Aalto University
School of Science
Master's Programme in Mathematics and Operations Research

Teemu Seeve

A Structured Method for Identifying and Visualizing Scenarios

Master's Thesis
Espoo, May 22, 2018

Supervisor: Professor Ahti Salo
Advisor: Professor Eeva Vilkkumaa

Aalto University
School of Science
Master's Programme in Mathematics and Operations Research

ABSTRACT OF MASTER'S THESIS

Author: Teemu Seeve

Title:
A Structured Method for Identifying and Visualizing Scenarios

Date: May 22, 2018 **Pages:** vii + 80

Major: Systems and Operations Research **Code:** SCI3055

Supervisor: Professor Ahti Salo

Advisor: Professor Eeva Vilkkumaa

To retain their competitive edge, companies and other organizations must develop methods to increase their understanding of changes in the future operational environment. These changes can be characterized with a set of scenarios, which provide plausible depictions of the future. Scenarios can be modelled as combinations of levels of uncertainty factors describing, e.g., alternative political or technological developments. However, the number of possible scenarios grows exponentially in the number of these factors. To date, there are few systematic methods to support the selection of a small set of plausible but mutually dissimilar scenarios from an exponentially large set of candidate scenarios.

In this thesis, we develop a method for identifying a set of internally consistent and sufficiently dissimilar scenarios. The method filters an exponentially large set of candidate scenarios to a smaller set of most plausible scenarios, as assessed by the consistencies of the pairs of uncertainty factor levels in the candidate scenarios. By applying Multiple Correspondence Analysis to this set, the most consistent scenarios are visualized by a Scenario Map, from which a set mutually dissimilar consistent scenarios can be selected. As a part of this thesis, an interactive software tool was developed, which implements the scenario identification and visualization method. This thesis also presents a case study in which this tool was used to identify a set of plausible futures for a Finnish National Emergency Supply Organization to support their strategic decision making.

Our method provides mathematically sound means for building scenarios efficiently based on numerous uncertainty factors. By guiding the building process through effective visualizations, the method leaves room for explorative thinking. Moreover, advancing in transparent and accessible steps, the method fosters trust in the developed scenarios. Based on the positive feedback from the case study, the method can provide valuable support in other scenario exercises as well.

Keywords: scenario planning, General Morphological Analysis, dimensionality reduction, Multiple Correspondence Analysis

Language: English

Aalto-yliopisto
Perustieteiden korkeakoulu
Matematiikan ja operaatiotutkimuksen maisteriohjelma

DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Teemu Seeve

Työn nimi:
Strukturoitu menetelmä skenaarioiden identifioimiseksi ja visualisoimiseksi

Päiväys:	22. toukokuuta 2018	**Sivumäärä:**	vii + 80
Pääaine:	Systeemi- ja operaatiotutkimus	**Koodi:**	SCI3055
Valvoja:	Professori Ahti Salo		
Ohjaaja:	Professori Eeva Vilkkumaa		

Kilpailukykynsä säilyttämiseksi yritysten ja muiden organisaatioiden on kehitettävä menetelmiä tulevaisuuden toimintaympäristön muutosten ymmärtämiseksi. Näitä muutoksia voi luonnehtia skenaarioilla, jotka kuvaavat uskottavia näkemyksiä tulevaisuudesta. Skenaarioita voidaan mallintaa kombinaatioina epävarmuustekijöistä, jotka kuvaavat esimerkiksi vaihtoehtoisia poliittisia tai teknologisia kehityskulkuja. Mahdollisten skenaarioiden määrä kuitenkin kasvaa eksponentiaalisesti tekijöiden määrän suhteen. Tähän mennessä on kehitetty vain vähän menetelmiä muutaman, mutta keskenään erilaisen skenaarion valinnan tueksi eksponentiaalisen monesta skenaariokandidaatista.

Tässä diplomityössä kehitetään menetelmä sisäisesti konsistenttien ja keskenään erilaisten skenaarioiden tunnistamiseksi. Menetelmä suodattaa eksponentiaalisen suuresta määrästä skenaariokandidaatteja pienemmän joukon uskottavia skenaarioita, kun uskottavuutta arvioidaan skenaarioiden sisältämien epävarmuustekijöiden pareittaisten konsistenssien perusteella. Moniulotteista korrespondenssianalyysiä soveltamalla konsistenteimmat skenaariot voidaan kuvata Skenaariokartalla, josta voidaan valita joukko keskenään erilaisia skenaarioita. Osana diplomityötä kehitettiin menetelmän pohjalta ohjelmistotyökalu, joka tukee skenaarioiden identifiointia ja visualisointia. Tässä työssä esitetään myös tapaustutkimus, jossa kehitetyllä työkalulla tunnistettiin onnistuneesti vaihtoehtoisia tulevaisuuksia Suomen Huoltovarmuusorganisaation strategisessa päätöksenteossa.

Menetelmä tarjoaa matemaattisesti perustellun ja tehokkaan keinon useista tekijöistä muodostuvien skenaarioiden rakentamiseksi. Opastamalla rakentamisprosessia vaikuttavilla visualisoinneilla menetelmä edistää luovaa ajattelua. Lisäksi menetelmän läpinäkyvyys ja selkeys edistävät skenaarioihin kohdistuvaa luottamusta. Tapaustutkimuksesta saadun positiivisen palautteen perusteella menetelmä voi tarjota arvokasta tukea muissakin skenaariohankkeissa.

Asiasanat:	skenaariosuunnittelu, Yleinen Morfologinen Analyysi, dimensionredusointi, Moniulotteinen korrespondenssianalyysi
Kieli:	Englanti

Acknowledgements

First and foremost, I thank my advisor Professor Eeva Vilkkumaa for her devotion and guidance in my research. Her integrity and diligence have been the most impactful driving forces for the quality of this thesis. I would also like thank my supervisor Professor Ahti Salo for the valuable advice he provided at the end of this thesis.

I am grateful to the personnel at Capful for the opportunity to work on this intriguing thesis topic. Jari, Arto, Nette, Kimmo, Katja: the software development project part of this thesis was a most exciting journey to share with you, and the real case with which your work has provided this thesis is of utmost value.

Friends and staff at the Systems Analysis Laboratory and the Management Science research group: thank you for the encouraging and professional environment that you bring to my workplace.

My special thanks go to my friends and family for their support during the many years of my studies. Finally, I thank my dear Katja for her constant support, as my partner and best friend.

Espoo, May 22, 2018

Teemu Seeve

Notation

N	Total number of possible scenarios
I	Number of consistent scenarios
M	Number of mutually dissimilar consistent scenarios
j	Index of an uncertainty factor
J	Number of uncertainty factors
K_j	Number of levels of uncertainty factor j
K	Total number of levels in all uncertainty factors
$\mathbf{K}$	Vector of numbers K_j of uncertainty factor levels
k_j	A level of uncertainty factor j
$\mathcal{K}_j$	Set of possible levels of uncertainty factor j
$\mathbf{s}$	Vector of levels k_j of factors j, i.e., a scenario
$\mathcal{S}$	Set of all possible scenarios
(k_{j_1}, k_{j_2})	A pair of uncertainty factor levels
$c^{j_1 j_2}(k_{j_1}, k_{j_2})$	Consistency of pair (k_{j_1}, k_{j_2}) of levels of uncertainty factors j_1, j_2
$\mathcal{C}$	Set of possible consistencies of pairs of uncertainty factors
$\mathbf{C}^{j_1 j_2}$	Consistencies of the levels of j_1th and j_2th factor as a matrix
$\mathbf{C}$	Consistency table: a block matrix of all matrices $\mathbf{C}^{j_1 j_2}, j_1 \neq j_2$
$\alpha(\mathbf{s})$	Overall consistency of all factor level pairs in scenario $\mathbf{s}$
$\mathcal{S}^I$	A subset of $\mathcal{S}$ with I scenarios of the greatest overall consistency
$\mathbf{S}$	Scenario matrix containing all scenarios of the set $\mathcal{S}$ as its rows
$\mathbf{s}_n$	nth row of scenario matrix $\mathbf{S}$
k_{nj}	Level of jth uncertainty factor of row n of $\mathbf{S}$
$\mathbf{1}_q$	A $q \times 1$ vector of ones
$\mathbf{c}^{j_1, j_2}$	Vector of consistencies of factors j_1, j_2 of scenario matrix $\mathbf{S}$ rows
$\boldsymbol{\alpha}$	Overall consistencies of rows of $\mathbf{S}$
$\mathbf{n}^I$	Row indices of I most consistent scenarios in $\mathbf{S}$
$\mathbf{S}^I$	Matrix of the I rows of $\mathbf{S}$ with the greatest consistencies
$\mathbf{s}_i^I$	ith row of matrix $\mathbf{S}^I$
$\boldsymbol{\alpha}^I$	Overall consistencies of rows of $\mathbf{S}^I$
η	Computational parameter for consistency value extraction

$\mathbf{x}_i$	Disjunctive vector of scenario $\mathbf{s}_i^I$
k	Category, i.e., a column index of vectors $\mathbf{x}_i$
$\mathbf{X}$	Complete disjunctive table of I disjunctive scenario vectors $\mathbf{x}_i$
I_k	Number of scenarios in $\mathbf{X}$ belonging to category k
p_k	Proportion of scenarios in $\mathbf{X}$ belonging to category k
$\mathbf{P}$	Diagonal matrix of proportions p_k
$d(\mathbf{x}_{i_1}, \mathbf{x}_{i_2})$	Distance of consistent scenarios i_1 and i_2
$\mathcal{X}$	Point cloud of consistent scenarios in K-dimensional space
$\mathbf{g}$	Center of gravity of point cloud $\mathcal{X}$
Inertia($\mathcal{X}$)	Inertia of point cloud $\mathcal{X}$
$\mathbf{Z}$	Transformed complete disjunctive table of consistent scenarios
$\mathbf{z}_i$	ith row of $\mathbf{Z}$
$\mathbf{\Sigma}$	Diagonal matrix of singular values of $\mathbf{Z}$
σ_r	rth singular value of $\mathbf{Z}$
$\mathbf{U}$	Matrix of left singular vectors of $\mathbf{Z}$
$\mathbf{u}_r$	rth column of $\mathbf{U}$
$\mathbf{V}$	Matrix of right singular vectors of $\mathbf{Z}$
$\mathbf{v}_r$	rth column of $\mathbf{V}$, i.e., rth principal dimension
ρ	Number of principal dimensions selected
$\mathcal{V}$	Plane onto which the point cloud $\mathcal{X}$ is projected
$\mathbf{Y}$	Matrix of ρ principal coordinates of $\mathbf{Z}$
$\mathbf{y}_i$	ith row of $\mathbf{Y}$, i.e., ith vector of ρ principal coordinates
$\mathcal{Y}$	Set of vectors of ρ principal coordinates
$PI(r)$	Percentage of inertia explained by rth principal dimension

Contents

Chapter 1

Introduction

The future operational environment of companies and other organizations is inherently uncertain. Moreover, this operational environment is complex and multidimensional, affected by possibly unexpected developments on, e.g., political, economic, and social dimensions. To retain their competitive edge, companies must develop methods and strategies to increase their understanding of this uncertain and complex future so that they can respond to changes in the operational environment with high-quality strategic decisions [Vilkkumaa et al. (2018)].

Traditional strategic planning approaches rely on the extrapolation of historical data without considering the effect of the unprecedented [van der Heijden (1996)]. Complex and highly uncertain environments render such approaches inadequate [Bunn and Salo (1993)], because political, environmental, economic and/or societal changes with the greatest impacts on businesses are often those that were not deemed as the most likely. For this reason, instead of focusing on the most expected future and searching for the one 'optimal' strategy in this regard, *scenario planning* considers a set of internally consistent but sufficiently diverse plausible futures, called *scenarios* [Schoemaker (1993)]. Specifically, scenario planning acknowledges that no single scenario can provide an adequate description of the future; instead, the uncertainties in the alternative futures call for a more robust approach and consideration of multiple trends.

One approach to constructing scenarios is by depicting them with regard to several *uncertainty factors* that reflect key uncertainties of future change [van der Heijden (1996)]. These uncertainty factors have multiple different outcomes, that is, factor *levels*, which give detailed descriptions of the development of change. Scenarios are then constructed as a combination

of factor-specific levels. The internal consistency of each scenario can be assessed by evaluating the consistencies between the levels of each pair of uncertainty factors. Then, a small number of internally consistent and sufficiently dissimilar scenarios can be selected, from which *scenario narratives* are written to create tangible depictions of the future, often amplified with images that help understand the type of future each scenario depicts.

Using this approach, the number of alternative scenarios grows exponentially with the number of uncertainty factors. For example, if there are 10 uncertainty factors that each have 4 different levels, there are $4^{10} \approx 1\ 000\ 000$ possible combinations of factor specific levels, thus rendering the inspection of all possible scenarios by hand impossible. This notwithstanding, there are but a few systematic methods to support the selection of a small number of internally consistent and mutually dissimilar scenarios from an exponentially large set of alternative scenarios. Hence, many authors recommend limiting the number of uncertainty factors and their levels [Pillkahn (2008),van der Heijden (1996)]. Such a limitation may, however, inhibit the scenario building process by restricting the exploration of possible futures [Wright et al. (2013), Lord et al. (2016)].

This thesis develops efficient methods for identifying the most consistent and mutually dissimilar scenarios when the number of uncertainty factors is arbitrarily large. In particular, the thesis develops an explicit enumeration algorithm that solves a very large consistency value extraction problem efficiently and filters the most consistent ones (e.g., 1000 or 10 000) out of exponentially many candidate scenarios. Then, to help select a small number (e.g., 3-6) of mutually dissimilar scenarios from the set of consistent scenarios, the thesis develops a method to visualize the consistent scenarios using a *Scenario Map*, which links each consistent scenario to a coordinate in a two-dimensional plane. The two-dimensional visualization is based on Multiple Correspondence Analysis, which is the optimal dimensionality reduction method for high-dimensional, categorical data [Greenacre and Blasius (2006), Greenacre (2017), Husson et al. (2017)]. Finally, the methods developed in this thesis are applied to a real case in which a set of plausible futures is identified for the Finnish National Emergency Supply Organization to aid their strategic decision making.

The rest of this thesis is structured as follows. Chapter 2 presents relevant scenario planning and dimensionality reduction literature for this thesis. Chapter 3 provides a mathematical formalization of scenarios and presents algorithms for evaluating the consistencies of scenarios and filtering out

the most consistent ones. Chapter 4 presents the dimensionality reduction method for visualizing the most consistent scenarios in a two-dimensional plane. In Chapter 5, we present the case study in which the methodology was successfully used to a identify set of scenarios for the National Emergency Supply Organization of Finland. Chapter 6 discusses the advantages and limitations of the method and also presents ways to extend our method. Chapter 7 concludes.

Chapter 2

Literature review

2.1 Approaches to identify scenarios

Traditionally, different scenario planning methodologies have been divided into three schools [Bradfield et al. (2005)], which are the *intuitive logics* school, the *probabilistic modified trends school*, and *La Prospective*. The intuitive logics model is a workshop-focused approach, which provides a good mix of sophistication and ease of use, and is hence popular among scenario practitioners and academics [Wright et al. (2013), Bowman (2016)]. The second school, the probabilistic modified trends, seeks to quantify expert opinions with different methodologies, such as Trend-Impact Analysis (TIA) and Cross-Impact Analysis (CIA). These methodologies attempt to evaluate probabilities of events that cause unprecedented deviations from the historical data [Bradfield et al. (2005)]. The third school, La Prospective, seeks to develop positive images of the future which can serve as a basis for action to guide policy makers [van Vught (1978)]. From a methodological point of view, this school can be viewed as a 'blend of tools and systems analysis' [Godet (2000)] or, alternatively, as a blending of the intuitive logics and probabilistic modified trends methodologies [Bradfield et al. (2005)].

While the division into these three schools is useful when considering the history of scenario analysis, this division is somewhat crude from a methodological point of view. In particular, the most popular school in the literature [Bradfield et al. (2005)], the intuitive logics, suffers from a 'methodological chaos' [Martelli (2001)], because there are almost as many different scenario planning approaches as there are practitioners. For this reason, we study the scenario planning literature from a methodological point of view, that is, by considering a division into *deductive* and *inductive* approaches [Ogilvy and

Schwartz (1998), Bowman et al. (2013), McBride et al. (2017)]. We compare these approaches with regard to four important aspects of the scenario development technique selection, as identified by McBride et al. (2017). These aspects are presented in Table 2.1.

Table 2.1: Four important aspects of the selection of an adequate scenario development technique [McBride et al. (2017)].

Objectives of scenario development process selection	
(i) Explorativeness:	To enable thinking of the unprecedented, the process must be sufficiently explorative.
(ii) Trustworthiness:	To establish trust in the developed scenarios, the process must be transparent and consensus supporting.
(iii) Efficiency:	To develop scenarios in a timely and efficient manner, the process must be well structured.
(iv) Accessibility:	To ensure that the stakeholders can contribute their best expertise, the process should not push the stakeholders too far from their capacity.

2.1.1 Deductive approach

In the general sense, deductive scenario planning methods can be considered as processes that build the scenarios in a top-down manner. In other words, in deductive methods, the big picture of the developed scenarios is sketched out in an early phase of the process. McBride et al. (2017) specify deductive approaches as general-to-specific techniques to identify key uncertainties. Similarly, Bowman et al. (2013) describe deductive processes beginning with the broad framework of the scenarios, then refining and inserting data in these scenarios.

In practice, deductive scenario planning approaches most often utilize a specific 2x2 scenario matrix technique [Bradfield et al. (2005)], originated by Shell [van der Heijden (1996)]. In this technique, two key uncertainties are represented by so-called scenario axes, the extremes of which correspond to the two possible outcomes of these uncertainties. Then, four skeletal scenarios are developed as combinations of the outcomes of these uncertainties [Schwartz (1991)]. The selection of the two key uncertainties from a large number of factors is based on their uncertainty and relevance. Ogilvy and Schwartz (1998) present the so-called poker chips narrowing exercise as an appropriate method for selecting the two key uncertainties. In this method, the

participants of the scenario workshop assign, e.g., 25 poker chips on different factors, based on their uncertainty and relevance. Then, the two factors with the most chips are selected for building the four skeletal scenarios in the 2x2 scenario matrix. These scenarios are further developed by considering what outcomes of each of the remaining uncertainty factors are deemed plausible in the four quadrants of the matrix.

