uncertainty factors that are marginally independent, such as A and B in Figure 2, but have other uncertainty factors dependent on them, should be included before the dependencies. This is because while introducing additional *irrelevant* uncertainty factors in expanding the set of partial scenarios cannot turn *irrelevant* uncertainty factors into relevant ones, the same does not hold for adding relevant uncertainty factors, which can under specific circumstances change other uncertainty factors from *irrelevant* into relevant.

3.4 | Bayesian networks

Conditional probability information and the conditional independence structure can also be combined to form a Bayesian network (Pearl & Paz, 2022). In the network, the conditional independence information is represented by a directed acyclic graph, like the one in Figure 3. The nodes (circles) represent uncertainty factors and the edges (arrows) between nodes represent conditional dependencies. Because the graph is directed and acyclic, there exists at least one total ordering of nodes such that node u precedes node v if there is an edge from node u to node v in the graph. This is called a *topological ordering* of the graph. In Figure 3 alphabetical order is one possible *topological ordering*. If two nodes do not share an edge they are conditionally independent given a subset of other nodes that precede either of the two. Conversely, if two nodes share an edge, they are not conditionally independent.

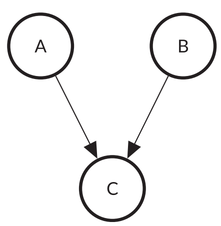


FIGURE 2 Two independent uncertainty factors A and B which share a dependent uncertainty factor C.

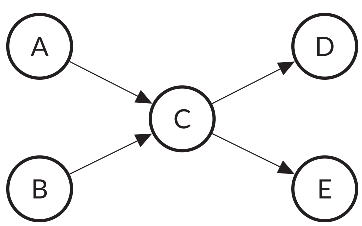


FIGURE 3 A simple conditional independence network.

Probability information is incorporated in the Bayesian network in the form of conditional probability tables like the one seen in Figure 4. Assuming the indexing of uncertainty factors follows a *topological ordering* of the Bayesian network, the probability table of uncertainty factor i contains the conditional probabilities $p\left(k|\mathbf{s}_{D_i}\right)\ \forall\ k \in \{1, \ldots, n_i\},\ \ \mathbf{s}_{D_i} \in S_{D_i}$. When combined, the conditional probability distributions $p(k|\mathbf{s}_{D_i})$ of all the uncertainty factors can be used to calculate the probability of any scenario $p(\mathbf{s})$ in S. This is because of the chain rule of probability (10) and conditional independence (23)

$$
\begin{aligned}
p(s) &= p(s_N|s_{1:N-1})p(\mathbf{s}_{1:N-1}) = p(s_N|s_{D_N})p(\mathbf{s}_{1:N-1}) \\
&= p(s_N|s_{D_N})p(s_{N-1}|s_{1:N-2})p(s_{1:N-2}) = p(s_N|s_{D_N})p(s_{N-1}|s_{D_{N-1}}) \\
&\qquad p(\mathbf{s}_{1:N-2})\ldots \\
&= p(s_N|s_{D_N})p(s_{N-1}|s_{D_{N-1}})\ldots p(s_2|s_{D_2})p(s_1),
\end{aligned}
\tag{28}
$$

which makes these conditional probability distributions a very memory-efficient way of storing the probability distribution.

Constructing the Bayesian network based on cross-impacts and conditional independence information is quite straightforward because the computational estimates for $p(k|\mathbf{s}_{D_i})$ can be obtained by solving the optimization problem (24)–(27). Indexing the uncertainty factors following a *topological ordering* of the conditional dependence graph is not a problem, because the cross-impact multipliers only measure probabilistic dependence and not causality, and thus they do not limit the order in any way.

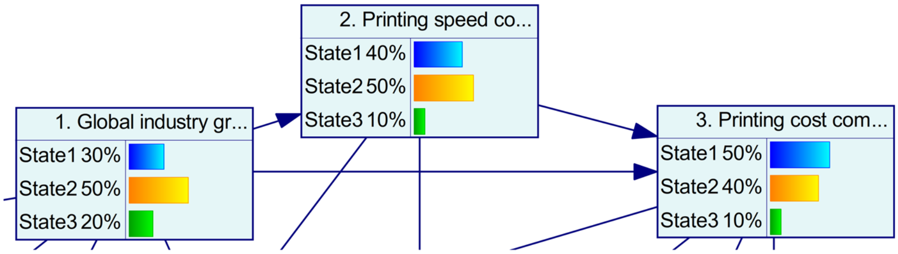
3.5 | Computational considerations

The results of the iterative method may differ slightly from those obtained by solving a single optimization problem to determine all scenario probabilities which would represent the best fit to all the elicited expert judgments about marginal probabilities and cross-impacts. Specifically, in the notation of Section 3.2, the problem of fitting scenario probabilities directly can be stated as the linear least squares problem

$$
\min_{q(\mathbf{s})} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{(k,l)\in R_{ij}} \left[\left(\sum_{\substack{\mathbf{s}\in S_{1:N} \\ s_i=k, s_j=l}} q(\mathbf{s}) \right) - \hat{C}_{kl}^{ij} \hat{p}_k^i \hat{p}_l^j \right]^2 \tag{29}
$$

$$
\sum_{\substack{\mathbf{s}\in S_{1:N} \\ s_i=k}} q(\mathbf{s}) = \hat{p}_k^i,\ \forall\ i \in \{1, \ldots, N\}, k \in \{1, \ldots, n_i\} \tag{30}
$$

$$
0 \le q(\mathbf{s}) \le 1,\ \forall\ \mathbf{s} \in S_{1:N}. \tag{31}
$$



		Global industry growth								
		s_1^1			s_2^1			s_3^1		
		Printing speed								
		s_1^2	s_2^2	s_3^2	s_1^2	s_2^2	s_3^2	s_1^2	s_2^2	s_3^2
	$\mathbf{s}_{D_3}$	(s_1^1, s_1^2)	(s_1^1, s_2^2)	(s_1^1, s_3^2)	(s_2^1, s_1^2)	(s_2^1, s_2^2)	(s_2^1, s_3^2)	(s_3^1, s_1^2)	(s_3^1, s_2^2)	(s_3^1, s_3^2)
Printing cost	$p(s_1^3 \mid \mathbf{s}_{D_3})$	0.88	0.81	0.65	0.55	0.31	0.33	0.31	0.09	0.40
	$p(s_2^3 \mid \mathbf{s}_{D_3})$	0.12	0.19	0.35	0.40	0.68	0.33	0.34	0.09	0.34
	$p(s_3^3 \mid \mathbf{s}_{D_3})$	0	0	0	0.05	0.01	0.33	0.35	0.82	0.26

FIGURE 4 One conditional probability table from the Bayesian network in the case study. The rows of the table show the outcome probabilities of the third uncertainty factor, printing cost, conditioned on the relevant preceding uncertainty factor's outcomes.

where the next to last constraint ensures that the scenario probabilities are matched to the marginal probabilities. The difference between the iterative process and this direct fitting stems from the fact that the iterative method does not weigh all cross-impact statements simultaneously. However, our computational experiments suggest that when the statements about marginal probabilities and cross-impacts are consistent (i.e., the sum (29) is zero for some scenario probabilities $q(s)$), both approaches produce similar distributions, which also fit all the marginals and cross-impacts perfectly by construction.

However, while the formulation (29) is conceptually simpler than the iterative approach in Section 3.2, it has a major caveat in that it leads to much bigger optimization problems. Specifically, the size of the optimization problem in (29) grows exponentially and in our computational tests, workstations with 16–32 GB RAM ran out of memory when the number of scenarios reached tens of thousands. In particular, the case study in Section 4 with its $3^{11} = 177,147$ scenarios proved too large for this approach. In contrast, the iterative approach which exploits judgments about conditional independence was able to construct the scenario probability distribution for the case study in 3.9 s using a laptop with 2.40 GHz I5 processor and 16 GB RAM.

In short, the iterative method performs better when analyzing large systems with more uncertainty factors. The reason for this is that the size of the optimization problems does not depend on the total number of scenarios but, rather, only on the size of the partial scenario set which contains the relevant uncertainty factors. To illustrate this point, if the new uncertainty factors can be added without increasing the average number of relevant factors per each new factor, the computational complexity of the iterative process increases linearly with the number of total uncertainty factors. This can be contrasted with the direct fitting approach in which the size of the optimization problems grows exponentially so that these problems become quickly unsolvable. In other words, while direct scenario probability fitting can still be used in smaller problems, the iterative method will prove indispensable in many problems of realistic size.

4 | CASE STUDY

We next present a case study on analyzing the developing 3D-printing technologies and their future impact on the Finnish Defence Forces (FDF). The aim of this case study was to (i) identify uncertainty factors that have a large impact on, how 3D printing would be applied in the Finnish military in view of possible developments over the next 15 years, (ii) specify ranges of possible realizations for these uncertainty factors, (iii) characterize dependencies between the uncertainty factors, and (iv) build structured scenario framework which would capture these multiple inputs, by doing so, and (v) serve as a tool for offering insights into questions which are of focal concern to the FDF in the context of 3D printing.

4.1 | Uncertainty factors

The identification of uncertainty factors was preceded by a systematic literature review and preliminary interviews with experts, resulting in an initial set of 10 key uncertainty factors with three outcomes for each. These uncertainty factors were discussed at length in a 4-h remote workshop which was organized by using video conferencing tools and attended by a panel consisting of half eight 3D-printing experts from the Finnish military and research community. The specific fields of expertise represented by the panelists covered military logistics, 3D-printing business, and 3D-printing technology.

In the workshop, the experts reached a consensus that the factor *Progress in 3D manufacturing* should be separated into two factors, representing *Printing speed* and *Printing costs*. This lead to the final list of uncertainty factors in Table 6.

In the workshop, the outcomes of every uncertainty factor were discussed together with the experts. A verbal description was developed for each, including numerical bounds where appropriate. For each outcome of every uncertainty factor, the corresponding marginal probability was assigned. This represented the baseline probability of this outcome in the absence of information about the outcomes of other uncertainty factors.

Next, the experts were asked to characterize cross-impacts using a seven-point scale from −3 to 3. This scale was employed to record statements about how the probability of a given outcome for an uncertainty factor would change from its baseline probability as a result of knowing that another uncertainty factor will have a specific outcome. For example, how much more likely it is that the global 3D industry maintains the growth speed of 2019–2020 if the costs associated with printing fall by 50%–90%? A small part of the cross-impact matrix is in Table 7.

The statements recorded on this ordinal scale were converted into estimates about cross-impact multipliers through the transformation

$$C^{ij}_{kl} = \sqrt{2}^{V^{ij}_{kl}}, \tag{32}$$

where C^{ij}_{kl} is the cross-impact multiplier derived from the statement V^{ij}_{kl}. Thus, responses from the range −3 to 3 were mapped to numerical values cross-impact multipliers $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, 1, $1\frac{1}{2}$, 2, and 3, and this information was available to the experts during the evaluations. The experts only estimated the cross-impacts of those uncertainty factor pairs they deemed to provide significant information about each other. The remaining pairs were deemed conditionally independent given the preceding uncertainty factors, as is shown in Table 8.