Van't Klooster and van Asselt (2006) conducted a study in which they followed professional futurists who applied the 2x2 scenario matrix technique in their qualitative scenario projects. In this study, three distinct views of the scenario axes were identified, namely, scenario axes as (i) a backbone, (ii) a building scaffold, and (iii) a foundation for the scenario development. According to the first view (backbone) the two most uncertain and relevant factors are 'out there', and that as such the scenario matrix provides plausible frames for the developed scenarios. Based on the second view (building scaffold), the scenario axes can be abandoned during scenario development in order to provide room for more integrated scenarios. The third view (foundation) is a mixture of the former two in that it does not consider the scenario axes as blueprint, but instead advocates the axes with social and methodological arguments. That is, although the scenario axes could even lead to implausible scenarios, the axes serve a functional meaning as enablers of discussion on four divergent scenarios.

The deductive scenario development approach has several limitations with respect to the aspects in Table 2.1. With respect to the explorativeness aspect, the deductive 2x2 scenario matrix approach is limited in that (a) focusing on the extremes of axes can drive unnecessary polarization in thinking and (b) limiting the number of uncertainty factors at an early stage may pre-emptively restrict the exploration of the future possibility space [Wright et al. (2013), Lord et al. (2016)]. With respect to the trustworthiness aspect, the deductive approach can be limited as well. More specifically, reaching consensus on the selection of the two scenario axes may be difficult, which may lead to a lack of trust by the process stakeholders in the scenarios resulting from these axes [van der Heijden (1996)]. For example, van't Klooster and van Asselt (2006) observed controversy in the selection of the scenario axes in the scenario projects they followed. Stakeholders of scenario workshops felt a lack of trust in the selected axes, their critiques including, for example, that 'the scenario axes represented a very classical scheme', and that the selection criteria of the two key uncertainties were not transparent.

2.1.2 Inductive approach

In contrast to deductive approaches, inductive scenario planning methods use specific-to-general techniques for building scenarios [McBride et al. (2017)]. In other words, inductive scenario development processes identify scenarios in a bottom-up manner such that the uncertainty factors are not restricted to a small manageable amount. Instead, inductive approaches recognize the scale of the scenario planning problem and seek to include all the relevant aspects of the problem at hand throughout the whole scenario exercise. Hence, the inductive approaches are less prone to restricting explorative thinking, which is evidenced by many successful inductive scenario planning studies [Ritchey (2011), Bowman et al. (2013), Vilkkumaa et al. (2018), Johansen (2018)]. In these studies, entire solution spaces and all possible solutions of scenario problems have been systematically explored [Johansen (2018)]. Consequently, broad thinking and the retaining of an explorative mindset have been supported [Wright et al. (2013)], because the point of view to the future is not narrowed down to just two key uncertainties of change. In the following subsections, we explicate some inductive scenario planning processes used in the literature.

Unstructured approach

Typically, inductive methods for scenario development are relatively unstructured. This is because the strengths of the inductive approach have been seen to arise from the process being open-ended and exploratory, such that scenarios emerge from in-depth discussions about individual events, and the more broad scenario storylines are then developed organically [McBride et al. (2017)]. McBride et al. (2017) note that by having a broad range of plot elements available, the inductive approaches yield compelling plot lines that can focus on the relevant strategic decisions at hand, depending on the particular case study. Moreover, they explain that by having direct connection to plausible events, unstructured inductive methods can effectively link the developed scenarios to relevant strategic decisions in the present [van Vliet et al. (2012)].

Bowman et al. (2013) used an unstructured inductive approach in developing two scenarios for Scottish Local Authorities to improve community planning. They describe their approach as '(a) process (that) begins with the granularity of available data and allows the scenarios to emerge incrementally'. This process began with teasing out key trends and discussing the merits and inherent uncertainties of these trends until consensus was reached. Later, the

scenario narratives were written with the help of professional storytellers. The scenario process of Bowman et al. (2013) was highly participative and engagement-oriented, which is characteristic to inductive processes [Wright (2005)].

Although the inductive approaches by McBride et al. (2017) and Bowman et al. (2013) are relatively unstructured, they can be characterized as variants of the Emblematic events approach [Ogilvy and Schwartz (1998)]. In this approach, the group of scenario workshop participants brainstorms various typical events for different scenarios. Then, these events work as plot elements, from which larger stories are created. The stories can be developed by considering, for example, what might lead up to such events, or, what plausible consequences the events might have.

McBride et al. (2017) acknowledge that inductive methods, such as the Emblematic events approach, are effective with respect to the exploratiVeness aspect in the selection of scenario development technique, as presented in Table 2.1. This notwithstanding, they preferred the traditional deductive 2x2 matrix approach, because of the limited timeframe of their scenario exercise and concerns that the inductive process might be unsuccessful due to its unstructured nature. If the inductive scenario development exercises are carried out in an unstructured manner [as in, e.g., Bowman et al. (2013)], the process can become more opaque and dependent on the creativity and imagination of the participants. This unstructured nature of inductive processes can cause greater time and facilitation demands and even risk the success of the scenario exercise [Volkery and Ribeiro (2009)]. Thus, inductive scenario development approaches can be inefficient, having limitations with respect to the third aspect (efficiency) of scenario technique selection in Table 2.1. Moreover, opaqueness in building scenarios can undermine trust in the developed scenarios, and thus an unstructured inductive process can have limitations with respect to the second aspect (trustworthiness) of Table 2.1 as well.

General Morphological Analysis

One approach to introduce structure to an inductive scenario planning exercise is General Morphological Analysis (GMA) [Ritchey (2006)]. In the first uses of morphological analysis [in, e.g., Zwicky (1948)], the method was used in the investigation of physical shapes, such as astrophysical objects. Later, Zwicky (1969) proposed to generalize the concept of morphological research to study the shapes of more abstract structural interrelations. In the

same year, Ayres (1969) pointed out that morphological analysis could be employed to provide a systematic approach for scenario development. In this context, GMA is used to model scenarios as multidimensional combinations of qualitative, non-quantifiable levels of different factors that characterize future uncertainty. Rhyne (1971) was the first to employ GMA as a scenario development technique, although under a different name, Field Anomaly Relaxation (FAR).

Today, GMA is used in modern scenario planning studies to provide structure to inductive scenario development processes [Ritchey (2011), Vilkkumaa et al. (2018), Johansen (2018)]. Johansen (2018) illustrates the use of GMA in scenario planning with an example case from Norwegian defence planning. They succeeded in establishing a diverse set of scenarios that encompass all possible security challenges in Norway. In this process, GMA was complemented with cross consistency assessment where scenarios containing inconsistent combinations of uncertainty factor levels are ruled out, thus reducing the plausible scenarios to a manageable number. Vilkkumaa et al. (2018) use GMA in developing scenarios to support the strategy process of a group of Nordic steel and engineering companies looking to establish a multi-sided platform ecosystem. In their study, GMA was augmented by consistency analysis in which the strengths of consistencies between pairs of different uncertainty factor levels were assessed as well.

Some authors [e.g., Lord et al. (2016)] distinguish between morphological analysis and inductive scenario planning methods. In particular, Lord et al. (2016) define the inductive approach as a method that generates scenarios in an emergent, organic fashion, so that scenario narratives are identified first, and only then structure is brought to the scenario set [Wilkinson and Kupers (2014)]. Nevertheless, even when defined this way, the inductive method bears important similarities with GMA in that the scenarios are developed from individual plot elements and that the diversity of the scenario set is considered in terms of the solution space as a whole.

Formative scenario planning methods

In GMA, the size of the scenario space to consider quickly grows too large to be inspected by hand. For example, a typical GMA problem with 6-10 uncertainty factors can result in 50 000 or even 5 000 000 candidate scenarios [Ritchey (2006)]. Classical GMA approaches use cross-consistency assessment to filter out a large number of scenarios. Nevertheless, even after such filtering, the number of candidate scenarios can remain excessively large, as

the goal of scenario planning interventions is to produce only a handful of scenarios. Some scenario planning methods that utilize GMA are *formative* [Lord et al. (2016), Carlsen et al. (2016), Tietje (2005), Brauers and Weber (1988)] in that they seek to fully define a set 3-6 of consistent and mutually dissimilar scenarios. Such formative methods use quantitative assessments given by subject experts on the consistencies or compatibilities of the levels of different uncertainty factors, and mathematical methods of optimization [Lord et al. (2016), Carlsen et al. (2016), Tietje (2005)] and cluster analysis [Brauers and Weber (1988)] to identify the final set of scenarios.

Lord et al. (2016) present an optimization framework called OLDFAR (Optimized Linear Diversity Field Anomaly Relaxation) for selecting n sets of k scenarios from morphological analysis. In this approach, experts are first asked to fill in a FAR compatibility matrix, the values of which reflect the compatibilities between the levels of different uncertainty factors. This matrix is then used to define a set of plausible scenarios, which may contain hundreds or thousands of candidate scenarios. Finally, a linear optimization problem is formulated to find those n scenario sets with k plausible scenarios that maximize (i) the internal diversity within each scenario set and (ii) the diversity between the n scenario sets. A similar approach is developed by Carlsen et al. (2016) who use optimization to identify scenario sets that are maximally diverse in a methodology which they refer to as Scenario Diversity Analysis (SDA).

Tietje (2005) proposes several procedures for formative scenario analysis to build scenario sets that are internally consistent and contain diverse scenarios. These procedures start from a consistency matrix that specifies the consistencies of all pairs of levels of different uncertainty factors. Scenarios are defined as vectors that specify one level for each uncertainty factor, and a consistency of a scenario is evaluated by aggregating the consistencies of pairs of factor levels in the scenario with, for example, addition or multiplication. Then, Tietje (2005) presents approaches to filter a large scenario space - the set product of all levels in all uncertainty factors - according to overall consistencies of the scenarios. From the filtered set of scenarios, methods are presented to further narrow the set to only a handful of scenarios by compromising between maximizing the distances and consistencies of the final set of scenarios. The scenario narratives, amplified with effective visual graphical representations, can then be developed from this small set of scenarios.

Brauers and Weber (1988) present a three-stage scenario identification method for determining plausible scenarios. In the first stage, a compatibility ma-

trix is used to identify those scenarios with least incompatibilities, which are deemed plausible. The next stage of their method calculates the probabilities of the plausible scenarios by solving a linear optimization problem, which is based on the compatibility matrix and marginal probabilities of uncertainty factor levels. In the third stage, cluster analysis is performed in the set of plausible scenarios to identify centres of the scenario clusters, which then define the main scenarios. These main scenarios are often combinations of multiple plausible scenarios and are not necessarily representative of any single compatible scenario. The probabilities of the main scenarios are the sums of the probabilities of all scenarios in the cluster.

The formative methods discussed above seek to find dissimilar scenarios. In these methods, a scenario is represented by a vector or a tuple, whereby the dissimilarity between two scenarios can be represented by the distance between these scenarios, defined through a suitable metric [Kaplansky (1972)]. The definition of such a metric is not at all trivial, because the levels of the factors in the scenarios are usually non-quantifiable [Ritchey (2006)]. In other words, the variables reflecting the levels of uncertainty factors in a scenario are often *categorical*, and hence they bear no intrinsic numerical values. Yet, many authors evaluate the distances between scenarios with a metric that requires quantifiable scenario variables. For instance, Lord et al. (2016) evaluate distances of scenarios using differences in the numerical values assigned to each level of the uncertainty factors. As a result, the numerical values given to the levels have an effect on the dissimilarity assessment, even though these values are only assigned for the purposes of problem formulation and are not related to the levels of the morphological analysis in any meaningful way. Similarly, Brauers and Weber (1988) use Euclidean distances of vectors of factor levels in their cluster analysis, which causes the numerical values assigned to the uncertainty factor levels to affect the analysis. While Carlsen et al. (2016) acknowledge that the scenario variables are categorical, they still use a similar distance metric as Lord et al. (2016), thus implying ordinality of the variables. A proper distance metric for categorical variables is used by Tietje (2005). In this metric, the sizes of differences in the numerical values assigned to scenario variables do not matter, but instead, it only matters whether the variables differ from one another.

The above formative approaches also seek to find internally consistent scenarios, usually by evaluating the consistency of a scenario by aggregating the pairwise consistencies between the uncertainty factor levels included in the scenario. However, evaluating these consistencies for a large number of scenarios can be computationally intensive, because typical GMA problems

with around 10 uncertainty factors can result in millions of candidate scenarios. As a result, the number of consistency evaluations of all pairs of factor levels for all scenarios can be hundreds of millions, which can be very time consuming even for modern computers. Many scenario planning studies that use GMA and formative methods circumvent this issue by either (i) using an exclusive overall consistency measure that discards large amounts of scenarios based on discrepancies between only one pair of factors [e.g, Tietje (2005), Johansen (2018)] or (ii) limiting the number of considered uncertainty factors to a small manageable amount [e.g., Lord et al. (2016), Brauers and Weber (1988)]. These approaches are problematic, however, first because excluding scenarios based on one pair of factors can effectively make it redundant to assess strengths of consistencies, because even if nearly all factors pairs in a scenario are strongly consistent, one low consistency pair will render a scenario implausible. Secondly, limiting the number of uncertainty factors at an early stage effectively causes the formative inductive process to restrict the exploration of the future possibility space in the same way as in the deductive 2x2 scenario matrix approach, although to a lesser extent.

Another challenge with formative scenario planning methods is related to the transparency and accessibility of these methods. While mathematical methods are necessary to provide structure and reliability into the analysis, formative scenario planning methods can in their purest form make the analysis opaque. Not all scenario practitioners and stakeholders of scenario interventions have backgrounds in mathematical fields, and hence a purely formative planning method may hide the reasoning behind the analysis from many participants of a scenario workshop, thus potentially reducing the transparency of the scenario exercise. This reduced transparency can create a lack of trust in the developed scenarios, which may reduce the effectiveness of the scenario planning intervention. Moreover, McBride et al. (2017) argue that complex opaque modelling processes can limit adequate stakeholder understanding and engagement in the process. Stakeholders that are unable to understand the process may not be able to contribute their best expertise. Hence, purely formative scenario planning methods can be limited with regard to the second and fourth aspects (trustworthiness and accessibility) of the scenario development technique selection in Table 2.1. Nevertheless, means to effectively communicate the results to the scenario intervention participants can improve the transparency of mathematical modelling. For example, comprehensive visualizations can provide the crucial means to create trust in the results of the mathematical models.

2.2 Dimensionality reduction

In GMA, scenarios are represented by multidimensional vectors of uncertainty factor levels. Hence, it is not possible to directly visualize the scenario space by, for example, representing a set of scenarios as a point cloud using a two-dimensional scatter plot. Similar issues are faced by statisticians and machine learning practitioners when they seek important properties of large, high-dimensional data sets. One common approach is to find a representation for the high-dimensional data set in a lower dimension by using *dimensionality reduction methods* [Alpaydin (2014), Leskovec et al. (2014), Greenacre and Blasius (2006), Greenacre (2017), Husson et al. (2017)]. While the high-dimensional point clouds cannot be perfectly displayed in a much lower-dimensional (e.g., two-dimensional) space, methods have been developed to produce representations that capture the most important properties of the data sets in their original space. As a result, the high-dimensional point cloud of a data set can be inspected in a two-dimensional plane using a scatter plot, and inference can be made based on the distances of points in the projected clouds.

Different dimensionality reduction methods can be divided into two main categories: dimension selection and dimension extraction [Alpaydin (2014)]. In dimension selection, few of the multiple dimensions of the data with most information are selected. In dimension extraction, the objective is to find new dimensions that are combinations of the original dimensions of the data.

Examples of dimension selection methods are subset selection and CUR decomposition. In subset selection, the goal is to find a subset of dimensions in the data that performs best in a regression problem [Kohavi and John (1997)]. In CUR decomposition, combinations of columns and rows of the data matrix are randomly selected to construct two matrices C and R such that rows and columns with greatest diversity have a greater chance of being selected. A third matrix, U, is then constructed such that the original data can be approximated with the product of the three matrices, and in this sense the method seeks to preserve the most important properties of the original data [Leskovec et al. (2014)].

Dimension extraction methods can be linear or nonlinear in that the new dimensions can be either linear or nonlinear combinations of the original dimensions. The most common, linear dimension extraction method is principal component analysis (PCA) [Alpaydin (2014), Leskovec et al. (2014),

Greenacre and Blasius (2006), Husson et al. (2017)], in which the goal is to find a linear map from the data in the original d-dimensional space to a new k-($< d$)dimensional space so that the distances between data points are distorted as little as possible. The data of PCA is represented with Cartesian coordinates, and the linear map is done by projecting the data points to a plane that passes the centroid of the data. The plane is defined by the first k eigenvectors of the correlation matrix of the data, or alternatively, by the first k right singular vectors of the data matrix, both corresponding to the k greatest eigenvalues or singular values.

Linear dimension extraction methods assume that the data is well representable in a linear subspace, which assumption may not hold in all cases. For example, if the data defines manifolds in a d-dimensional space, a linear projection will not capture the distances between data points with good precision [Alpaydin (2014)]. In such cases, various nonlinear dimensionality reduction methods will perform better. Isometric feature mapping [Tenenbaum et al. (2000)] calculates distances among a manifold, instead of Euclidean distances, and then applies dimensionality reduction such that these distances are best preserved. Laplacian Eigenmap is another nonlinear dimensionality reduction method, which only considers local similarities in the data, that is, it seeks a nonlinear subspace of the original space that has minimal (nonzero) local distances within neighborhoods of data points [Belkin and Niyogi (2003)].

Dimension extraction methods also differ based on whether the data is quantitative or qualitative. Methods such as PCA, Isometric feature mapping, and Laplacian Eigenmaps utilize distance metrics that are most suitable for quantitative data, which is why it is unsound to apply of these methods to non-quantifiable, categorical data, such as a set of scenarios represented as vectors of factor-specific levels. A widely used dimension extraction method for non-quantitative sets of data is Correspondence Analysis (CA) [Greenacre and Blasius (2006), Greenacre (2017), Husson et al. (2017)]. CA analyses the two-way cross-tabulation of levels in a data set of two categorical variables, and this cross tabulation is analysed simultaneously as a set of rows and a set of columns. Visualizations of the point cloud of the rows and point cloud of the columns are then created by projecting them onto planes that preserve the maximal amount of inertia, defined as the mass weighted squared distance of the points from their center of gravity. As such, CA is close to PCA, with the difference mainly being in how the point clouds are constructed. Nevertheless, because CA is based on a two-dimensional table, data sets of at most two categorical variables can be represented. Hence, in

scenario analysis, CA can only be used for GMA with two uncertainty factors.