The usual scale from −3 to 3 was chosen (instead of asking about the cross-impact multipliers directly) to expedite the elicitation process while recognizing that the resulting estimates would not necessarily be consistent answers. Indeed, there were some inconsistencies in the resulting estimates. In Table 7, the highlighted cross-impact between stable industry growth and 10%–50% printing cost leads to a situation where the conditional probabilities associated with the row for global industry growth would not sum up to one. This inconsistency is easy to spot, because all terms on this row are nonnegative, meaning that stable industry growth would invariably preserve or increase the probabilities of all outcomes of printing costs; but this is impossible because these outcomes are (meant to be) mutually exhaustive (Salo et al., 2021). Similar, including less apparent inconsistencies appear all over the cross-impact matrix. Indeed, the fact that such inconsistencies in expert judgments are likely to surface in the cross-impact analysis is one of the reasons which motivated us to develop a method that could derive scenario probabilities even when the estimates are not perfectly consistent. This can make the elicitation process both faster and less arduous.

TABLE 6 Uncertainty factors, their outcomes, and corresponding marginal probabilities estimated for the year 2035.

Uncertainty factor	Outcome	Probability
1. Global industry growth	Decreases	0.3
	Remains same	0.5
	Increases	0.2
2. Printing speed compared to present	Up to 2 times faster	0.4
	2–10 times faster	0.5
	Over 10 times faster	0.1
3. Printing cost compared to present	Up to 50% cheaper	0.5
	50%–90% cheaper	0.4
	Over 90% cheaper	0.1
4. Finnish industry growth	Decreases	0.3
	Remains same	0.5
	Increases	0.2
5. Graduates with 3D-printing expertise	Up to 100 (current)	0.2
	100–300	0.6
	Over 300	0.2
6. Legal regulation of 3D-printing in Finland	Limits strongly	0.05
	Similar to other manufacturing	0.9
	No regulation	0.05
7. Standardization of processes and models	No standardization	0.35
	Includes technical requirements	0.45
	Full automation possible	0.2
8. Use of 3D-printed objects in FDF	Just individual items	0.1
	Common and has purchase procedures	0.5
	Access to 3D-printing capacity on demand	0.4
9. FDF access to 3D-printing model files	Just individual items	0.2
	Relevant models included in system purchases	0.7
	Models available for most new and old systems	0.1
10. FDF 3D-printing spare parts in peacetime	Low importance	0.7
	Significant and well planned	0.29
	Crucial and strictly controlled	0.01
11. FDF 3D-printing spare parts in crisis times	Low importance	0.45
	Significant and well planned	0.45
	Crucial and strictly controlled	0.1

4.2 | Results

We used the presented iterative method to compute the scenario probability distribution for all the scenarios that can be formed from the uncertainty factors in Table 6. The entire distribution could not be included here, because it has $3^{11} = 177,147$ probabilities. Thus, we are offering some (hopefully) interesting observations instead.

TABLE 7 Part of the cross-impact estimate matrix on probabilistic dependencies between pairs of outcomes for uncertainty factors.

			Printing cost		
			50-100%	10-50%	Max10%
			0.5	0.4	0.1
Global industry growth	Slow	0.3	2	0	-2
	Stable	0.5	0	2	0
	Fast	0.2	-2	0	3
Printing speed	100-200%	0.4	1	-1	-1
	200-1000%	0.5	-1	1	0
	Over 1000%	0.1	-1	0	1

The cross-impact judgments provided by the experts indicated that the role of 3D-printing in the future of spare parts logistics is quite uncertain. There is a 41% probability that it will not have a great role in either the peace or crisis time logistics. From the scenario probability distribution, we calculated that the probability of any scenario where spare parts production in either peace or crisis time is significant or crucial is practically zero, if the use of 3D printed objects in FDF is limited to just individual items or access to 3D-printing models is extremely limited. Collecting a library of 3D-printing model files and building processes to order and use 3D-printed items takes a significant amount of time and effort, so it would be advisable to start as soon as possible if the 3D printing of spare parts is seen as worth pursuing.

Using the conditional probability distributions and conditional independence information (Table 8), we also constructed a Bayesian network using the GeNIe Modeler software (BayesFusion, LLC, 2021), seen in Figure 5. The uncertainty factors were introduced starting with exogenous factors that would not be affected by choices the FDF makes, followed by exogenous factors that could be affected in limited ways in cooperation with Finnish government entities and industry, and the last factors included were endogenous to the Finnish military. Thus, factors 1–3 describe the state of the 3D-printing industry globally, 4–7 describe the situation in Finland, and 8–11 describe the situation inside the FDF. The constructed network can be used to illuminate various what-if (partial) scenarios (Fenton & Neil, 2001).

To give a concrete example, looking at the partial scenarios consisting of uncertainty factors 1–7, that is, the exogenous factors, the most probable is the one in which every factor gets the second outcome. Its probability is 8.62%, which is quite high considering these uncertainty factors can produce $3^7 = 2187$ different partial scenarios. Figure 6 shows how the probabilities of the FDF endogenous uncertainty factors change when the outcomes of other

TABLE 8 The X:s denote the uncertainty factors whose cross-impacts were evaluated.

	1	2	3	4	5	6	7	8	9	10	11
1. Global industry growth		X	X	X	X	X	X				
2. Printing speed compared to present			X	X							X
3. Printing cost compared to present				X				X	X	X	
4. Finnish industry growth					X	X	X	X	X	X	X
5. Number of graduates with 3D-printing expertise									X		
6. Legal regulation of 3D-printing in Finland								X	X	X	
7. Standardization of processes and models									X		X
8. Use of 3D-printed objects in FDF									X	X	X
9. FDF access to 3D-printing model files										X	X
10. FDF 3D-printing spare parts in peacetime											X
11. FDF 3D-printing spare parts in crisis times											

Note: The empty white cells are conditionally independent uncertainty factor pairs.

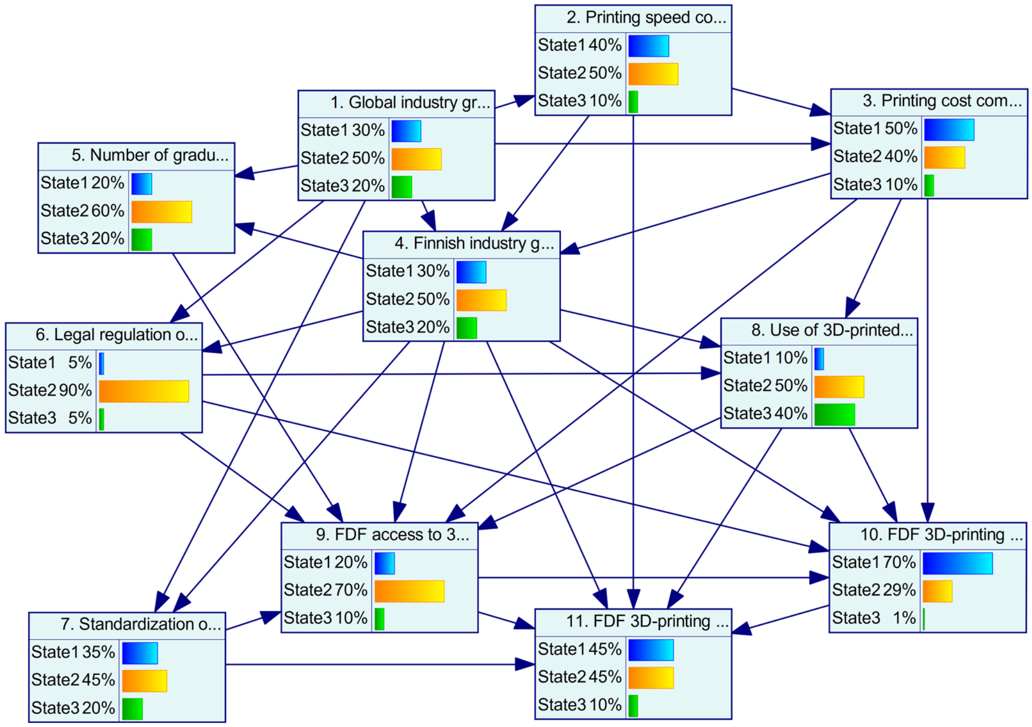


FIGURE 5 The constructed Bayesian network in GeNIe Modeler software (BayesFusion, LLC, 2021).

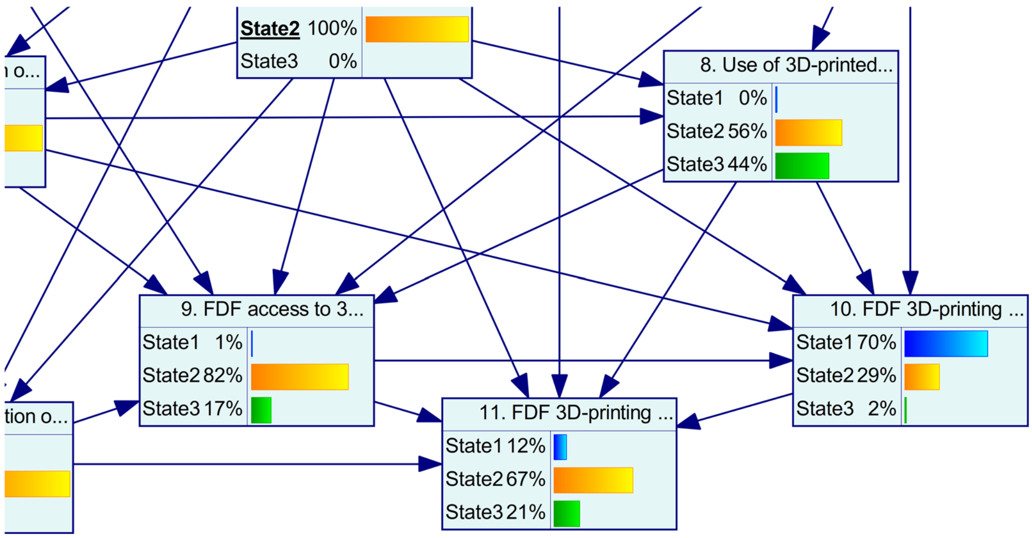


FIGURE 6 The probability distributions of uncertainty factors describing 3D-printing in Finnish Defence Forces in the most likely exogenous partial scenario.

uncertainty factors are locked in place. 3D-printed parts are very likely to have at least significant importance in crisis time operations. At the same time, they are quite unlikely to be that important during peacetime. Because crisis capabilities are developed during peacetime, this means that special attention should be focused on both training and developing processes to support this 3D-printing spare parts in a crisis, because it seems unlikely to develop on its own.

Figure 7 shows how the endogenous probabilities change yet again when the exogenous uncertainty factors are locked into another relatively high probability (2.28%) partial scenario in which uncertainty factors 1–5 and 7 all obtain the first outcome while uncertainty factor 6 obtains its second outcome. This is a more pessimistic partial scenario for the 3D-printing industry as a whole and represents growth and technological development slowing down