An effective dimensionality reduction method that handles multidimensional categorical data is Multiple Correspondence Analysis (MCA) [Greenacre and Blasius (2006), Greenacre (2017), Husson et al. (2017)]. MCA is an extension of Correspondence Analysis that can be considered as the analysis of all two-way cross-tabulations among the variables. In this method, separate dimensions are not only given to the categorical, non-quantifiable, variables, but instead separate dimensions are defined to represent each distinct category. For instance, in the context of scenarios that are represented as vectors of levels of uncertainty factors, each distinct level of each uncertainty factor would be represented as a separate dimension in MCA. With this encoding of the data, similar means as in Correspondence Analysis can be taken to represent the high-dimensional, categorical data in a lower dimension.

MCA provides a way to effectively visualize a large number of scenarios in a two-dimensional plane. Such visualizations can then be used to guide the process of selecting the final scenarios for analysis. Yet, to our knowledge, no studies exist that would have explored the possibilities of MCA in scenario visualization.

Chapter 3

Identifying consistent scenarios

In this chapter, we present the model framework for identifying the most consistent scenarios out of a large set of so-called *skeletal scenarios*. In particular, Section 3.1 presents the General Morphological Analysis framework for formally defining the set of skeletal scenarios, the internal consistencies of these scenarios, and the most internally consistent scenarios among this set. Efficient algorithms for finding the set of I (e.g., $I = 100, 1\ 000, 10\ 000$) most consistent scenarios are presented in Section 3.2.

3.1 Modelling framework

In Subsection 3.1.1, we present the modelling framework for scenarios using GMA, and explain how the set of all possible scenarios can be generated. In Subsection 3.1.2, we show how the plausibility of the possible scenarios can be studied by assessing the consistencies between the levels of different uncertainty factors in these scenarios. The rationale for identifying the set of most consistent scenarios based on these consistency assessments is presented in Subsection 3.1.3.

3.1.1 All possible scenarios

The process for generating the set of all possible scenarios through GMA begins with the definition of key uncertainty factors that drive the change in the operational environment. Each uncertainty factor is associated with a set of levels that represent the possible outcomes of these factors. A *morphological field* is constructed by tabulating these levels in consecutive columns.

Let us illustrate this process through an example on building scenarios for a Finnish electricity sales company. The company is analysing the external uncertainty factors affecting the development of their product and service portfolio and business model. Here, uncertainty related to the electricity markets is seen to be characterized by the six uncertainty factors shown in Table 3.1. The possible levels of these uncertainty factors are presented in Table 3.2.

Table 3.1: Uncertainty factors of the electricity sales scenario planning example.

A. Energy regulation's focus
B. Electricity price
C. Competitive field
D. Activity of switching electricity supplier
E. Digitalization and technology
F. Finnish economy

Table 3.2: One skeletal scenario of the electricity sales morphological field.

Energy regulation's focus	**Electricity price**	**Competitive field**	**Activity of switching supplier**	**Digigalization & technology**	**Finnish economy**
Environment & renewable energy	Low, under 30 €/MWh	Traditional: private & municipal	Low, under 8 %/year	Digital evolution	Deep recession
Energy security & reliability	Moderate, 30-45 €/MWh	Consolidation	Moderate, 9-14%/year	Fast digitalization	Zero growth
Market-based energy industry	High, over 45 €/MWh	International competitive field	High, over 15%/year	Digital revolution	Strong growth
Citizens: empowerment & protection	Turbulent, 0-200 €/MWh	New players from different industries			

A skeletal scenario is represented by a combination of levels on different uncertainty factors. One such skeletal scenario is marked by purple cells in

Table 3.2. The actual scenarios developed in the scenario exercise as a whole would be fleshed out from these kinds of skeletal scenarios using narratives and visualizations of the future, and the scenarios resulting from the following analyses should simply serve as starting points for storytelling. For brevity, we shall refer to the skeletal scenarios as scenarios from this point forward.

To obtain a mathematical representation for the scenarios in GMA, we denote the uncertainty factors with indices $j = 1, 2, \ldots, J$. The level of uncertainty factor j is represented by a *categorical variable* k_j, defined as follows.

Definition 1. ***(Categorical variable of an uncertainty factor)*** *Let K_j be the number of levels of uncertainty factor j. The level of uncertainty factor j is represented by categorical variable*

$$k_j \in \mathcal{K}_j = \{1, 2, \ldots, K_j\}. \tag{3.1}$$

It should be specified that neither the ordinarity nor the distances of the assigned values $1, 2, \ldots, K_j$ bear any meaning. The vector of the numbers of uncertainty factors is denoted by $\mathbf{K} = \begin{bmatrix} K_1 & K_2 & \cdots & K_J \end{bmatrix}$, and the total number of uncertainty factor levels is $K = \sum_{j=1}^{J} K_j$.

A scenario is modelled as a combination of levels on different uncertainty factors j. More formally:

Definition 2. ***(Scenario)*** *A scenario* $\mathbf{s}$ *is a vector of values of categorical variables k_j representing the levels of uncertainty factors $j = 1, 2, \ldots, J$:*

$$\mathbf{s} = \begin{bmatrix} k_1 & k_2 & \cdots & k_J \end{bmatrix} \in \mathcal{S}, \tag{3.2}$$

where $\mathcal{S}$ is the set of all possible scenarios.

The set of all possible scenarios $\mathcal{S}$ is defined as follows.

Definition 3. ***(Set of all possible scenarios)*** *Let $\mathcal{K}_j$ be the sets of levels of each uncertainty factor $j = 1, 2, \ldots, J$. Then, the set of all possible scenarios is the set product*

$$\mathcal{S} = \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_J. \tag{3.3}$$

The total number of possible scenarios is the product of all numbers K_j of levels $N = |\mathcal{S}| = \prod_{j=1}^{J} K_j$.

As an example, consider the morphological field in Table 3.2. Three of the six uncertainty factors have four possible levels $K_j = 4, j = 1, 2, 3$ and the rest have three possible levels $K_j = 3, j = 4, 5, 6$. Thus, the set all possible scenarios is

$$\mathcal{S} = \bigtimes_{j=1}^{3} \{1, 2, 3, 4\} \times \bigtimes_{j=4}^{6} \{1, 2, 3\}.$$

The scenario illustrated in Table 3.2 is $\mathbf{s} = \begin{bmatrix} 4 & 3 & 4 & 3 & 2 & 3 \end{bmatrix}$, and the total number possible scenarios is $N = \prod_{j=1}^{6} K_j = 1728$.

3.1.2 Consistency assessment

To assess the internal consistency of a scenario, the consistencies between each pair of levels on two uncertainty factors are rated. Then, the consistency of the scenario as a whole is assessed based on these pairwise consistencies. In particular, let us denote by $c^{j_1 j_2}(k_{j_1}, k_{j_2}) \in \mathcal{C}$ the consistency between levels (k_{j_1}, k_{j_2}) of uncertainty factors $j_1, j_2 = 1, 2, \ldots, J, j_1 \neq j_2$. These consistencies are rated on scale $\mathcal{C} \in \{-3, -2, -1, 0, 1, 2, 3\}$ [Scholz and Tietje (2001)], the interpretation of which is given in Table 3.3.

Table 3.3: Definitions of consistency indicators. The interpretations follow those given in Scholz et al. (1999).

Consistency	Explanation	
-3	Strongly inconsistent:	It is very unlikely that the levels occur at the same time
-2	Inconsistent:	It is moderately unlikely that the levels occur simultaneously
-1	Slightly inconsistent:	The levels hinder each other, but they can occur simultaneously
0	Independent:	The levels occur independently; they have no direct relation
1	Slightly consistent:	The occurrence of one level supports the occurrence of the other
2	Consistent:	The occurrence of a level strongly supports the other
3	Strongly consistent:	The occurence of a level induces the other

Thus, each pair of uncertainty factors j_1, j_2 is associated with a $K_{j_1} \times K_{j_2}$ rectangular matrix of pairwise consistency values.

Definition 4. ***(Table of pairwise consistencies)*** *Let $c^{j_1 j_2}(k_{j_1}, k_{j_2}) \in \mathcal{C}$ be the consistency of the pair (k_{j_1}, k_{j_2}) of levels $k_{j_1} \in \{1, 2, \ldots, K_{j_1}\}$ and $k_{j_2} \in \{1, 2, \ldots, K_{j_2}\}$ of uncertainty factors $j_1, j_2 = 1, 2, \ldots, J, j_1 \neq j_2$. The table of pairwise consistencies between all levels of uncertainty factors j_1 and j_2 is*

$$\mathbf{C}^{j_1 j_2} = \left[c^{j_1 j_2}(k_{j_1}, k_{j_2})\right] \in \mathcal{C}^{K_{j_1} \times K_{j_2}}, \tag{3.4}$$

where k_{j_1} refers to the row index and k_{j_2} refers to the column index of an element in matrix $\mathbf{C}^{j_1 j_2}$.

Once the consistency ratings of all pairs of uncertainty factor levels have been assessed, the ratings can be summarized in a *consistency table* $\mathbf{C}$. Because the consistency indicator is symmetric, the consistency table reduces to an upper triangular $(J-1) \times (J-1)$ block matrix, which is defined below.

Definition 5. ***(Consistency table)*** *Let $\mathbf{C}^{j_1 j_2}$ be the tables of pairwise consistencies of uncertainty factors $j_1 = 1, 2, \ldots, J-1, j_2 = 2, 3, \ldots, J, j_1 \neq j_2$. The consistency table of the J uncertainty factors is the $(J-1) \times (J-1)$ upper triangular block matrix of all consistency values:*

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^{12} & \mathbf{C}^{13} & \cdots & \mathbf{C}^{1J} \\ & \mathbf{C}^{23} & \cdots & \mathbf{C}^{2J} \\ & & \ddots & \vdots \\ & & & \mathbf{C}^{J-1J} \end{bmatrix}. \tag{3.5}$$

Thus, the total number of rows in the consistency table is $\sum_{j_1=1}^{J-1} K_{j_1}$, and the total number of columns is $\sum_{j_2=2}^{J} K_{j_2}$. Each block $\mathbf{C}^{j_1 j_2}$ in the j_1th block row and j_2th block column of the consistency table is of size $K_{j_1} \times K_{j_2}$, and these blocks define the consistencies of the level pairs in uncertainty factors j_1 and j_2.

Consider again the example of analysing the future business environment of the electricity sales company. An illustrative consistency table for this example is given in Table 3.4. For example, consider two uncertainty factors $j_1 = 3$ and $j_2 = 4$, which are C. Competitive field and D. Activity of switching electricity supplier, respectively. The levels $k_3 = 4$ (New players from different industries) and $k_4 = 3$ (High, over 15%/year) of these factors are strongly consistent, because $c^{34}(4, 3) = 3$. On the other hand, the levels

$k_3 = 3$ (International competitive field) and $k_4 = 1$ (Low, under 8%/year) are strongly inconsistent, because $c^{34}(3,1) = -3$.

Table 3.4: Consistency indicators of the electricity sales example.

		B				C				D			E			F		
		1	2	3	4	1	2	3	4	1	2	3	1	2	3	1	2	3
A	1	1	0	0	1	0	1	0	1	-1	-1	0	-2	1	2	-2	0	3
	2	0	-1	-1	1	2	1	-2	-2	1	-1	-1	0	0	0	1	0	1
	3	0	-1	0	1	1	0	2	1	-1	0	1	-2	1	2	2	0	3
	4	-1	-1	1	1	1	0	0	1	0	0	1	0	2	1	2	0	2
B	1					-2	2	-2	1	3	1	-2	1	2	3	3	2	-2
	2					1	1	1	1	1	1	1	2	1	1	1	1	2
	3					1	-2	2	1	-2	2	3	1	1	-3	-1	-1	3
	4					0	0	0	1	-2	1	2	3	2	-2	0	0	0
C	1									3	2	-1	3	1	1	-2	1	2
	2									2	2	2	1	1	1	3	2	1
	3									-3	1	2	-1	2	3	-2	1	2
	4									-3	1	3	-3	2	3	-1	1	2
D	1												3	2	-2	0	0	0
	2												2	3	1	0	0	0
	3												-2	2	2	2	0	0
E	1															3	2	-2
	2															-1	2	2
	3															-3	-1	2

The overall consistency of a scenario can be calculated as the arithmetic mean of the consistencies of all factor level pairs in the scenario [Scholz and Tietje (2001)].

Definition 6. ***(Overall consistency)*** *Let $c^{j_1 j_2}(k_{j_1}, k_{j_2})$ be the elements of the consistency table as in Definitions 4 and 5. Then, the overall consistency of scenario $\mathbf{s} \in \mathcal{S}$ is*

$$\alpha(\mathbf{s}) = \frac{2}{J(J-1)} \sum_{j_1=1}^{J-1} \sum_{j_2=j_1+1}^{J} c^{j_1 j_2}(k_{j_1}, k_{j_2}), \tag{3.6}$$

where constant $2/[J(J-1)]$ divides the sum by the number of its elements.

3.1.3 Most consistent scenarios

Once we have defined the set of all possible scenarios $\mathcal{S}$ and the consistencies of each factor level pair with the consistency table $\mathbf{C}$, the goal is to seek a subset $\mathcal{S}^I$ of I scenarios $\mathbf{s}$ of $\mathcal{S}$ with the greatest overall consistencies $\alpha(\mathbf{s})$. In other words, we seek a partition of the set of all possible scenarios

$$\mathcal{S} = \mathcal{S}^I \cup \complement\left(\mathcal{S}^I\right),$$

where $\complement\left(\mathcal{S}^I\right) = \mathcal{S} \setminus \mathcal{S}^I$ is the complement of the set $\mathcal{S}^I$ containing $\left|\complement\left(\mathcal{S}^I\right)\right| = N - I$ scenarios. Moreover, this partition is made such that no scenario $\mathbf{s}'$ in the complementary set $\complement\left(\mathcal{S}^I\right)$ has a greater consistency $\alpha(\mathbf{s}')$ than any scenario $\mathbf{s}$ in set $\mathcal{S}^I$. Set $\mathcal{S}^I$ is formally defined below.

Definition 7. ***(Set of the** I **most consistent scenarios)*** *Let $\mathcal{S}$ be the set of all possible scenarios and let $\alpha(\mathbf{s})$ be the overall consistency of scenario $\mathbf{s} \in \mathcal{S}$ as in Definition 6. Then, the set of the I most consistent scenarios is*

$$\mathcal{S}^I = \left\{\mathbf{s} \in \mathcal{S} \mid \alpha(\mathbf{s}) \geq \alpha(\mathbf{s}') \forall \mathbf{s}' \in \complement\left(\mathcal{S}^I\right), \left|\mathcal{S}^I\right| = I\right\}. \tag{3.7}$$

For illustrative purposes, we demonstrate the computation of overall consistencies by focusing only on the three following uncertainty factors of the electricity sales example.

C. Competitive field

D. Activity of switching electricity supplier

E. Digitalization and technology

The smaller consistency table for this subproblem is presented in Table 3.5.

Table 3.5: Consistency indicators of a subproblem of the electricity sales example.

		D			E		
		1	2	3	1	2	3
C	1	3	2	-1	3	1	1
	2	2	2	2	1	1	1
	3	-3	1	2	-1	2	3
	4	-3	1	3	-3	2	3
D	1				3	2	-2
	2				2	3	1
	3				-2	2	2

Consider two scenarios $\mathbf{s} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ and $\mathbf{s}' = \begin{bmatrix} 4 & 1 & 1 \end{bmatrix}$. From Table 3.5 and using Equation (3.6), we can evaluate the overall consistencies of these scenarios to be

$$\alpha(\mathbf{s}) = \frac{2}{3(3-2)}(3+3+3) = 3 \text{ and}$$
$$\alpha(\mathbf{s}') = \frac{2}{3(3-2)}(-3-3+3) = -1.$$

Thus, we observe that the pairs of factor levels in scenario $\mathbf{s}$ are strongly consistent on average, and the pairs of factor levels in scenario $\mathbf{s}'$ are slightly inconsistent on average (cf. Table 3.3).

By evaluating the overall consistencies of all possible scenarios in a similar way, we can find the subset $\mathcal{S}^5$ of five scenarios with the greatest overall consistency. This subset is

$$\mathcal{S}^5 = \left\{ \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 4 & 3 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 3 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 3 & 2 \end{bmatrix} \right\}.$$

This subset is unique, because all scenarios with the sixth greatest overall consistencies have $\alpha(\mathbf{s}') = 2$ while for all scenarios $\mathbf{s} \in \mathcal{S}^5$ the overall consistency is $\alpha(\mathbf{s}) \geq 2.333$. However, a subset of $I = 6$ scenarios which have the greatest consistencies would no longer be unique, because there are 7 scenarios which have equal consistency of $\alpha(\mathbf{s}') = 2$, and thus the set $\mathcal{S}^6$ could be made by including any of these scenarios in $\mathcal{S}^5$ such that the requirement (3.7) would be satisfied. In such a situation, the scenarios to include in set $\mathcal{S}^I$ would be selected at random.

3.2 Scenario generation algorithms

Our measure $\alpha(\mathbf{s})$ for overall scenario consistency in Equation (3.6) is compensating in that even strong inconsistency between some pair of uncertainty factor levels can be compensated by the consistencies between other pairs in the scenario. Thus, no scenario can be excluded from the set of most consistent scenarios based simply on inconsistencies of few uncertainty factor levels. Consequently, defining the set $\mathcal{S}^I$ of the I most consistent scenarios requires evaluating the overall consistencies of all possible scenarios $\mathbf{s} \in \mathcal{S}$. Depending on the particular scenario planning case, evaluating the overall consistencies for all possible scenarios may be computationally intensive. For example, let the number of uncertainty factors be $J = 10$, each of which have $K_j = 4, j = 1, 2, \ldots, 10$ levels. Then there are

$N = \prod_{j=1}^{J} K_j = 4^{10} = 1\ 048\ 576$ different scenarios in the set $\mathcal{S}$ for which there are $J(J-1)/2 = 10(10-1)/2 = 45$ consistency values to aggregate, thus resulting in $[J(J-1)/2] \prod_{j=1}^{J} K_j = 47\ 185\ 920$ evaluations. Such a number of evaluations can render explicit enumeration methods impractical for a scenario planning exercise that is to be carried out in a timely manner, because generating each scenario and evaluating their overall consistencies separately may take from hours to days to compute.