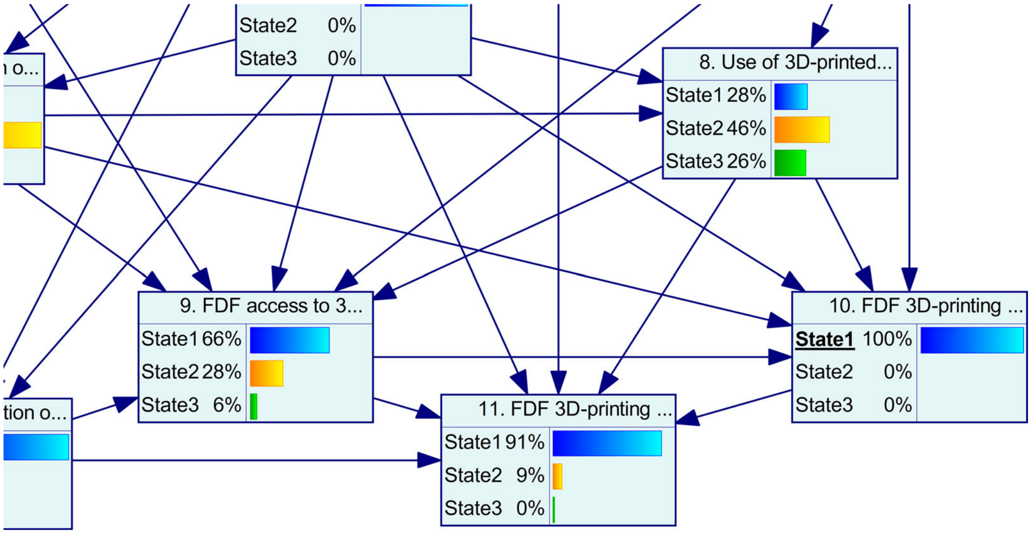


FIGURE 7 The probability distributions of uncertainty factors describing 3D-printing in Finnish Defence Forces in a pessimistic high probability exogenous partial scenario.

significantly. Here 3D-printing is unlikely to play any role at all in the spare parts logistics, and this helps showcase why there is such a high probability of them remaining unremarkable despite their seeming importance in the most probable scenario in Figure 6. The developments in the industry as a whole are going to have an impact on the usefulness of the technology for the FDF and should be monitored carefully.

5 | CONCLUSIONS AND DISCUSSION

Scenario analysis provides a structured framework for identifying and exploring how the future is shaped by key uncertainties. Because the range of qualitative and quantitative scenario methods is so wide, there are different perspectives on the rationales and suitable ways of carrying out scenario analysis (Millett, 2009), but qualitative and quantitative scenario methods are not fundamentally at odds. Rather, they are complementary: for instance, evocative narratives can be made even more compelling by accompanying them with numerical data while detailed quantitative analyses can be enriched with storylines to communicate the implications behind the numbers. In short, the choice of methodologies should be guided by how the scenarios are going to be used.

Against this backdrop, we have formulated a method to elicit expert judgments about cross-impact terms which are processed to infer the accompanying joint probability distribution over all possible scenarios. A notable benefit of this approach is that it facilitates the integration with other well-founded quantitative approaches—including expected utility theory, probabilistic risk assessment, and statistical inference—and thus expands the extant range of available techniques for foresight and strategic planning.

To our knowledge, our approach is the first to include conditional independence in cross-impact analysis. This increases the number of uncertainty factors that can be included in the analysis without overwhelming the experts from whom cross-impact estimates are elicited. Adding this new element to the elicitation process may necessitate some training when working with experts who are accustomed to earlier cross-impact methods. However, even if some conditional independence relationships are overlooked, as long as reasonable cross-impact estimates were provided instead, this should have a minimal impact on the calculated results.

Scenario probabilities and Bayesian networks depicting dependencies between the uncertainty factors have many practical uses in the military context. Scenario probability distributions facilitate a number of different analyses to assess the impacts of new technologies. Numerous simulation (Lappi, 2008; Rao et al., 1993), game-theoretic (Poropudas & Virtanen, 2010; Roponen et al., 2020) and dynamic (Gue, 2003) tools can be used to analyze the scenario-specific system performance, but their ability to support strategic analysis is limited without the underlying scenario probabilities.

The same also applies to technology forecasting beyond the military context to some extent. For example, climate models help generate well-founded scientific predictions concerning the rate of change in the global temperature and sea level, rise, but they are less apt at predicting what kinds of mitigation actions governments will take or how people will respond to changing environmental conditions; yet such behavioral would also need to be accounted for to address risks comprehensively. As a result, there is a need for scenario models such as ours which harnesses cross-impact statements to link technological changes to the key behavioral responses that are pivotal in shaping the future.

Among cross-impact methods, ours is purposely grounded on the estimation of all possible scenario probabilities. Most probabilistic cross-impact methods tend to rely on Monte Carlo simulation, which, however, may require an impractically large number of iterations to reach good accuracy when the number of scenarios is large. Computationally, our method scales well into problems with even dozens of uncertainty factors, especially if the number of probabilistic dependencies between the uncertainty factors is not too large (see Appendix B). The problems caused by large dependency sets can be mitigated to some extent by choosing the included uncertainty factors and iteration order in the right way, but eventually, a limit is reached on how much can be expressed with just pairwise dependency statements. This is a limitation shared by all cross-impact techniques because the number of possible scenarios grows faster than the number of cross-impacts.

We have employed unconditional cross-impact multipliers (4) in which the relative change in the probability of a given outcome level does not explicate assumptions about the realizations of uncertainty factors beyond the two that are considered in the comparison. Mathematically, however, one could elicit conditional cross-impact multipliers which would explicate such assumptions with no reason why the cross-impact multipliers could not be even extended into triplets or a larger number of uncertainty factors. We have chosen not to explore it beyond conditional independence in this paper, because the number of triplets and beyond grows so much faster than the number of pairs, that collecting such information for all uncertainty factors would be practically infeasible in most cases. However, introducing individual optimization constraints based on higher-level dependencies would be straightforward if desired.

Although our case study has focused on 3D-printing, the proposed method is generic and can be readily applied across numerous contexts in which it is of interest to build a comprehensive model that retains all possible scenarios. Thus, its advantages lie in countering the risk that the focus is, perhaps prematurely, placed on a small subset of scenarios, as opposed to capturing the full breadth of possible scenarios that can be built as combinations of outcomes of several uncertainty factors. In such contexts, cross-impact analysis offers a pragmatic, relatively straightforward, and cognitively manageable approach to assessing dependencies between uncertainty factors. Furthermore, the models proposed in this paper are computationally efficient and make it possible to provide informative insights based on the interactive exploration of the implications of all model inputs, including judgments about the marginal outcome probabilities and cross-impact statements.

ACKNOWLEDGMENTS

The experts who participated in our workshop and provided their valuable insights for the case study are gratefully acknowledged. This research has been financially supported by MATINE—The Scientific Advisory Board for Defence of the Finnish Ministry of Defence.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data for the case study can be found in the Supporting Information excel-file.

ORCID

Juho Roponen http://orcid.org/0000-0002-7921-0072
Ahti Salo http://orcid.org/0000-0003-1526-7923

ENDNOTE

[1] An uncertainty factor refers here to a salient aspect of the future which is uncertain and which can be addressed by dividing it into two or more mutually exclusive and jointly exhaustive outcomes. In the literature, there are several terms for analogous concepts, for example, random variable (Kallenberg, 1997), random event (Harsanyi, 1967), key factor (Bunn & Salo, 1993), lottery (Myerson, 1997; Raiffa, 1968), distinction (Howard & Abbas, 2016), and uncertainty factor (Seeve & Vilkkumaa, 2022). Here, the term uncertainty factor emphasizes that the perceived randomness need not arise from inherent stochasticity but, rather, the subjective lack of information. The interpretation of this term is consistent with (Salo et al., 2021).

REFERENCES

BayesFusion, LLC. (2021). *GeNIe modeler*. Retrieved May 31, 2022, from https://www.bayesfusion.com/

Ben-Tal, A., ElGhaoui, L., & Nemirovski, A. (2009). *Robust optimization* (Vol. 28). Princeton University Press.

Bolger, F., & Rowe, G. (2015). The aggregation of expert judgment: Do good things come to those who weight? *Risk Analysis*, *35*, 5–11.

Bolger, F., & Wright, G. (2011). Improving the Delphi process: Lessons from social psychological research. *Technological Forecasting and Social Change*, *78*, 1500–1513.

Booth, J., Edwards, E., Whitley, M., Kranz, M., Seif, M., & Ruffin, P. (2018). Military comparison of 3D printed vs commercial components. In V. K. Varadan (Ed.), *Nano-, bio-, info-tech sensors, and 3D systems II* (Vol. 10597, pp. 52–66). SPIE.

Bowman, G. (2016). The practice of scenario planning: An analysis of inter-and intra-organizational strategizing. *British Journal of Management*, *27*, 77–96.

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

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

Carlsen, H., Eriksson, E. A., Dreborg, K. H., Johansson, B., & Bodin, Ö. (2016). Systematic exploration of scenario spaces. *Foresight*, *18*, 59–75.

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

Dalkey, N. C. (1971). An elementary cross-impact model. *Technological Forecasting and Social Change*, *3*, 341–351.

Derbyshire, J. (2017). The siren call of probability: Dangers associated with using probability for consideration of the future. *Futures*, *88*, 43–54.

Edwards, W., Miles, R. F., & Von Winterfeldt, D. (2007). *Advances in decision analysis*. Citeseer.

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

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.

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

Godet, M. (1986). Introduction to la prospective: Seven key ideas and one scenario method. *Futures*, *18*, 134–157.

Godet, M. (1994). *From anticipation to action: A handbook of strategic prospective*. UNESCO Publishing.

Gordon, T. J. (1994). *Cross-impact method* (Vol. 4). American Council for the United Nations University.

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

Gue, K. R. (2003). A dynamic distribution model for combat logistics. *Computers & Operations Research*, *30*, 367–381.

Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players, I-III part I. The basic model. *Management Science*, *14*, 159–182.

Heinen, J. J., & Hoberg, K. (2019). Assessing the potential of additive manufacturing for the provision of spare parts. *Journal of Operations Management*, *65*, 810–826.

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

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

Howard, R. A., & Abbas, A. E. (2016). *Foundations of decision analysis*. Pearson Education Limited.

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

Jiménez, M., Romero, L., Domínguez, I. A., Espinosa, M. D. M., & Domínguez, M. (2019). Additive manufacturing technologies: An overview about 3d printing methods and future prospects. *Complexity*, *2019*, 30.

Kallenberg, O. (1997). *Foundations of modern probability*. Springer.

Kemp-Benedict, E. (2004). From narrative to number: A role for quantitative models in scenario analysis. In C. Pahl-Wostl, S. Schmidt, A. E. Rizzoli, & A. J. Jakeman (Eds.), *International Congress on Environmental Modelling and Software*, 22.

Kietzmann, J., Pitt, L., & Berthon, P. (2015). Disruptions, decisions, and destinations: Enter the age of 3-D printing and additive manufacturing. *Business Horizons*, *58*, 209–215.

Kosow, H., & Gaßner, R. (2008). *Methods of future and scenario analysis: Overview, assessment, and selection criteria* (Vol. 39). DEU.

Lappi, E. (2008). Sandis military operation analysis tool. In *2nd Nordic Military Analysis Symposium* (pp. 17–18).

Lehr, T., Lorenz, U., Willert, M., & Rohrbeck, R. (2017). Scenario-based strategizing: Advancing the applicability in strategists' teams. *Technological Forecasting and Social Change*, *124*, 214–224.

Meyer, M. A., & Booker, J. M. (2001). *Eliciting and analyzing expert judgment: A practical guide*. SIAM.

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

Myerson, R. B. (1997). *Game theory: Analysis of conflict*. Harvard University Press.

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

Pearl, J., & Paz, A. (2022). Reasoning about relevance relations. *Probabilistic and Causal Inference: The Works of Judea Pearl*, (pp. 189).

Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P., Fernandez-Manjarrés, J. F., Araújo, M. B., Balvanera, P., Biggs, R., Cheung, W. W., Chini, L., Cooper, H. D., Gilman, E. L., Guénette, S., Hurtt, G. C., Huntington, H. P., Mace, G. M., Oberdorff, T., Revenga, C., … Walpole, M. (2010). Scenarios for global biodiversity in the 21st century. *Science*, *330*, 1496–1501.