In this section, we present an explicit enumeration procedure that generates the set $\mathcal{S}^I$ of the I most consistent scenarios efficiently. The efficiency of this procedure is based on the regular structure for the set of all scenarios, which is presented in Subsection 3.2.1. Utilizing this regular structure, the consistencies of all possible scenarios can be evaluated with an explicit enumeration algorithm presented in Subsection 3.2.2. Subsection 3.2.3 then explicates the efficiency of this evaluation procedure with simulation results.

3.2.1 Scenario matrix

To determine a regular structure for all scenarios in $\mathcal{S}$, we define a $N \times J$ matrix, the *scenario matrix* $\mathbf{S}$, the rows of which each represent a distinct combination of factor-specific levels. In the first $\prod_{t=2}^{J} K_t$ rows of this matrix, the level of the first uncertainty factor is fixed at $k_1 = 1$, while the levels of the rest of the factors span all their $\prod_{t=2}^{J} K_t$ possible combinations. Similarly, on the next $\prod_{t=2}^{J} K_t$ rows, the level of the first uncertainty factor is fixed at $k_1 = 2$, and the levels of the remaining factors again span all their possible combinations. The levels of these remaining factors $j = 2, 3, \ldots, J$ are specified such that given the fixed level $k_{j-1} \in \{1, 2, \ldots, K_{j-1}\}$ for uncertainty factor $j-1$, the level of factor j is fixed at some level $k_j \in \{1, 2, ..., K_j\}$ for $\prod_{t=j+1}^{J} K_t$ consecutive rows. In particular, the last uncertainty factor $j = J$ changes its level $k_J \in \{1, 2, \ldots, K_J\}$ on each row. In Equation (3.8), we illustrate the scenario matrix $\mathbf{S}$ for the subproblem of the electricity sales scenario planning problem, where the numbers of levels are $\mathbf{K} = \begin{bmatrix} 4 & 3 & 3 \end{bmatrix}$. For illustrative purposes, we show the row indices $n = 1, 2, \ldots, N, N = 4 \cdot 3 \cdot 3 = 36$ next to the matrix.

$$\text{When } \mathbf{K} = \begin{bmatrix} 4 & 3 & 3 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 3 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 3 & 2 \\ 1 & 3 & 3 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \\ 2 & 1 & 3 \\ 2 & 2 & 1 \\ 2 & 2 & 2 \\ 2 & 2 & 3 \\ 2 & 3 & 1 \\ 2 & 3 & 2 \\ 2 & 3 & 3 \\ 3 & 1 & 1 \\ 3 & 1 & 2 \\ 3 & 1 & 3 \\ 3 & 2 & 1 \\ 3 & 2 & 2 \\ 3 & 2 & 3 \\ 3 & 3 & 1 \\ 3 & 3 & 2 \\ 3 & 3 & 3 \\ 4 & 1 & 1 \\ 4 & 1 & 2 \\ 4 & 1 & 3 \\ 4 & 2 & 1 \\ 4 & 2 & 2 \\ 4 & 2 & 3 \\ 4 & 3 & 1 \\ 4 & 3 & 2 \\ 4 & 3 & 3 \end{bmatrix}_{\begin{smallmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \\ 17 \\ 18 \\ 19 \\ 20 \\ 21 \\ 22 \\ 23 \\ 24 \\ 25 \\ 26 \\ 27 \\ 28 \\ 29 \\ 30 \\ 31 \\ 32 \\ 33 \\ 34 \\ 35 \\ 36 \end{smallmatrix}}. \tag{3.8}$$

In general, scenario matrix $\mathbf{S}$ is defined as follows.

Definition 8. ***(Scenario matrix)*** *Let $\mathcal{S}$ be the set of all scenarios as in Definition 3. The scenario matrix $\mathbf{S}$ is the matrix with all possible scenarios $\mathbf{s} \in \mathcal{S}$ as its rows such that*

$$\mathbf{S} = [k_{nj}] \in \mathbb{N}^{N \times J},$$

where

$$k_{nj} = 1 + \left\lfloor \left(\prod_{t=j+1}^{J} K_t \right)^{-1} \left(\left(\left(\cdots \left(\left((n-1) \mod \prod_{t=2}^{J} K_t \right) \mod \prod_{t=3}^{J} K_t \right) \cdots \right) \mod \prod_{t=j-1}^{J} K_t \right) \mod \prod_{t=j}^{J} K_t \right) \right\rfloor. \tag{3.9}$$

In Equation (3.9), $\bmod$ is the modulo operation, that is, the remainder of the Euclidean division, e.g., $19 \mod 4 = 3$, because $q \cdot 4 + 3 = 19$, where $q = 4$. For example, the first and last column of this matrix are

$$k_{n1} = 1 + \left\lfloor \left(\prod_{t=2}^{J} K_t \right)^{-1} (n-1) \right\rfloor, \text{ and}$$

$$k_{nJ} = 1 + \Bigg(\Bigg(\Bigg(\cdots \Bigg(\Bigg((n-1) \mod \prod_{t=2}^{J} K_t \Bigg) \mod \prod_{t=3}^{J} K_t \Bigg) \cdots \Bigg) \mod \prod_{t=J-1}^{J} K_t \Bigg) \mod K_J \Bigg).$$

This structure is illustrated in Equation (3.10) below. In the equation, $\mathbf{1}_{\prod_{t=j+1}^{J} K_t}$ denotes a $\prod_{t=j+1}^{J} K_t \times 1$ column vector of ones.

$$\mathbf{S} = \begin{bmatrix}
 & \ddots & \mathbf{1}_{\prod_{t=j+1}^{J} K_t} & \begin{matrix} \iddots & 1 \\ \cdots & \vdots \\ \ddots & K_J \end{matrix} \\
\mathbf{1}_{\prod_{t=2}^{J} K_t} & \cdots & \vdots & \begin{matrix} \iddots & \vdots \\ \cdots & \vdots \\ \ddots & \vdots \end{matrix} \\
 & \ddots & K_j \cdot \mathbf{1}_{\prod_{t=j+1}^{J} K_t} & \begin{matrix} \iddots & 1 \\ \cdots & \vdots \\ \ddots & K_J \end{matrix} \\
 & \ddots & \mathbf{1}_{\prod_{t=j+1}^{J} K_t} & \begin{matrix} \iddots & 1 \\ \cdots & \vdots \\ \ddots & K_J \end{matrix} \\
2 \cdot \mathbf{1}_{\prod_{t=2}^{J} K_t} & \cdots & \vdots & \begin{matrix} \iddots & \vdots \\ \cdots & \vdots \\ \ddots & \vdots \end{matrix} \\
 & \ddots & K_j \cdot \mathbf{1}_{\prod_{t=j+1}^{J} K_t} & \begin{matrix} \iddots & 1 \\ \cdots & \vdots \\ \ddots & K_J \end{matrix} \\
 & \ddots & \cdots & \begin{matrix} \iddots & \vdots \\ \cdots & \vdots \\ \ddots & \vdots \end{matrix} \\
\vdots & \cdots & \cdots & \begin{matrix} \iddots & \vdots \\ \cdots & \vdots \\ \ddots & \vdots \end{matrix} \\
 & \ddots & \cdots & \begin{matrix} \iddots & \vdots \\ \cdots & \vdots \\ \ddots & \vdots \end{matrix} \\
 & \ddots & \mathbf{1}_{\prod_{t=j+1}^{J} K_t} & \begin{matrix} \iddots & 1 \\ \cdots & \vdots \\ \ddots & K_J \end{matrix} \\
K_1 \cdot \mathbf{1}_{\prod_{t=2}^{J} K_t} & \cdots & \vdots & \begin{matrix} \iddots & \vdots \\ \cdots & \vdots \\ \ddots & \vdots \end{matrix} \\
 & \ddots & K_j \cdot \mathbf{1}_{\prod_{t=j+1}^{J} K_t} & \begin{matrix} \iddots & 1 \\ \cdots & \vdots \\ \ddots & K_J \end{matrix}
\end{bmatrix} \tag{3.10}$$

The scenario matrix can be enumerated with the recursive Algorithm 1. In ScenarioMatrix, $\mathbf{S}_j$ denotes the jth column of $\mathbf{S}$. Scenario matrix $\mathbf{S}$ is generated by calling the algorithm from the top of the recursion stack, that is, $\mathbf{S} \leftarrow \text{ScenarioMatrix}(\mathbf{K}, 1)$.

Algorithm 1 Generating the scenario matrix $\mathbf{S}$.

Input: $\mathbf{K}, j$ ▷ Numbers of levels, current recursion depth
Output: $\mathbf{S}$ ▷ Scenario matrix in the current recursion

function ScenarioMatrix($\mathbf{K}, j$)
  if $j \neq J$ **then**
    $j \leftarrow j + 1$
    $\mathbf{S} \leftarrow \text{ScenarioMatrix}(\mathbf{K}, j)$ ▷ Recursion
  else
    $\mathbf{S} \leftarrow [\ ]$ ▷ Initialize $\mathbf{S}$
  end if
  For uncertainty factor j, repeat each level $\prod_{t=j+1}^{J} K_t$ times (once, if $j = J$).
  $\mathbf{k} \leftarrow [\ ]$ ▷ Initialize $\mathbf{k}$
  for $k_j = 1, 2, \ldots, K_j$ **do**
    $\mathbf{k}_j \leftarrow k_j \cdot \mathbf{1}^{\mathrm{T}}_{\prod_{t=j+1}^{J} K_t}$
    $\mathbf{k} \leftarrow \begin{bmatrix} \mathbf{k} & \mathbf{k}_j \end{bmatrix}$
  end for
  $\mathbf{S}_j \leftarrow \Big[\overbrace{\mathbf{k} \quad \mathbf{k} \quad \cdots \quad \mathbf{k}}^{\text{Repeat } \prod_{t=1}^{j-1} K_t \text{ times}} \Big]^{\mathrm{T}}$
  $\mathbf{S} \leftarrow \begin{bmatrix} \mathbf{S}_j & \mathbf{S} \end{bmatrix}$
  return S
end function

Let us denote the rows of scenario matrix $\mathbf{S}$ by $\mathbf{s}_n = \begin{bmatrix} k_{n1} & k_{n2} & \cdots & k_{nJ} \end{bmatrix}$, $n = 1, 2, \ldots, N$, and the vector of the overall consistencies of these rows (i.e., scenarios) by

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha(\mathbf{s}_1) \\ \alpha(\mathbf{s}_2) \\ \vdots \\ \alpha(\mathbf{s}_N) \end{bmatrix}. \tag{3.11}$$

From this vector, we define the vector $\mathbf{n}^I$ of scenario indices with the greatest overall consistencies as follows.

Definition 9. ***(Index vector of the I most consistent scenarios)*** *Let $\mathbf{S}$ be the scenario matrix with rows $\mathbf{s}_n, n = 1, 2, \ldots, N$. Let $\alpha(\mathbf{s}_n)$ be the overall consistency of scenario $\mathbf{s}_n$ as in Definition 6, and let $\mathcal{S}^I$ be the set of I most consistent scenarios as in Definition 7. The index vector of the I most consistent scenarios is*

$$\mathbf{n}^I = \begin{bmatrix} n_1^I \\ n_2^I \\ \vdots \\ n_I^I \end{bmatrix}, \quad \textit{such that} \tag{3.12}$$
$$\mathbf{s}_{n_i^I} \in \mathcal{S}^I, \forall i = 1, 2, \ldots, I, \textit{ and}$$
$$\alpha(\mathbf{s}_{n_1^I}) \geq \alpha(\mathbf{s}_{n_2^I}) \geq \cdots \geq \alpha(\mathbf{s}_{n_I^I}).$$

Using the index vector $\mathbf{n}^I$, we can define the matrix $\mathbf{S}^I$ of the I most consistent scenarios as follows.

Definition 10. ***(Matrix of the I most consistent scenarios)*** *Let $\mathcal{S}^I$ be the set of the I most consistent scenarios as in Definition 7. Let $\mathbf{S}$ be the scenario matrix and let $\mathbf{n}^I$ be the index vector of the I most consistent scenarios. The matrix of the I most consistent scenarios is*

$$\mathbf{S}^I = \begin{bmatrix} \mathbf{s}_{n_1^I}^I \\ \mathbf{s}_{n_2^I}^I \\ \vdots \\ \mathbf{s}_{n_I^I}^I \end{bmatrix} = [k_{n_i^I j}] \in \mathbb{N}^{I \times J}, \tag{3.13}$$

where $k_{n_i^I j}$ is the element of the scenario matrix $\mathbf{S}$ in the n_i^Ith row and jth column as in Equation (3.9).

We denote the rows of $\mathbf{S}^I$ by $\mathbf{s}_i^I = \mathbf{s}_{n_i^I}^I, i = 1, 2, \ldots, I$, and the vector of the overall consistencies of these rows by $\boldsymbol{\alpha}^I$.

3.2.2 Consistency Value Extraction algorithm

In keeping with Definition 10, to evaluate the set of I most consistent scenarios, it suffices to evaluate the indices $\mathbf{n}^I$ of the most consistent scenarios. Then, the most consistent scenarios can be retrieved according to the regular structure of the scenario matrix. However, to find the indices $\mathbf{n}^I$, it is necessary to explicitly evaluate the consistencies $\boldsymbol{\alpha}$ of all the rows of $\mathbf{S}$, of which there are exponentially many.

Nevertheless, although the number of possible scenarios (i.e., the number of rows in $\mathbf{S}$) grows exponentially, the number of unique consistency values in the consistency table $\mathbf{C}$ is much smaller. For each pair of uncertainty factors (j_1, j_2), there is a $K_{j_1} \times K_{j_2}$ matrix of consistency values corresponding to different combinations of levels on these factors, and hence the number of unique consistency values in the consistency table $\mathbf{C}$ is $\sum_{j_1=1}^{J-1} \sum_{j_2=j_1+1}^{J} K_{j_1} K_{j_2}$. For example, let the number of uncertainty factors be $J = 10$, each of which have $K_j = 4$ levels. Then there are $N = 1\ 048\ 576$ scenarios, but only $10 \cdot 9/2 \cdot 4^2 = 720$ elements in the consistency table that have to be extracted.

To extract the consistency values efficiently, we utilize the regular structure of the scenario matrix. This regularity is reflected by repetitions of these consistency values that follow the same pattern as the levels in each column of scenario matrix $\mathbf{S}$. We utilize the repeated level pairs by creating a number $J(J-1)/2$ of $N \times 1$ vectors $\mathbf{c}^{j_1 j_2} \in \mathcal{C}^N$ corresponding to each block $\mathbf{C}^{j_1 j_2}$ of the consistency table $\mathbf{C}$. In these vectors, each element $c_{j_1 j_2}(k_{j_1}, k_{j_2})$ of the block $\mathbf{C}^{j_1 j_2}$ is repeated in those rows $n \in \{1, 2, \ldots, N\}$ where factor j_1 has level k_{j_1} and factor j_2 has level k_{j_2}. Thus, the nth element of vector $\mathbf{c}^{j_1 j_2}$ is

$$\begin{aligned} c_n^{j_1 j_2} &= c^{j_1 j_2}(k_{j_1}, k_{j_2}), \quad \text{where} \\ k_{j_1} &= k_{n j_1}, \quad k_{j_2} = k_{n j_2} \end{aligned}$$

such that k_{nj_1} and k_{nj_2} are the levels of uncertainty factors j_1 and j_2 in the row n of the scenario matrix $\mathbf{S}$, as defined in Equation (3.9). The vector of overall consistencies of the rows $n = 1, 2, \ldots, N$ of scenario matrix $\mathbf{S}$ can be calculated as:

$$\boldsymbol{\alpha} = \frac{2}{J(J-1)} \sum_{j_1=1}^{J-1} \sum_{j_2=j_1+1}^{J} \mathbf{c}^{j_1 j_2}.$$

As an example, consider the subproblem of the electricity sales scenario planning example, where the first, fourth, and sixth uncertainty factors are considered. The scenario matrix of this problem with $\mathbf{K} = \begin{bmatrix} 4 & 3 & 3 \end{bmatrix}$ is presented in Equation (3.8) and the consistency table is presented in Table 3.5.

1. Consider the pair of uncertainty factors $(j_1, j_2) = (1, 2)$. Because the levels of these factors are repeated, there are scenarios (i.e., rows of $\mathbf{S}$) for which the consistency values from this factor pair are identical. For example, the first three scenarios all have same levels corresponding to factor level pair $(k_1, k_2) = (1, 1)$, and hence the consistency values are the same $c_n^{12} = c_n^{j_1 j_2} = c^{j_1 j_2}(k_{j_1}, k_{j_2}) = c^{12}(1, 1) = 3$ for $n = 1, 2, 3$.

2. Consider the factor pair $(j_1, j_2) = (1, 3)$. Due to level repetition, the pair of levels of these factors is $(k_1, k_3) = (1, 1)$ in rows $n = 1, 4, 7$, resulting in the corresponding consistency values being identical $c_n^{13} = c^{13}(1, 1) = 3$ for $n = 1, 4, 7$. As a similar example, the pair of levels $(k_1, k_3) = (3, 1)$ corresponds to rows $n = 19, 22, 25$, and hence $c_n^{13} = c^{13}(3, 1) = -1$ for these n.

3. Consider the factor pair $(j_1, j_2) = (2, 3)$. For example, the levels $(k_2, k_3) = (1, 1)$ correspond to rows $n = 1, 10, 19, 28$ and the levels $(k_2, k_3) = (3, 1)$ correspond to rows $n = 8, 17, 26, 35$, which is why we have $c_n^{23} = c^{23}(1, 1) = 3$ for $n = 1, 10, 19, 28$ and $c_n^{23} = c^{23}(3, 2) = -2$ for $n = 7, 16, 25, 34$.

By extending this approach to the complete consistency value extraction problem, the overall consistencies of all rows $n = 1, 2, \ldots, N$ of the scenario matrix presented in Equation (3.8) can be calculated as in Equation (3.14).

$$\boldsymbol{\alpha} = \frac{1}{3}\left(\mathbf{c}^{12} + \mathbf{c}^{13} + \mathbf{c}^{23}\right) = \frac{1}{3}\left(\begin{bmatrix} 3\\3\\3\\2\\2\\2\\-1\\-1\\-1\\2\\2\\2\\2\\2\\2\\2\\2\\-3\\-3\\-3\\1\\1\\1\\2\\2\\2\\-3\\-3\\-3\\1\\1\\1\\3\\3\\3 \end{bmatrix} + \begin{bmatrix} 3\\1\\1\\3\\1\\1\\3\\1\\1\\1\\1\\1\\1\\1\\1\\1\\-1\\2\\3\\-1\\2\\3\\-1\\2\\3\\-3\\2\\3\\-3\\2\\3\\-3\\2\\3 \end{bmatrix} + \begin{bmatrix} 3\\2\\-2\\2\\3\\1\\-2\\2\\2\\3\\2\\-2\\2\\3\\1\\-2\\2\\2\\3\\2\\-2\\2\\3\\1\\-2\\2\\2\\3\\2\\-2\\2\\3\\1\\-2\\2\\2 \end{bmatrix}\right) = \begin{bmatrix} 3\\2\\0.67\\2.33\\2\\1.33\\0\\0.67\\0.67\\2\\1.67\\0.33\\1.67\\2\\1.33\\0.33\\1.67\\1.67\\-0.33\\0.33\\-0.67\\0.67\\2\\1.67\\-0.33\\2\\2.33\\-1\\0.33\\-0.67\\0\\2\\1.67\\-0.67\\2.33\\2.67 \end{bmatrix} \tag{3.14}$$

The overall consistencies of the rows of scenario matrix $\mathbf{S}$ can be evaluated with the recursive CVX algorithm, presented in Algorithm 2. In each recursion j of this algorithm, the pairwise consistencies are repeated for each factor level pair corresponding to the factor of the current recursion depth j and all preceding uncertainty factors $t = 1, 2, \ldots, j-1$. The consistencies are then repeated in a structure which corresponds to the repetitions of levels in rows 10 and 13 of Algorithm 1, corresponding to recursion depths j and $t = 1, 2, \ldots, j-1$.

Algorithm 2 The Consistency Value Extraction (CVX) algorithm.

Input: $\mathbf{K}, \mathbf{C}, j$ ▷ Numbers of levels, consistency table, current recursion depth

Output: $\boldsymbol{\alpha}$ ▷ Overall consistencies in the current recursion

1: **function** CVX($\mathbf{K}, \mathbf{C}, j$)
2: **if** $j \neq J$ **then**
3: $\boldsymbol{\alpha} \leftarrow \text{CVX}(\mathbf{K}, \mathbf{C}, j+1)$ ▷ Recursion
4: **else**
5: $\boldsymbol{\alpha} \leftarrow \prod_{t=1}^{J} K_t \times 1$ vector of zeros ▷ Initialize A
6: **end if**
For factor j, repeat consistency values with this factor as column factor in consistency table in a similar structure as levels in $\mathbf{S}$. Then, add these values to $\boldsymbol{\alpha}$.
7: **for** $t = 1, 2, \ldots, j-1$ **do**
8: $\mathbf{c}_{kk}^{tj} \leftarrow [\ ]$ ▷ Initialize $\mathbf{c}_{kk}^{tj}$
9: $\mathbf{c}_{k}^{tj} \leftarrow [\ ]$ ▷ Initialize $\mathbf{c}_{k}^{tj}$
10: **for** $k_t = 1, 2, \ldots, K_t$ **do**
11: **for** $k_j = 1, 2, \ldots, K_j$ **do**
12: $\mathbf{c}_{kk}^{tj} \leftarrow \left[\mathbf{c}_{kk}^{tj} \quad c^{tj}(k_t, k_j) \cdot \mathbf{1}^{\mathrm{T}}_{\prod_{u=j+1}^{J} K_u}\right]$ ▷ Repeat $c^{tj}(k_t, k_j)$
13: **end for**
14: $\mathbf{c}_{k}^{tj} \leftarrow \left[\mathbf{c}_{k}^{tj} \quad \overbrace{\mathbf{c}_{kk}^{tj} \quad \mathbf{c}_{kk}^{tj} \quad \cdots \quad \mathbf{c}_{kk}^{tj}}^{\text{Repeat } \prod_{u=t+1}^{j-1} K_u \text{ times}}\right]$
15: **end for**
16: $\mathbf{c}^{tj} \leftarrow \left[\overbrace{\mathbf{c}_{k}^{tj} \quad \mathbf{c}_{k}^{tj} \quad \cdots \quad \mathbf{c}_{k}^{tj}}^{\text{Repeat } \prod_{u=1}^{t-1} K_u \text{ times}}\right]^{\mathrm{T}}$
17: $\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} + \mathbf{c}^{tj}$ ▷ Accumulate overall consistency
18: **end for**
19: **return** $\boldsymbol{\alpha}$
20: **end function**

Once the consistencies $\boldsymbol{\alpha}$ of the rows of $\mathbf{S}$ have been evaluated with the CVX algorithm, the index vector $\mathbf{n}^I$ of I most consistent scenarios [see Equation (3.12)] can be evaluated by (i) partitioning the vector $\boldsymbol{\alpha}$ with respect to the Ith greatest element and then (ii) sorting the list of the I greatest values. The indices of the I greatest values of vector $\boldsymbol{\alpha}$ can be obtained, for example, with the introselect algorithm [Musser (1997)] and the I greatest elements of $\boldsymbol{\alpha}$ can be sorted with, e.g., the quicksort algorithm [Cormen et al. (2017)]. A procedure to evaluate the consistencies $\boldsymbol{\alpha}$ by calling CVX and then constructing $\mathbf{S}^I$ from $\mathbf{n}^I$ is presented in Algorithm 3.

Algorithm 3 Evaluating scenarios $\mathbf{S}^I$ and their overall consistencies.

Input: $\mathbf{K}, \mathbf{C}, I$ ▷ Levels, consistency table, number of consistent scenarios
Output: $\mathbf{S}^I, \boldsymbol{\alpha}^I$ ▷ The most consistent scenarios, their overall consistencies

function ConsistentScenarios($\mathbf{K}, \mathbf{C}, I$)
  $\boldsymbol{\alpha} \leftarrow \mathrm{CVX}(\mathbf{K}, \mathbf{C}, 1)$ ▷ Call CVX
  $\boldsymbol{\alpha}^I,\ \mathbf{n}^I \leftarrow$ PartitionDescending($\boldsymbol{\alpha}, I$) ▷ Get $\boldsymbol{\alpha}^I$ and $\mathbf{n}^I$ unsorted
  $\boldsymbol{\alpha}^I,\ \mathbf{i} \leftarrow$ SortDescending($\boldsymbol{\alpha}^I$) ▷ Sort $\boldsymbol{\alpha}^I$
  $\mathbf{n}^I \leftarrow \mathbf{n}^I_{\mathbf{i}}$ ▷ Sort $\mathbf{n}^I$
  $\boldsymbol{\alpha}^I \leftarrow \boldsymbol{\alpha}^I / [J(J-1)/2]$ ▷ Scale $\boldsymbol{\alpha}^I$
  $\mathbf{S}^I \leftarrow [\ ]$
  for n **in** $\mathbf{n}^I$ **do** ▷ Construct $\mathbf{S}^I$
    $\mathbf{s}_i \leftarrow [\ ]$
    for $j = 1, 2, \ldots, J$ **do**
      $\mathbf{s}_i \leftarrow \begin{bmatrix} \mathbf{s}_i & k_{nj} \end{bmatrix}$ ▷ Evaluate k_{nj} using (3.9)
    end for
    $\mathbf{S}^I \leftarrow \begin{bmatrix} (\mathbf{S}^I)^{\mathrm{T}} & \mathbf{s}_i^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$
  end for
  return $\mathbf{S}^I, \boldsymbol{\alpha}^I$
end function

The efficiency of Algorithm 2 is based on the ability of the processing units of modern computers to simultaneously perform multiple elementary operations, such as insertion and addition. However, if the number of possible scenarios N is large, vector $\boldsymbol{\alpha}$ and the vectors $\mathbf{c}^{j_1 j_2}$ can overload the memory of the computer. In such a situation, the efficiency of the computation is reduced, and moreover, the size of vectors $\boldsymbol{\alpha}$ can be too large to be stored in rapid access memory for very large J, thus requiring hard drive operations for processing. For example, if $J = 15, K_j = 4 \forall j = 1, 2, \ldots, 15$ and if 16-bit integers store the elements of vector $\boldsymbol{\alpha}$ and vectors $\mathbf{c}^{j_1 j_2}$, the sizes of $\boldsymbol{\alpha}$ and $\mathbf{c}^{j_1 j_2}$'s are $2 \cdot 4^{15} \approx 2 \cdot 10^9$ bytes, i.e., roughly 2 GB each.

Due to such memory issues, evaluating the consistency values by calling CVX for the full enumeration problem can be inefficient. Instead, efficiency can be improved by solving smaller partial problems and calling the algorithm several times, the number of calls depending on the size of the enumeration problem at hand. In particular, we define a computational parameter η, which defines the number of uncertainty factors on which calling CVX is parallelized. More specifically, (i) if $\eta = 0$, the complete problem is evaluated at once, (ii) if $\eta = 1$, CVX is called K_1 times, once for each level of the first uncertainty factor $k_1 = 1, 2, \ldots, K_1$, and (iii) if $\eta = 2$, CVX is called $K_1 \cdot K_2$ times, once for each combination of pairs of levels (k_1, k_2). In general, for any $\eta \in \{0, 1, \ldots, J-1\}$, CVX is called once for each row of an auxiliary scenario matrix $\mathbf{S}^\circ$ corresponding to the numbers of factors $\mathbf{K}^\circ = \begin{bmatrix} K_1 & K_2 & \cdots & K_\eta \end{bmatrix}$. During these calls, the effective number of uncertainty factors is $J - \eta$, because the η first factors have fixed levels, and hence these factors do not increase the scales of the partial problems.

While solving these partial enumeration problems, not all of the exponentially many consistency values have to be stored in the memory of the computer, because only I ($<< N$) scenarios with the greatest consistency values are of interest. Low consistency values can be discarded sequentially, and only the greatest consistency values and the corresponding scenario indices must be stored. With this approach, we present a parallelized scheme to calculate the most consistent scenarios, presented below in Algorithm 4. In this Algorithm 4, the outer for loop is parallelizable, because the iterations are independent of each other, and hence multiple cores of modern computers can be utilized to evaluate this algorithm. Algorithm 4 can be used as the general scheme to call CVX, because this calling procedure reduces to Algorithm 3 when $\eta = 0$.

We explain some operations in Algorithm 4 in more detail. Line 3 creates the auxiliary scenario matrix $\mathbf{S}^\circ$ to access those rows and columns of $\mathbf{C}$ that are ignored when evaluating consistencies such that the first η factors have constant levels $k_{n'j}, j = 1, 2, \ldots, \eta$. The constant levels are accessed by calling CVX in line 16 such that the rows and columns of $\mathbf{C}$ that are ignored are deleted from an auxiliary consistency table $\mathbf{C}'$ in lines 8-15. In line 18, the indices of $\mathbf{n}'$ are shifted so that they correspond to the scenario matrix $\mathbf{S}$ of the full problem with the numbers of factor levels $\mathbf{K}$, not the subproblem with the numbers of factor levels $\mathbf{K}'$. In line 22, the resulting indices of the partition correspond to the vector $\boldsymbol{\alpha}^{\eta I}$, so the indices of the most consistent scenarios need to be retrieved from $\mathbf{n}^{\eta I}$ in line 23, where the subscript $\mathbf{n}''$ denotes retrieving the elements of $\mathbf{n}^{\eta I}$ in positions $\mathbf{n}''$. Moreover, line 24 sorts

the I greatest consistencies in a descending order, and then line 25 sets the same order on the vector of indices $\mathbf{n}^I$ of rows of scenario matrix $\mathbf{S}$.

Algorithm 4 Calling procedure of CVX algorithm.

Input: $\mathbf{K}, \mathbf{C}, \eta, I$ ▷ η is number of factors to parallelize
Output: $\mathbf{S}^I, \boldsymbol{\alpha}^I$ ▷ I most consistent scenarios and their overall consistencies

1: **function** CVXScheme($\mathbf{K}, \mathbf{C}, \eta, I$)
2: $\quad\mathbf{K}^\circ \leftarrow \begin{bmatrix} K_1 & K_2 & \cdots & K_\eta \end{bmatrix}$
3: $\quad\mathbf{S}^\circ \leftarrow$ ScenarioMatrix($\mathbf{K}^\circ, 1$)
4: $\quad\boldsymbol{\alpha}^{\eta I}, \mathbf{n}^{\eta I} \leftarrow [\ ], [\ ]$ ▷ Initialize $\boldsymbol{\alpha}^{\eta I}$ and $\mathbf{n}^{\eta I}$
5: $\quad\mathbf{K}' \leftarrow \begin{bmatrix} 1 & \cdots & 1 & K_{\eta+1} & \cdots & K_J \end{bmatrix}$
6: $\quad$**for** $n' = 1, 2, \ldots, \prod_{j=1}^{\eta} K_j$ **do** ▷ If $\eta = 0$, $\prod_{j=1}^{\eta} K_j = 1$
7: $\quad\quad\mathbf{C}' \leftarrow \mathbf{C}$ ▷ Copy $\mathbf{C}$
8: $\quad\quad$**for** $j = 1, 2, \ldots, \eta$ **do** ▷ Keep only 1 level in factors $1, 2, \ldots, \eta$
9: $\quad\quad\quad j_1 \leftarrow k^\circ_{n'j} + (j-1)$ ▷ $k^\circ_{nj'}$ is an element of $\mathbf{S}^\circ$
10: $\quad\quad\quad j_2 \leftarrow K_j + (j-1)$
11: $\quad\quad\quad\mathbf{C}'_{j:j_1-1,:}, \mathbf{C}'_{j_1+1:j_2,:} \leftarrow [\ ], [\ ]$ ▷ Delete rows
12: $\quad\quad\quad$**if** $j > 1$ **then**
13: $\quad\quad\quad\quad\mathbf{C}'_{:,j-1:j_1-2}, \mathbf{C}'_{:,j_1:j_2-1,} \leftarrow [\ ], [\ ]$ ▷ Delete columns
14: $\quad\quad\quad$**end if**
15: $\quad\quad$**end for**
16: $\quad\quad\boldsymbol{\alpha}' \leftarrow \mathrm{CVX}(\mathbf{K}', \mathbf{C}', 1)$ ▷ Get consistencies
17: $\quad\quad\boldsymbol{\alpha}',\ \mathbf{n}' \leftarrow$ PartitionDescending($\boldsymbol{\alpha}', I$) ▷ Partition $\boldsymbol{\alpha}'$
18: $\quad\quad\mathbf{n}' \leftarrow \mathbf{n}' + (n'-1)\prod_{j=\eta+1}^{J} K_j$ ▷ Shift $\mathbf{n}'$
19: $\quad\quad\boldsymbol{\alpha}^{\eta I} \leftarrow \begin{bmatrix} \left(\boldsymbol{\alpha}^{\eta I}\right)^{\mathrm{T}} & (\boldsymbol{\alpha}')^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ ▷ Store greatest overall consistencies
20: $\quad\quad\mathbf{n}^{\eta I} \leftarrow \begin{bmatrix} \left(\mathbf{n}^{\eta I}\right)^{\mathrm{T}} & (\mathbf{n}')^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ ▷ Store indices of greatest consistencies
21: $\quad$**end for**
22: $\quad\boldsymbol{\alpha}^I,\ \mathbf{n}'' \leftarrow$ PartitionDescending($\boldsymbol{\alpha}^{\eta I}, I$) ▷ Get $\boldsymbol{\alpha}^I$ unsorted
23: $\quad\mathbf{n}^I \leftarrow \mathbf{n}^{\eta I}_{\mathbf{n}''}$ ▷ Get $\mathbf{n}^I$ unsorted
24: $\quad\boldsymbol{\alpha}^I,\ \mathbf{i} \leftarrow$ SortDescending($\boldsymbol{\alpha}^I$) ▷ Sort $\boldsymbol{\alpha}^I$
25: $\quad\mathbf{n}^I \leftarrow \mathbf{n}^I_{\mathbf{i}}$ ▷ Sort $\mathbf{n}^I$
26: $\quad\boldsymbol{\alpha}^I \leftarrow \boldsymbol{\alpha}^I / [J(J-1)/2]$ ▷ Scale $\boldsymbol{\alpha}^I$
27: $\quad$**for** n **in** $\mathbf{n}^I$ **do** ▷ Construct $\mathbf{S}^I$
28: $\quad\quad\mathbf{s}_i \leftarrow [\ ]$
29: $\quad\quad$**for** $j = 1, 2, \ldots, J$ **do**
30: $\quad\quad\quad\mathbf{s}_i \leftarrow \begin{bmatrix} \mathbf{s}_i & k_{nj} \end{bmatrix}$ ▷ Evaluate k_{nj} using (3.9)
31: $\quad\quad$**end for**
32: $\quad\quad\mathbf{S}^I \leftarrow \begin{bmatrix} (\mathbf{S}^I)^{\mathrm{T}} & \mathbf{s}_i^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$
33: $\quad$**end for**
34: $\quad$**return** $\mathbf{S}^I, \boldsymbol{\alpha}^I$
35: **end function**

3.2.3 Efficiency tests

Two kinds of efficiency tests were run for the algorithms presented in the previous subsections. First, we compare the efficiency of the presented algorithms to the efficiency of a naive evaluation strategy. Second, we analyse the effect of the computational parameter η on computation time. In all tests, the numbers of uncertainty factor levels were kept constant $K_j = 4$ for all $j = 1, 2, \ldots, J$, and the consistency tables $\mathbf{C}$ were randomly generated. In the second set of tests, the number of consistent scenarios was $I = \min(0.1 \cdot N, 2000)$. The tests were implemented in Matlab and run on a standard laptop (dual core, 2.4 GHz, 8 GB memory), and both cores were utilized in parallel in the second set of tests.