Phadnis, S., Caplice, C., Singh, M., & Sheffi, Y. (2014). Axiomatic foundation and a structured process for developing firm-specific intuitive logics scenarios. *Technological Forecasting and Social Change*, *88*, 122–139.

Poropudas, J., & Virtanen, K. (2010). Game-theoretic validation and analysis of air combat simulation models. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *40*, 1057–1070.

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty*. Addison-Wesley.

Rao, A. S., Lucas, A., Morley, D., Selvestrel, M., & Murray, G. (1993). *Agent-oriented architecture for air-combat simulation*. Technical report, Technical Note 42, The Australian Artificial Intelligence Institute.

Roponen, J., Insua, D. R., & Salo, A. (2020). Adversarial risk analysis under partial information. *European Journal of Operational Research*, *287*, 306–316.

Salo, A., Tosoni, E., Roponen, J., & Bunn, D. W. (2021). Using cross-impact analysis for probabilistic risk assessment. *Futures & Foresight Science*, *4*, e2103.

Schneider, J., & Kirkpatrick, S. (2007). *Stochastic optimization*. Springer Science & Business Media.

Scholz, R. W., & Tietje, O. (2002). *Embedded case study methods: Integrating quantitative and qualitative knowledge*. Sage.

Schwartz, P. (2012). *The art of the long view: Planning for the future in an uncertain world*. Currency.

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

Siljander, R., & Ekholm, T. (2018). Integrated scenario modelling of energy, greenhouse gas emissions and forestry. *Mitigation and Adaptation Strategies for Global Change*, *23*, 783–802.

Tosoni, E., Salo, A., & Zio, E. (2018). Scenario analysis for the safety assessment of nuclear waste repositories: A critical review. *Risk Analysis*, *38*, 755–776.

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

Wright, G., & Cairns, G. (2011). *Scenario thinking: Practical approaches to the future*. Palgrave MacMillan.

How to cite this article: Roponen, J., & Salo, A. (2023). A probabilistic cross-impact methodology for explorative scenario analysis. *Futures & Foresight Science*, e165. https://doi.org/10.1002/ffo2.165

APPENDIX A: PROOFS

Definition Consider the set of partial scenarios $S_{1:i}$ such that the probabilities $p(\mathbf{s}_{1:i-1})$ are strictly positive. Then the uncertainty factor $a \in \{1, \ldots, i\}$ is *irrelevant* for uncertainty factor *i* if and only if $p(s_k^i|\mathbf{s}_{1:i-1}) = p\left(s_k^i|\mathbf{s}_{1:i-1\setminus a}\right)$, $\forall\ k = 1, \ldots, n_i$, $\forall\ \mathbf{s}_{1:i-1} \in S_{1:i-1}$, where $\mathbf{s}_{1:i-1\setminus a}$ denotes the partial scenario that is contained in $\mathbf{s}_{1:i-1}$ but does not include the uncertainty factor *a*.

We note that the requirement of positive probabilities for partial scenarios $\mathbf{s}_{1:i-1}$ is needed because otherwise the conditional probabilities as the conditional would not be well-defined.

Theorem 1 (Intersection). *Let a, b, and i be distinct uncertainty factors in* {1, ..., i} *such that* $a \neq b$, $b \neq i$, $i \neq a$. *If uncertainty factors a and b are* irrelevant *for uncertainty factor i in the set of partial scenarios* $S_{1:i}$, *then* $p(s_k^i|\mathbf{s}_{1:i-1}) = p\left(s_k^i|\mathbf{s}_{1:i-1\backslash a,b}\right)$, $\forall\ k = 1, ..., n_i$, $\forall\ \mathbf{s}_{1:i-1} \in S_{1:i-1}$.

Proof. By assumption,

$$p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1}\right) = p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash a}\right), \forall\ k = 1, ..., n_i, \forall\ \mathbf{s}_{1:i-1} \in S_{1:i-1} \tag{A1}$$

$$p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1}\right) = p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash b}\right), \forall\ k = 1, ..., n_i, \forall\ \mathbf{s}_{1:i-1} \in S_{1:i-1} \tag{A2}$$

and thus

$$p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash a}\right) = p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash b}\right), \forall\ k = 1, ..., n_i, \forall\ \mathbf{s}_{1:i-1} \in S_{1:i-1}. \tag{A3}$$

Now, let $k \in \{1, ..., n_i\}$ and $\mathbf{s}_{1:i-1} \in S_{1:i-1}$. Using the law of total probability, we get

$$p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash a,b}\right) = \sum_{l=1}^{n_a} p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash a,b}, s_l^a\right)p\left(s_l^a\middle|\ \mathbf{s}_{1:i-1\backslash a,b}\right) \tag{A4}$$

Because $\left(\mathbf{s}_{1:i-1\backslash a,b} \wedge s_l^a\right)$ is a partial scenario in $S_{1:i-1\backslash b}$, the equality (A3) gives

$$p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash a,b}\right) = \sum_{l=1}^{n_a} p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash a}\right)p\left(s_l^a\middle|\ \mathbf{s}_{1:i-1\backslash a,b}\right) \tag{A5}$$

$$= p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash a}\right)\sum_{l=1}^{n_a} p\left(s_l^a\middle|\ \mathbf{s}_{1:i-1\backslash a,b}\right) \tag{A6}$$

$$= p\left(s_k^i\middle|\ \mathbf{s}_{1:i-1\backslash a}\right). \tag{A7}$$

Thus, combining this equality $p\left(s_k^i\middle|\mathbf{s}_{1:i-1\backslash a,b}\right) = p\left(s_k^i\middle|\mathbf{s}_{1:i-1\backslash a}\right)$ with the assumption $p\left(s_k^i\middle|\mathbf{s}_{1:i-1}\right) = p\left(s_k^i\middle|\mathbf{s}_{1:i-1\backslash a}\right)$ leads to the stated conclusion $p\left(s_k^i\middle|\mathbf{s}_{1:i-1}\right) = p\left(s_k^i\middle|\mathbf{s}_{1:i-1\backslash a,b}\right)$, $\forall\ k = 1, ..., n_i$, $\forall\ \mathbf{s}_{1:i-1} \in S_{1:i-1}$. □

It would also be straightforward to extend this proof to any number of *irrelevant* uncertainty factors.