Figure 3.1 demonstrates the efficiency of the CVX algorithm by comparing the efficiency of Algorithm 3 to the efficiency of the naive approach, presented in Appendix A. The naive approach utilizes nested for loops to generate the set $\mathcal{S}$ and then evaluates the overall consistencies of all scenarios generated in this way. Both evaluation strategies have exponential time complexity with respect to the number of factors J, which can be observed from the relatively straight lines in the semi-logarithmic plots of their evaluation times. However, the efficient approach has a lower slope, thus resulting in much lower evaluation times. For example, the evaluation time with $J = 10$ was 0.45 s for the efficient algorithm and 41.57 s for the naive algorithm. By extrapolating the curve of evaluation times of the naive algorithm, we estimate the evaluation time for $J = 15$ to be approximately 10^5s (i.e., roughly one day).

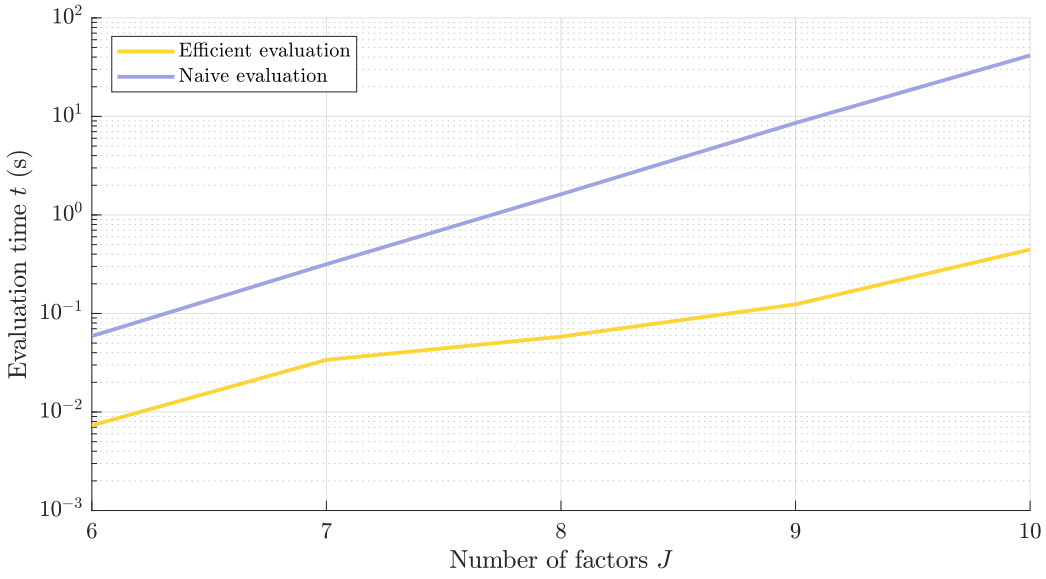


Figure 3.1: Average evaluation times from 10 test runs of algorithms 3 and 7, referred to as efficient and naive approach, respectively.

Figure 3.2 presents the effect of varying the computational parameter η for small to moderate numbers of uncertainty factors $J = 6, 7, 8, 9, 10$. For nearly all problem sizes, an unparallelized setup with $\eta = 0$ yields the fastest computation time. However, when the number of uncertainty factors becomes larger, the optimal value of η increases. For instance, for $J = 9$, the optimal value of η is 1, and for $J = 10$, this optimal value is 2.

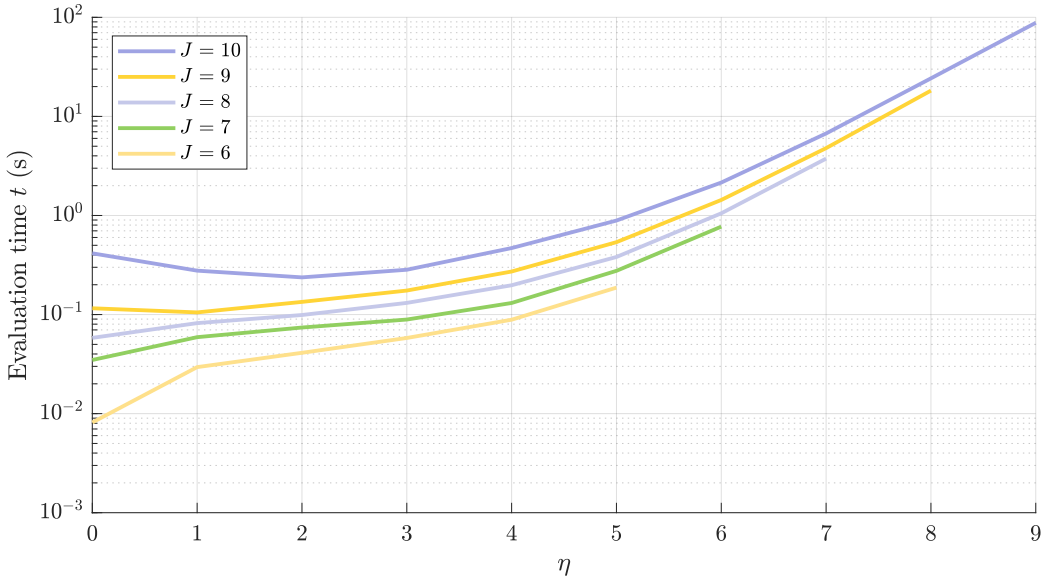


Figure 3.2: Average evaluation times from 10 test runs of Algorithm 4 with small to moderate numbers of uncertainty factors J.

Figure 3.3 presents similar tests for large numbers of uncertainty factors $J = 11, 12, 13, 14, 15$. For larger problem sizes, paralellizing the runs of CVX becomes much more important, e.g., reducing the computation time from 177.42s to 49.34s with $J = 14$, and making the evaluation possible with $J = 15$, for which full enumeration with Algorithm 3 was not possible due to memory overload. In Figure 3.3, as well as in Figure 3.2, the optimal effective number of uncertainty factors with which to call CVX was $J - \eta^* = 8$ for all J from 8 to 15, where η^* denotes the value of η with which the evaluation time was the shortest. Any exceeding number of factors is parallelized in the optimal evaluation strategy. With more factors, processing excessively large vectors $\mathbf{c}^{j_1 j_2}$ becomes more computationally expensive than to compute multiple smaller partial enumeration problems. We note, however, that this optimality is strongly related to the type of hardware used in the computation.

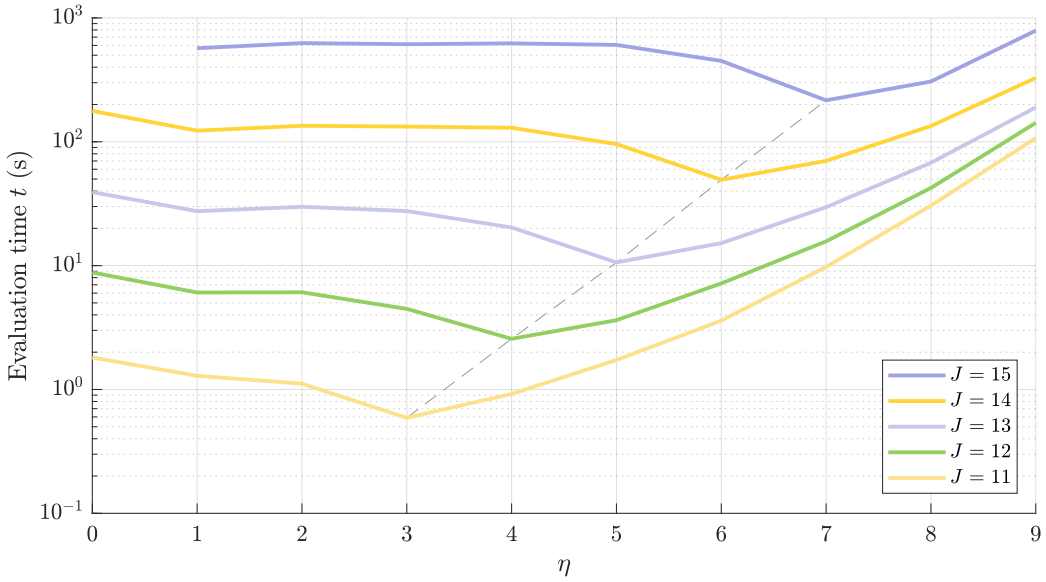


Figure 3.3: Evaluation times from single test runs of Algorithm 4 with large numbers of uncertainty factors J. The dashed line connects the values η^* with the smallest computation time for each J.

Our test results demostrate that the computation of the set $\mathcal{S}^I$ of I most consistent scenarios for any consistency value extraction problem of a reasonable size can be quickly solved with Algorithm 4. With $K_j = 4, j = 1, 2, \ldots, J = 15$, the consistency table already requires $4^2 \cdot 15(15-1)/2 = 1680$ consistency assessments. For any greater J, the evaluation of the pairwise consistencies becomes very arduous. Nevertheless, even for very massive consistency value extraction problems with, e.g., $K_j = 4, J = 20$, Algorithm 4 can be used to evaluate the set of I most consistent scenarios, although in such cases the computation might require computation time from hours to days.

Chapter 4

Visualizing dissimilarities of scenarios

In this chapter, we present a method for visualizing the set of the I most consistent scenarios in a two-dimensional plane. From this visualization, a set of M mutually dissimilar consistent scenarios can be efficiently identified (e.g., $M = 3, 4, 5, 6$). In particular, Section 4.1 presents the Multiple Correspondence Analysis dimensionality reduction algorithm for projecting the high-dimensional cloud of the most consistent scenarios onto a lower-dimensional space that most preserves the structure of the scenario cloud. Section 4.2 presents the Scenario Map, which is a visualization method for exploring the I consistent scenarios and their dissimilarity to assist in the selection of the final set of M internally consistent and mutually dissimilar scenarios.

4.1 Multiple Correspondence Analysis

In this section, we present Multiple Correspondence Analysis (MCA), which is used to visualize the set $\mathcal{S}^I$ of the most consistent scenarios. MCA is a dimensionality reduction method, in which a categorical data set is explored by simultaneously studying the data vectors and the levels of this set with the goal of seeking structural interrelations between category levels. In this thesis, however, we use MCA to illustrate the dissimilarities of scenarios (cf. data vectors) and give less weight to the consideration of the closeness of the factor levels. Moreover, although we present the theory of MCA as in most MCA textbooks, the notation and terminology is aligned with the scenario analytic formulations presented in the preceding chapter.

We begin in Subsection 4.1.1 by presenting the *complete disjunctive table*, which is an indicator matrix for studying the structural interrelations of a categorical data set. In Subsection 4.1.2, we present a distance metric that is suitable for such data. Finally, in Subsection 4.1.3, we (i) transform the complete disjunctive table to a form in which the Euclidean distances between rows correspond to the previously defined distance metric, and (ii) apply Singular Value Decomposition to optimally project the transformed table onto a lower-dimensional space, such that maximal amount of variability of the original table is retained.

4.1.1 Complete disjunctive table

For the purposes of applying MCA, we present the matrix $\mathbf{S}^I$ of the I most consistent scenarios in the table form shown in Figure 4.1 [Husson et al. (2017)]. Each row $i = 1, 2, \ldots, I$ of this table corresponds to a consistent scenario $\mathbf{s}_i^I = \begin{bmatrix} k_{i1} & k_{i2} & \cdots & k_{iJ} \end{bmatrix} \in \mathcal{S}^I$. We refer to these scenarios by their row indices i and vectors $\mathbf{s}_i^I$ of factor-specific levels interchangeably.

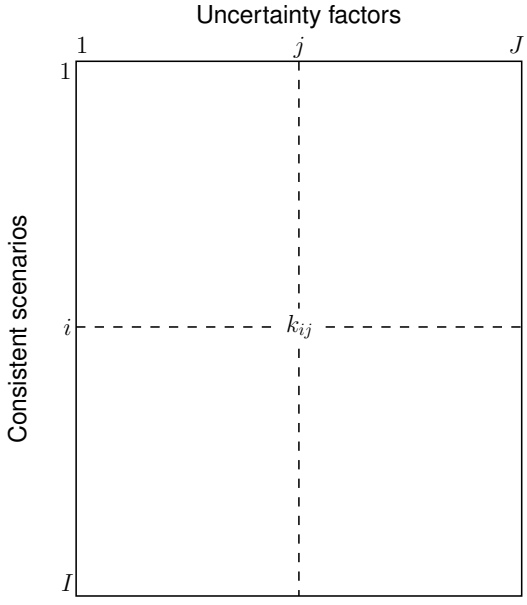


Figure 4.1: Table representation of the matrix $\mathbf{S}^I$ of the I most consistent scenarios.

From table $\mathbf{S}^I$, we can construct an indicator matrix $\mathbf{X}$ with elements $x_{ik} \in \{0, 1\}$. The rows $i = 1, 2, \ldots, I$ of this matrix correspond to the consistent

scenarios, and the columns $k = 1, 2, \ldots, K, K = \sum_{j=1}^{J} K_j$ to the levels k_j of the categorical variables representing all uncertainty factors $j = 1, 2, \ldots, J$. If the level of factor j of scenario i is k_j, then the element x_{ik} corresponding to this factor-specific level is 1, otherwise this element is 0. For brevity, the columns k are referred to as *categories*, and if the binary element x_{ik} of the indicator matrix $\mathbf{X}$ is equal to one, we say that scenario i *belongs to category* k.

In each scenario i, each uncertainty factor must attain exactly one level. As a result, each row i of the indicator matrix $\mathbf{X}$ has exactly one non-zero entry in the columns corresponding to the jth uncertainty factor. Hence, for each row there are J constraints, one for each factor j to impose that the columns corresponding to these factors each have exactly one non-zero entry. Hence, the rows $\mathbf{x}_i$ of $\mathbf{X}$ lie in a $K - J$ dimensional space, because they have K elements and J constraints, each of which reduces the dimensionality by one.

Matrix $\mathbf{X}$ is called a *complete disjunctive table* (CDT), and this table is more formally defined in Definitions 11 and 12.

Definition 11. ***(Disjunctive scenario vector)*** *Let* $\mathbf{s}_i^I = \begin{bmatrix} k_{i1} & k_{i2} & \cdots & k_{iJ} \end{bmatrix}$ *be a row of the matrix of the* I *most consistent scenarios* $\mathbf{S}^I$ *as in Definition 10. The disjunctive scenario vector* $\mathbf{x}_i$ *that corresponds to scenario* $\mathbf{s}_i^I$ *is*

$$\begin{aligned} \boldsymbol{x}_i &= \begin{bmatrix} x_{i1} & x_{i1} & \cdots & x_{iK} \end{bmatrix}, \quad \textit{where} \\ x_{ik} &= 1 \quad \textit{for all} \quad k = k_{ij} + \sum_{t=1}^{j-1} K_t, j = 1, 2, \ldots, J \quad \textit{and} \\ x_{ik} &= 0, \quad \textit{otherwise.} \end{aligned} \tag{4.1}$$

Definition 12. ***(Complete disjunctive table)*** *Let* $\mathbf{x}_i, i = 1, 2, \ldots, I$ *be the disjunctive scenario vectors of the rows of matrix* $\mathbf{S}^I$*. The complete disjunctive table* $\mathbf{X}$ *of consistent scenarios is*

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_I \end{bmatrix}. \tag{4.2}$$

We consider some basic properties of the CDT, starting from the sums along the rows (row margins). Because there are exactly J non-zero entries in each row of the CDT, the row sums are all equal to J. Moreover, because the number of scenarios in the data matrix is I, the sum of all row margins in the CDT is IJ. Considering the columns of the CDT, we observe that the

sum of the entries in the kth column (kth column margin) equals the number of scenarios I_k that belong to category k. On the other hand, the kth column sum is the number of scenarios I times the proportion of scenarios in the kth category $p_k = I_k/I$. Then, if we sum the K_j terms of the column margins corresponding to the jth categorical variable, we get I. Because there are in total J categorical variables, the sum of all column margins is also IJ. The properties of CDT table are illustrated in Table 4.2.

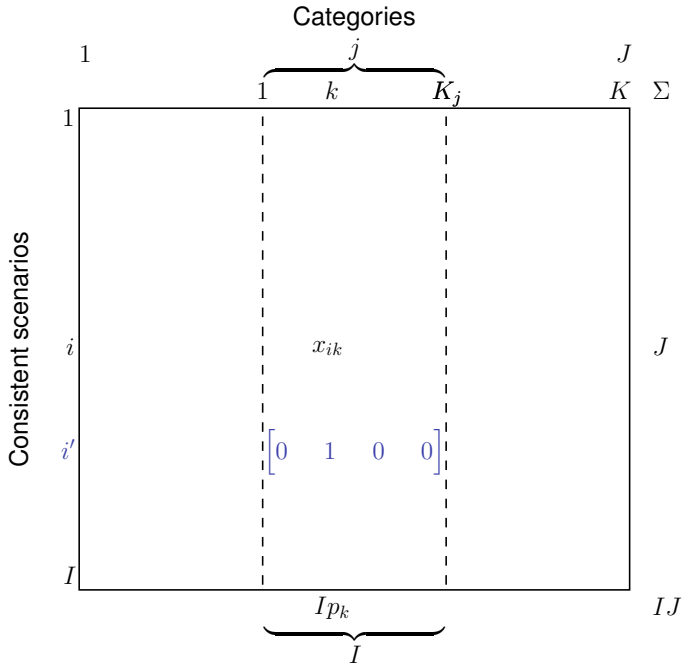


Figure 4.2: Complete disjunctive table (CDT) representation of the matrix $\mathbf{S}^I$ of the I most consistent scenarios.

4.1.2 Distance between two consistent scenarios

To study the similarity between two scenarios $i_1, i_2 = 1, 2, \ldots, I$, we define a distance metric $d(\mathbf{x}_{i_1}, \mathbf{x}_{i_2})$ for the set $\mathcal{X}$ of rows $\mathbf{x}_i = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{iK} \end{bmatrix}$ of the indicator matrix $\mathbf{X}$. This metric should satisfy the conditions presented in Table 4.1. These conditions are influenced by those presented in Husson et al. (2017), yet defined and formalized specifically for our scenario planning context.

Table 4.1: Three conditions for the distance metric of scenarios of $\mathbf{S}^I$ as represented by rows of $\mathbf{X}$.

	Explanation	Formal definition
(i)	If the categories of two scenarios $\mathbf{s}^I_{i_1}$ and $\mathbf{s}^I_{i_2}$ are the same, the distance is zero.	$\mathbf{x}_{i_1} = \mathbf{x}_{i_2} \Leftrightarrow d(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}) = 0$
(ii)	If scenario $\mathbf{s}^I_{i_1}$ shares more categories with scenario $\mathbf{s}^I_{i_2}$ than with scenario $\mathbf{s}^I_{i_3}$, then scenarios $\mathbf{s}^I_{i_1}$ and $\mathbf{s}^I_{i_2}$ are closer together than scenarios $\mathbf{s}^I_{i_1}$ and $\mathbf{s}^I_{i_3}$.	Let $\exists j$ s.t. $k_{i_1 j} = k_{i_2 j}, k_{i_1 j} \neq k_{i_3 j}$, but $\nexists j'$ s.t. $k_{i_1 j'} \neq k_{i_2 j'}, k_{i_1 j'} = k_{i_3 j'}$. Then $d(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}) \leq d(\mathbf{x}_{i_1}, \mathbf{x}_{i_3})$.
(iii)	If scenario $\mathbf{s}^I_{i_1}$ shares all but one category with scenarios $\mathbf{s}^I_{i_2}$ and $\mathbf{s}^I_{i_3}$, and scenarios $\mathbf{s}^I_{i_1}$ and $\mathbf{s}^I_{i_2}$ share a rare category that scenarios $\mathbf{s}^I_{i_1}$ and $\mathbf{s}^I_{i_3}$ do not share, this uniqueness will bring scenarios $\mathbf{s}^I_{i_1}$ and $\mathbf{s}^I_{i_2}$ closer together than scenarios $\mathbf{s}^I_{i_1}$ and $\mathbf{s}^I_{i_3}$.	Let $\exists! j$ s.t. $k_{i_1 j} = k_{i_2 j}, k_{i_1 j} \neq k_{i_3 j}$ and $\exists! j'$ s.t. $k_{i_1 j'} \neq k_{i_2 j'}, k_{i_1 j'} = k_{i_3 j'}$ and let k correspond to category $k_{i_1 j} = k_{i_2 j}$ of j and let k' correspond to category $k_{i_1 j'} = k_{i_3 j'}$ of j' according to (4.1). Then $p_k \leq p_{k'} \Leftrightarrow d(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}) \leq d(\mathbf{x}_{i_1}, \mathbf{x}_{i_3})$.

A metric that satisfies all three conditions is given in Definition 13.

Definition 13. ***(Distance of two disjunctive scenario vectors)*** *Let $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$ be the disjunctive scenario vectors corresponding to consistent scenarios $\mathbf{s}^I_{i_1}$ and $\mathbf{s}^I_{i_2}$. The distance between these two scenarios is*

$$d(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}) = \frac{1}{\sqrt{IJ}} \sqrt{\sum_{k=1}^{K} \frac{1}{p_k} \left(x_{i_1 k} - x_{i_2 k}\right)^2}, \tag{4.3}$$

where

$$p_k = \frac{1}{I} \sum_{i=1}^{I} x_{ik}.$$

The first condition in Table 4.1 is satisfied, because identical vectors $\mathbf{x}_{i_1} = \mathbf{x}_{i_2}$ have a distance equal to zero, which would be true for many other types of metrics as well. However, in this metric, summing squared differences in the binary elements x_{ik} guarantees that it matters only whether the categorical variables k_{ij} of two scenarios i_1 and i_2 differ or do not differ. As a result, the condition (ii) is also satisfied, because large differences in k_{ij}'s due to magnitudes of the numerical values of the categorical variables do not affect

the calculation of the distance according to Equation (4.3). Moreover, scaling the squared differences by weights that are inversely proportional to the proportion p_k of entries in each category k increases the sensitivity of the metric with respect to differences in rare categories, thus satisfying condition (iii).[1] Equation (4.3) also has a scaling constant $1/\sqrt{IJ}$, which is not related to the conditions in Table 4.1. Instead, this scaling constant enables many other types of useful properties in MCA [Greenacre and Blasius (2006)], such as symmetric transition relations that link the point cloud of the rows to the point cloud of the columns.

The goal of MCA is to find a projection of a point cloud (here, the set of rows $\mathbf{x}_i$ of the indicator matrix $\mathbf{X}$) from $K-J$ dimensions onto a lower-dimensional space, such that the diversity of the point cloud is best preserved. In MCA, the diversity of the point cloud is expressed with the (rotational) inertia of the cloud, which is the mass weighted sum of squared distances from the center of gravity of all the points. The center of gravity of the scenario point cloud is the weighted average of the coordinates of all scenarios, such that the weights of each individual is equal to $1/I$. Thus, the center of gravity $\mathbf{g}$ is at point

$$\begin{aligned}\mathbf{g} &= \sum_{i=1}^{I} \frac{1}{I}\mathbf{x}_i = \sum_{i=1}^{I} \frac{1}{I}\begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{iK}\end{bmatrix} = \frac{1}{I}\begin{bmatrix} I_1 & I_2 & \cdots & I_K\end{bmatrix} \\ \Rightarrow \mathbf{g} &= \begin{bmatrix} p_1 & p_2 & \cdots & p_K\end{bmatrix}.\end{aligned}$$

The squared distance of a point $\mathbf{x}_i$ from the center of gravity $\mathbf{g}$ is then

$$\begin{aligned} d^2(\mathbf{x}_i, \mathbf{g}) &= \frac{1}{IJ}\sum_{k=1}^{K} \frac{1}{p_k}\left(x_{ik} - p_k\right)^2 \\ &= \frac{1}{IJ}\sum_{k=1}^{K} \frac{1}{p_k}\left(x_{ik}^2 - 2x_{ik}p_k + p_k^2\right) && |x_{ik}^2 = x_{ik} \\ &= \frac{1}{IJ}\left(\sum_{k=1}^{K}\frac{x_{ik}}{p_k} - 2\sum_{k=1}^{K} x_{ik} + \sum_{k=1}^{K} p_k\right) && |\sum_{k=1}^{K} x_{ik} = \sum_{k=1}^{K} p_k = J \\ \Rightarrow d^2(\mathbf{x}_i, \mathbf{g}) &= \frac{1}{IJ}\left(\sum_{k=1}^{K}\frac{x_{ik}}{p_k} - J\right). \end{aligned}$$

Now, we can calculate the inertia of the point cloud of scenarios in the K-dimensional space, where the mass of each point is set equal to one.

[1] We assume here that $p_k > 0$ for all k. If there exists k such that no row of $\mathbf{X}$ belong to category k, this column can be removed from $\mathbf{X}$ for redundancy.

$$
\begin{aligned}
\text{Inertia}(\mathcal{X}) &= \sum_{i=1}^{I} d^2(\mathbf{x}_i, \mathbf{g}) = \sum_{i=1}^{I} \frac{1}{IJ} \left(\sum_{k=1}^{K} \frac{x_{ik}}{p_k} - J \right) \\
&= \frac{1}{IJ} \sum_{i=1}^{I} \sum_{k=1}^{K} \frac{I}{I_k} x_{ik} - \frac{1}{IJ} \cdot IJ \\
&= \frac{1}{IJ} \sum_{k=1}^{K} \frac{I}{I_k} \sum_{i=1}^{I} x_{ik} - 1 \qquad\qquad | \sum_{i=1}^{I} x_{ik} = I_k \\
\Rightarrow \text{Inertia}(\mathcal{X}) &= \frac{K}{J} - 1 \qquad\qquad (4.4)
\end{aligned}
$$

Thus, we observe that total inertia of the scenario cloud does not depend on the content of the CDT, but only on the dimensionality of the table, through the number of variables J and total number of categories K. The point cloud $\mathcal{X}$ of consistent scenarios and the distances of i_1th and i_2th consistent scenarios and the center of gravity are illustrated in Figure 4.3.

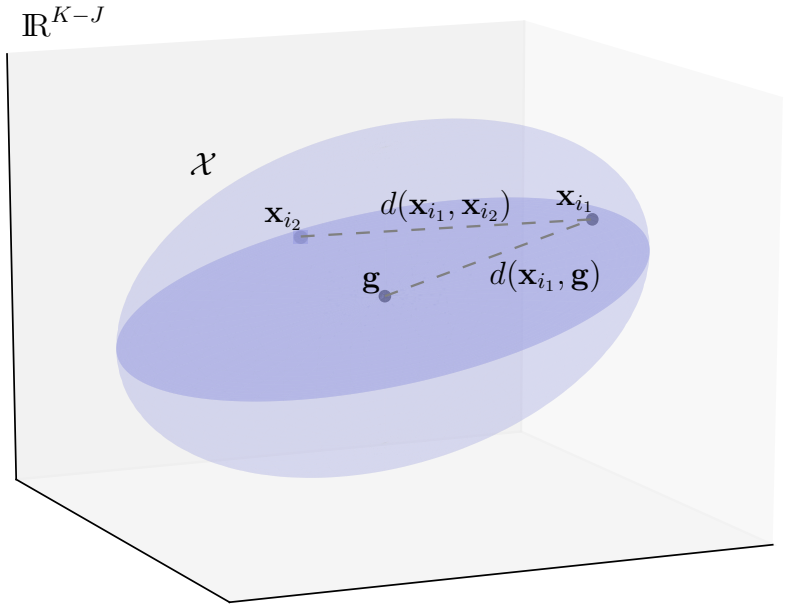


Figure 4.3: Illustration of the point cloud of consistent scenarios $\mathcal{X}$ and distances of two scenarios and the center of gravity.

4.1.3 Singular Value Decomposition

Next, we find the projection of point cloud $\mathcal{X}$ onto a lower-dimensional subspace that retains the most of the inertia of the original point cloud. To do this, we must first transform the complete disjunctive table $\mathbf{X}$ in a suitable way. In particular, the transformed table $\mathbf{Z}$ is obtained by (i) moving the center of gravity of the point cloud to the origin, and (ii) scaling the table in such a way that the distance metric defined in Equation (4.3) for row vectors $\mathbf{x}_i$ of $\mathbf{X}$ corresponds to the Euclidean distance metric for row vectors $\mathbf{z}_i$ of $\mathbf{Z}$. This transformation is formally defined as follows.

Definition 14. ***(Transformed complete disjunctive table)*** *Let* $\mathbf{X}$ *be the complete disjunctive table of scenarios as in Definition 12. Let* $\mathbf{P}$ *be a diagonal matrix of proportions* p_k *and* $\text{diag}(1/\sqrt{IJ})$ *be a diagonal matrix of size* $K \times K$ *with elements* $1/\sqrt{IJ}$ *in the diagonal. Then the transformed CDT* $\mathbf{Z}$ *is*

$$\mathbf{Z} = \left(\mathbf{X} - \mathbf{1}_I \mathbf{1}_K^{\mathrm{T}} \mathbf{P}\right) \text{diag}\left(1/\sqrt{IJ}\right) \mathbf{P}^{-1/2}. \tag{4.5}$$

With this transformation, the Euclidean distance corresponds the distance metric of Definition 13, because for two rows $\mathbf{z}_{i_1}, \mathbf{z}_{i_2}$ we see that

$$\begin{aligned} ||\mathbf{z}_{i_1} - \mathbf{z}_{i_2}||^2 &= \sum_{k=1}^{K} \left(\frac{1}{\sqrt{IJ}} \frac{1}{\sqrt{p_k}} \left(x_{i_1 k} - p_k\right) - \frac{1}{\sqrt{IJ}} \frac{1}{\sqrt{p_k}} \left(x_{i_2 k} - p_k\right) \right)^2 \\ \Rightarrow ||\mathbf{z}_{i_1} - \mathbf{z}_{i_2}||^2 &= \frac{1}{IJ} \sum_{k=1}^{K} \frac{1}{p_k} \left(x_{i_1 k} - x_{i_2 k}\right)^2 = d^2(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}). \end{aligned} \tag{4.6}$$

To obtain a lower-dimensional representation of $\mathbf{X}$, we can use singular value decomposition of the transformed CDT $\mathbf{Z}$.

Definition 15. ***(Singular value decomposition)*** *The singular value decomposition of matrix* $\mathbf{Z} \in \mathbb{R}^{I \times K}$ *is*

$$\mathbf{Z} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\mathrm{T}}, \tag{4.7}$$

where $\mathbf{\Sigma} \in \mathbb{R}^{R \times R}$ *is the diagonal matrix of singular values* σ_r *of* $\mathbf{Z}$*, and* $\mathbf{U} \in \mathbb{R}^{I \times R}$ *and* $\mathbf{V} \in \mathbb{R}^{K \times R}$ *are the orthogonal matrices of left and right singular vectors as columns, and* R *is the rank of* $\mathbf{Z}$*.*

The computation of the singular value decomposition is covered in various linear algebra textbooks [see, e.g., Hopcroft and Kannan (2012) or Strang (2016)]], and many mathematical programming languages (e.g, NumPy and

Matlab) have built-in functions to evaluate SVD.

The Frobenius norm $||\mathbf{Z}||_F$ of matrix $\mathbf{Z}$ is the square root of the sum of the squared elements z_{ik}^2 of $\mathbf{Z}$. It can be shown that the squared Frobenius norm of $\mathbf{Z}$ equals the total inertia of the point cloud $\mathcal{X}$:

$$||\mathbf{Z}||_F^2 = \sum_{i=1}^{I}\sum_{k=1}^{K}\left[\frac{1}{\sqrt{IJ}}\frac{1}{\sqrt{p_k}}\left(x_{i_1k}-p_k\right)\right]^2 = \sum_{i=1}^{I}\frac{1}{IJ}\sum_{k=1}^{K}\frac{1}{p_k}\left(x_{ik}-p_k\right)^2$$

$$\Rightarrow ||\mathbf{Z}||_F^2 = \sum_{i=1}^{I} d^2(\mathbf{x}_i,\mathbf{g}) = \text{Inertia}(\mathcal{X}). \tag{4.8}$$

Thus, the lower-dimensional representation of $\mathcal{X}$ that retains the most of its inertia is the representation of $\mathbf{Z}$ that maximizes the Frobenius norm. To find this representation, note that $||\mathbf{Z}||_F^2 = \text{tr}(\mathbf{Z}^\text{T}\mathbf{Z})$, where $\text{tr}(\cdot)$ is the trace of a matrix, i.e., the sum of its diagonal elements. Hence,

$$\begin{aligned}
||\mathbf{Z}||_F^2 &= \text{tr}(\mathbf{Z}^\text{T}\mathbf{Z}) = \text{tr}\left[\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\text{T}\right)^\text{T}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\text{T}\right] \\
&= \text{tr}\left(\mathbf{V}\mathbf{\Sigma}^\text{T}\mathbf{U}^\text{T}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\text{T}\right) && |\mathbf{U}\text{ is orthogonal} \\
&= \text{tr}\left(\mathbf{V}\mathbf{\Sigma}^\text{T}\mathbf{\Sigma}\mathbf{V}^\text{T}\right) && |\text{tr}(\mathbf{ABCD}) = \text{tr}(\mathbf{BCDA}) \\
&= \text{tr}\left(\mathbf{\Sigma}^\text{T}\mathbf{\Sigma}\mathbf{V}^\text{T}\mathbf{V}\right) && |\mathbf{V}\text{ is orthogonal} \\
\Rightarrow ||\mathbf{Z}||_F^2 &= \text{tr}\left(\mathbf{\Sigma}^\text{T}\mathbf{\Sigma}\right) = \sum_{r=1}^{R}\sigma_r^2.
\end{aligned} \tag{4.9}$$

Denoting the orthogonal columns of $\mathbf{U}$ and $\mathbf{V}$ by $\mathbf{u}_r$ and $\mathbf{v}_r$, $r = 1, 2, \ldots, R$, Equation (4.7) can be written in an alternative form:

$$\mathbf{Z} = \sum_{r=1}^{R}\sigma_r\mathbf{u}_i\mathbf{v}_i^\text{T}. \tag{4.10}$$

From Equations (4.9) and (4.10), we see that the contribution of the rth term in the summation (4.10) to the squared Frobenius norm of $\mathbf{Z}$ equals σ_r^2. Moreover, because the rank of an outer product of two vectors is one, including $\rho < R$ terms in the summation results in an approximation for $\mathbf{Z}$ with rank ρ. Hence, the best rank ρ approximation of $\mathbf{Z}$ in terms maximizing the Frobenius norm (and, thereby, the inertia of the original point cloud) is obtained by including those ρ terms in the summation with the largest corresponding singular values σ_r.

Let us denote by $\tilde{\mathbf{U}}, \tilde{\mathbf{\Sigma}}$ and $\tilde{\mathbf{V}}$ those matrices that have been obtained by choosing those columns from $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}$ that correspond to the ρ largest singular values σ_r, and by $\tilde{\mathbf{Z}}$ the corresponding approximation of $\mathbf{Z}$. Because the columns of $\tilde{\mathbf{V}}$ are orthonormal, we have $\tilde{\mathbf{Z}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^{\mathrm{T}} \Leftrightarrow \tilde{\mathbf{Z}}\tilde{\mathbf{V}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}} \in \mathbb{R}^{I\times\rho}$. Let us denote $\mathbf{Y} = \tilde{\mathbf{Z}}\tilde{\mathbf{V}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}$. The rows $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_I$ of matrix $\mathbf{Y}$ are the coordinates of the rows $\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2, \ldots, \tilde{\mathbf{z}}_I$ of the best rank ρ transformation $\tilde{\mathbf{Z}}$ of table $\mathbf{Z}$ in a ρ-dimensional subspace spanned by the column vectors of $\tilde{\mathbf{V}}$. These vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_\rho$ are referred to as the *principal dimensions*, and the vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_I$ as *principal coordinates*. The principal coordinates are defined more formally below.

Definition 16. ***(Principal coordinates of the most consistent scenarios)*** *Let $\mathbf{Z}$ be the transformed CDT of $\mathbf{X}$, and let $\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$ be the singular value decomposition of $\mathbf{Z}$. Let $\tilde{\mathbf{U}}$, $\tilde{\mathbf{\Sigma}}$, and $\tilde{\mathbf{V}}$ be matrices consisting of those columns of matrices $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}$ corresponding to the ρ largest singular values σ_r that define the best rank ρ approximation $\tilde{\mathbf{Z}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}$ of $\mathbf{Z}$. The matrix of ρ principal coordinates of $\mathbf{Z}$ is*

$$\mathbf{Y} = \tilde{\mathbf{Z}}\tilde{\mathbf{V}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}. \tag{4.11}$$

The set of principal coordinates $\mathcal{Y}$ is the set of rows $\mathbf{y}_i, i = 1, 2, \ldots, I$ of $\mathbf{Y}$:

$$\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_I\}. \tag{4.12}$$

The projection of the point cloud in Figure 4.3 onto a two-dimensional plane $\mathcal{V}$ spanned by vectors $\mathbf{v}_1, \mathbf{v}_2$ is illustrated in Figure 4.4.