APPENDIX B: COMPUTATIONAL EXPERIMENTS

To test how much the order of the uncertainty factors affected the results iterative approach from Section 3.2 and how much they differed from the direct fitting approach found in Section 3.5, we compared the results with several randomly generated distributions with exponentially distributed probability density using 8 uncertainty factors with three possible outcomes each. We used $\lambda = 1$ for generating the randomized probabilities, but because the probabilities need to be normalized to sum up to 1, the λ does not affect the end result. The setup was chosen to represent a worst-case where the probabilities cannot be represented well with cross-impacts. The exponential distribution was used because it produces probabilities that are similarly distributed in magnitude as, for example, our 3D-printing case study.

We tested the method with 10 different randomly generated distributions that had eight uncertainty factors with three possible outcomes each. The test was limited to eight uncertainty factors because we also wanted to test against the direct fitting approach from Section 3.5, and including more than eight uncertainty factors would have required more memory than was available on our test laptop. The cross-impact multipliers and marginal probabilities were calculated from the randomly generated probability distributions. We also used five different randomly chosen calculation orders for the iterative method for each distribution, to test whether the order of uncertainty factors made a difference.

The total absolute differences of the calculated probability distributions, that is, the sum of the absolute values of the differences between the scenario probabilities ranged from 0.0130 to 0.0273 when comparing the different computation orders for the iterative method, and from 0.0211 to 0.0304 when comparing the iterative method to the direct fitting approach. Considering that there were $3^8 = 6561$ scenarios in total, the average differences in individual scenario probabilities were less than 5×10^{-6}. Both methods also produced probability distributions that matched the given cross-impacts and marginal probability distributions exactly, as designed. The calculated distributions are so similar that any inaccuracies arising from computation orders and approaches are dwarfed by the inaccuracies in the expert judgments intended to be used for the calculations.

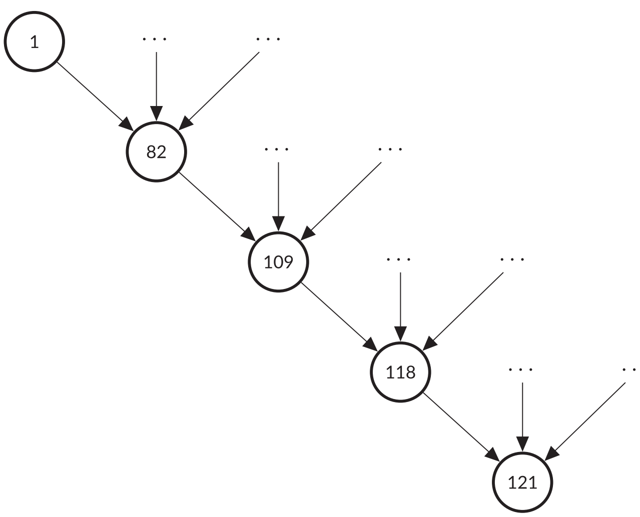


FIGURE B1 A 3-regular directed tree with a depth of 4. The number of nodes multiplies by three on each row when moving up. The first row has nodes 1–81, the second row has 27 nodes numbered 82–108, the third row nine nodes numbered 109–117, the fourth row three nodes numbered 118–120, and the fifth row only has one node numbered 121.

The calculated distributions, however, differed significantly from the original ones used to generate the marginals and cross-impacts, because all higher-degree dependence information was lost. The only similarity the calculated distributions had to the original was limited to the cross-impacts and marginals. This was expected because the original distributions were just a collection of random numbers that added up to 1.

To test the scalability of the iterative method, we created a conditional independence network of 121 uncertainty factors, whose underlying undirected graph was a 3-regular tree of depth 4 (Figure B1). This means that the last uncertainty factor, that is, number 121, had three incoming connections, and each of those three connected nodes also had three incoming connections and so on, repeated a total of four times. Each of the uncertainty factors had three possible outcomes.

The total number of scenarios that could be formed as combinations of 121 uncertainty factors with three outcomes each is $3^{121} \approx 5.4 \times 10^{57}$. This means that the entire scenario probability distribution is impossible to store on any device as a list of probabilities, so instead, we only computed the conditional probability distributions for all uncertainty factors, which could be used to easily calculate the probability of any scenario. Because we were testing the scalability of the method presented in this paper, both the marginal probabilities and cross-impact estimates were randomly generated. Thus, both the parameters and the output of the computation are nonsense.

A laptop with a 2.40 GHz I5 processor and 16 GB RAM calculated the conditional probabilities in 0.36 s. Considering that the 3D-printing case example takes 3.9 s to calculate on the same computer, this confirmed that the total number of uncertainty factors can be very large as long as the number of cross-impacts affecting individual uncertainty factors remains low. In the 3D-printing case example, most of the calculation time is spent on uncertainty factors 9 and 11, both of which are dependent on six preceding uncertainty factors and take 1.7–1.8 s to calculate each. Solving the least squares minimization problems takes up 90% of the total computation time in the 3D-printing case. With the large tree, the least squares problem solver only took around 60% of the total computation time, but this is mostly because of other inefficiencies in the code taking up relatively more time with the shorter total run time.