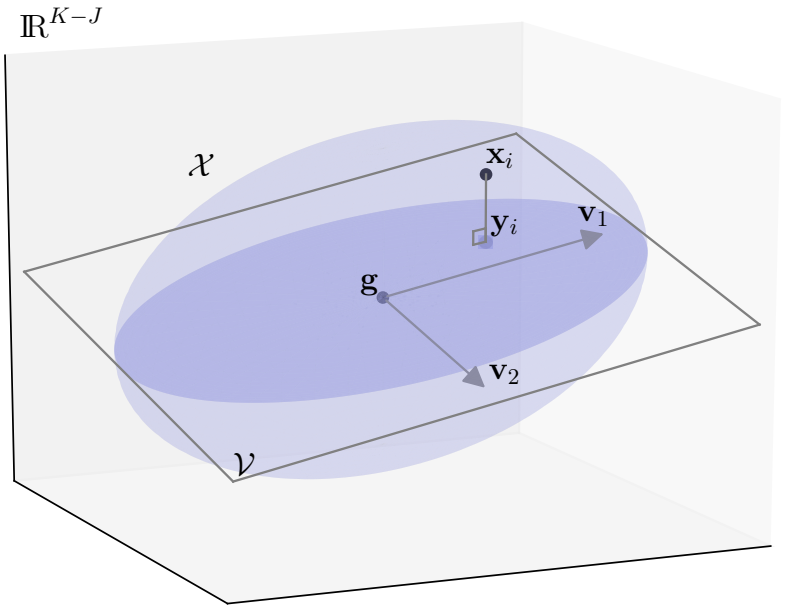


Figure 4.4: Illustration of the projection of the point cloud of the most consistent scenarios $\mathcal{X}$ onto a lower-dimensional space.

From Equations (4.8) and (4.9) we see that the inertia of dimension $\mathbf{v}_r$ is σ_r^2. Hence, the share of inertia explained by the rth principal dimension is

$$PI(r) = \frac{\sigma_r^2}{\sum_{r=1}^{R} \sigma_r^2}. \tag{4.13}$$

One property of matrix $\mathbf{Z}$ in MCA is that the maximal singular value is at most $\sigma_r = 1$ [Husson et al. (2017), Greenacre and Blasius (2006)]. The total inertia of point cloud $\mathcal{X}$ was found in Equation (4.4) to be $K/J - 1$, whereby

$$PI(r) = \frac{\sigma_r^2}{K/J - 1} \leq \frac{1}{K/J - 1} = \frac{J}{K - J}. \tag{4.14}$$

Thus, we have an upper bound for the percentage of inertia that can at most be explained by a single principal dimension, which is an intrinsic property of dimensionality reduction of categorical data. For example, in the electricity sales example, the upper bound of Equation (4.14) is $PI(r) \leq J/(K - J) = 6/(21 - 6) = 40\%$. Hence two principal dimensions can never explain more than $2 \cdot 40\% = 80\%$ of the total inertia of the point cloud in a problem with the same dimensionality as with the electricity sales example.

4.2 Scenario Map

Once we have evaluated the principal coordinates of the I most consistent scenarios, the projection $\mathcal{V}$ of the point cloud $\mathcal{X}$ of the consistent scenarios can be explored to find a small number M of scenarios from the set of consistent scenarios $\mathcal{S}^I$ which are mutually dissimilar. One approach could be to simply choose, e.g., $M = 4$ points from the principal coordinates representing three scenarios that maximize the sum of distances. However, because of the limitation of the projection quality in terms of the explained inertia, the selection of the M dissimilar scenarios requires a more interactive approach.

For this interactive approach, we define a mapping from the principal coordinates to the set of consistent scenarios. This mapping, which we call the Scenario Map, is defined as follows.

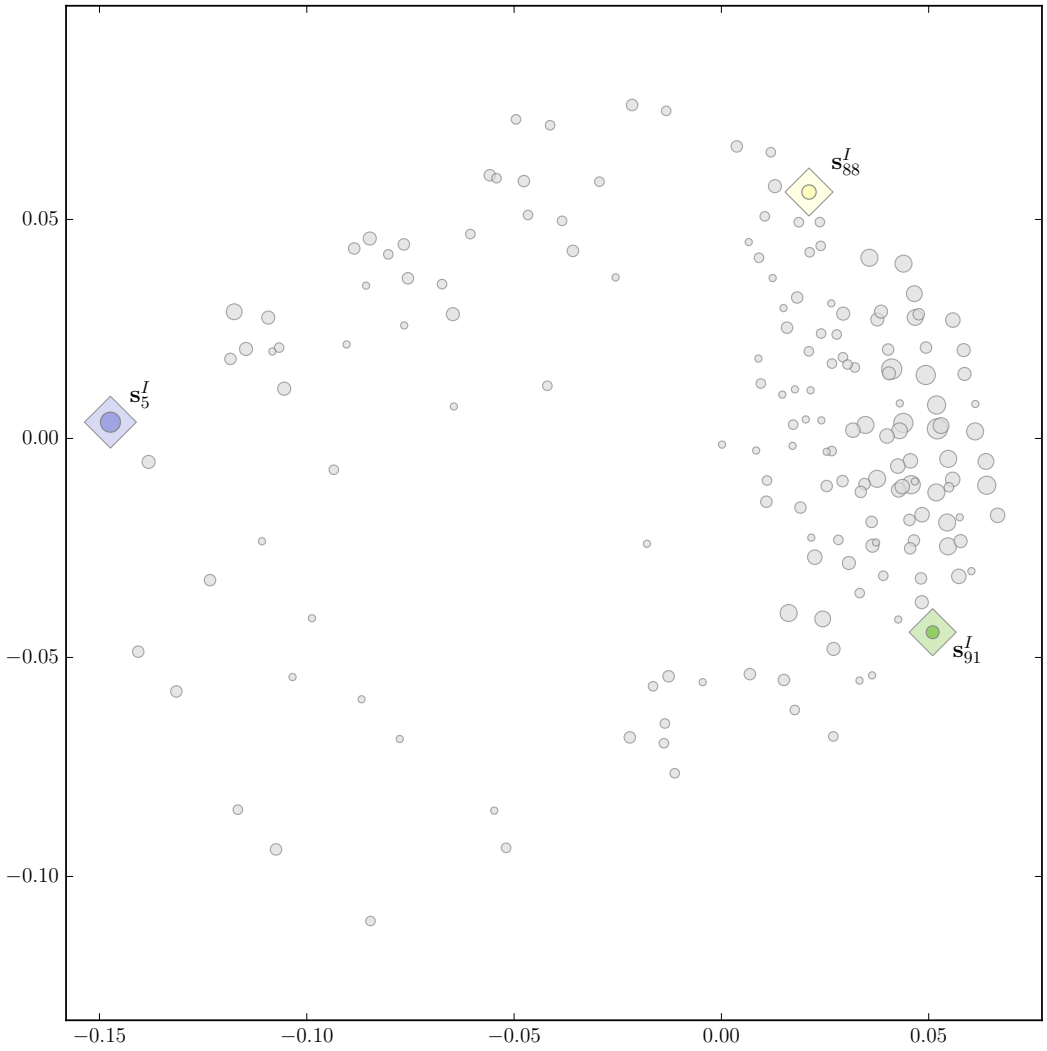
Definition 17. ***(Scenario Map)*** *Let $\mathcal{S}^I$ be the set of the I most consistent scenarios as in Definition 7 and let $\mathcal{Y}$ be the set of ρ principal coordinates of these scenarios as in Definition 16. Scenario Map f is a bijection from the set $\mathcal{Y}$ to the set $\mathcal{S}^I$:*

$$f : \mathcal{Y} \to \mathcal{S}^I. \tag{4.15}$$

The set $\mathcal{Y}$ of principal coordinates is the set of rows $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_I$ of $\mathbf{Y}$, each of which corresponds to the projected representation of scenario $\mathbf{s}_i^I, i = 1, 2, \ldots, I$ in the ρ-dimensional subspace. From this point forward, we only consider the set of $\rho = 2$ principal coordinates $\mathcal{Y}$ that can be illustrated with a 2D scatter plot. The scenarios in the set $\mathcal{S}^I$, in turn, can be illustrated with a morphological field. Hence, a Scenario Map with $\rho = 2$ principal coordinates can be constructed with a mapping from the scatter plot of two principal coordinates to the morphological field. This mapping can be done by utilizing various means to express similarity. As a similarity indicator, we use color coding that provides an identical color for the markers in the scatter plot of principal coordinates and to the representation of scenarios in the morphological field by coloured cells. Moreover, we use a different marker for those scenarios illustrated with Scenario Map. These kinds of mappings can be implemented with an interactive software, where selecting points in a scatter plot (e.g., with a mouse click) induces the illustration of the corresponding scenario.

Figure 4.5 illustrates the two-dimensional scatter plot for the Scenario Map corresponding to the set of the $I = 0.1 \cdot N = 172$ most consistent scenarios in the electricity sales example. The horizontal axis is the first principal dimension, explaining $PI(1) = 21.39\%$ of the inertia of the original scenario cloud [calculated through to Equation (4.13)], and the vertical axis is the second principal dimension, which explains $PI(2) = 10.34\%$ of the inertia. Thus, the scatter plot explains 31.73% of the inertia of the original $K - J = 15$ dimensional cloud in total. Here, each marker corresponds to one scenario such that the size of the marker is proportional to the overall consistency of the scenario. The mapping from the scatter plot (i.e., the principal coordinates $\mathcal{Y}$) to the morphological field (i.e., the set of consistent scenarios $\mathcal{S}^I$) for three scenarios $\mathbf{s}_5^I$, $\mathbf{s}_{88}^I$ and $\mathbf{s}_{91}^I$ is illustrated by the color codes such that

$$\begin{aligned}
\mathbf{s}_5^I &= \begin{bmatrix} 2 & 1 & 2 & 1 & 1 & 1 \end{bmatrix} \text{ (purple)}, \\
\mathbf{s}_{88}^I &= \begin{bmatrix} 3 & 4 & 4 & 3 & 2 & 2 \end{bmatrix} \text{ (light yellow)}, \\
\mathbf{s}_{91}^I &= \begin{bmatrix} 1 & 2 & 3 & 2 & 3 & 3 \end{bmatrix} \text{ (green)}.
\end{aligned}$$



Energy regulation's focus	**Electricity price**	**Competitive field**	**Activity of switching supplier**	**Digigalization & technology**	**Finnish economy**
Environment & renewable energy	Low, under 30 €/MWh	Traditional: private & municipal	Low, under 8%/year	Digital evolution	Deep recession
Energy security & reliability	Moderate, 30-45 €/MWh	Consolidation	Moderate, 9-14%/year	Fast digitalization	Zero growth
Market-based energy industry	High, over 45 €/MWh	International competitive field	High, over 15%/year	Digital revolution	Strong growth
Citizens: empowerment & protection	Turbulent, 0-200 €/MWh	New players from different industries			

Figure 4.5: Scenario Map from the principal coordinates to three consistent scenarios.

Because not all combinations of uncertainty factor levels correspond to consistent scenarios, a consistent scenario cannot be found directly from the morphological field by fixing one level of each uncertainty factors simultaneously. Nevertheless, fixing the levels of one or few uncertainty factors is likely to result in a nonempty subset of the set $\mathcal{S}^I$ of consistent scenarios, if I is sufficiently large. Following this idea, we define a mapping from a set of uncertainty factor levels to the set of consistent scenarios. This mapping, which we call the Uncertainty Factor Level Map, is defined as follows.

Definition 18. ***(Uncertainty Factor Level Map)*** *Let $\mathcal{S}^I$ be the set of the I most consistent scenarios. Let $\mathbf{j} = \begin{bmatrix} j_1 & j_2 & \cdots & j_{|\mathbf{j}|} \end{bmatrix}$ be a vector of uncertainty factors and let $\mathcal{K}^{\mathbf{j}} = \mathcal{K}_{j_1} \times \mathcal{K}_{j_2} \times \cdots \times \mathcal{K}_{|\mathbf{j}|}$ be the set of level combinations with factors $\mathbf{j}$. Uncertainty Factor Level Map (UFLM) g is a bijection from the set $\mathcal{K}^{\mathbf{j}}$ to the set $\mathcal{S}^I$:*

$$g : \mathcal{K}^{\mathbf{j}} \to \mathcal{S}^I \tag{4.16}$$

This mapping results in a subset of consistent scenarios $\mathcal{S}^{\mathcal{K}^{\mathbf{j}}} \subseteq \mathcal{S}^I$, each of which have a specific uncertainty factor level combination $\begin{bmatrix} k_{j_1} & k_{j_2} & \cdots & k_{j_{\mathbf{j}}} \end{bmatrix}$ corresponding to factors $\mathbf{j} = \begin{bmatrix} j_1 & j_2 & \cdots & j_{|\mathbf{j}|} \end{bmatrix}$. The subset can be found, e.g., by looking for those rows in matrix $\mathbf{S}^I$ that have value k_{j_1} in column j_1, value k_{j_2} in column j_2, and so on and so forth. This subset can then be illustrated in the two-dimensional scatter plot with (inverse) Scenario Map.

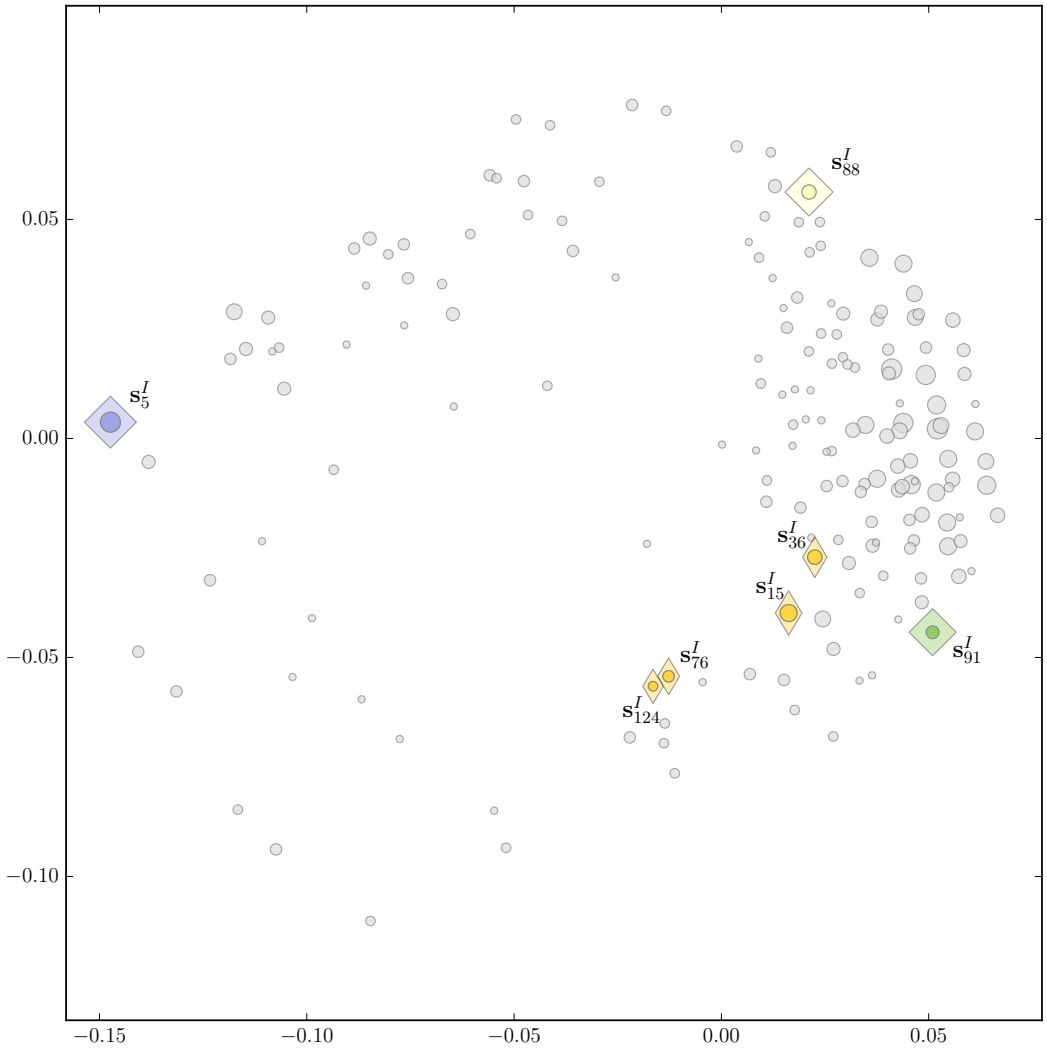
Using Scenario Map and UFLM, the set $\mathcal{S}^I$ of consistent scenarios can be explored iteratively by (i) inspecting the consistent scenarios and their dissimilarities according to their principal coordinates in the two-dimensional scatter plot and (ii) seeking scenarios corresponding to specific uncertainty factor level combinations in the morphological field. Consider, for instance, the Scenario Map for the electricity sales example shown in Figure 4.5. We seek $M = 4$ mutually dissimilar scenarios such that (a) each level of each uncertainty factor is present in at least one scenario, and (b) for each pair of scenarios, the number of uncertainty factors that have the same level is at most one. While these conditions are not always possible to satisfy, for small scenario planning problems they are feasible.

First, we note that scenarios $\mathbf{s}^I_5$, $\mathbf{s}^I_{88}$ and $\mathbf{s}^I_{91}$ illustrated in Figure 4.5 satisfy the aforementioned conditions. Then, to help select the fourth scenario, we use UFLM to find the subset of consistent scenarios in which the first, second, and third uncertainty factors (i.e., $\mathbf{j} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$) have levels $k_1 = 4, k_2 = 3$ and $k_3 = 1$, because these levels are not present in any of the three selected scenarios. This mapping from $\begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 1 \end{bmatrix} \in \mathcal{K}_1 \times \mathcal{K}_2 \times \mathcal{K}_3 = \mathcal{K}^{\mathbf{j}}$ to the

set of consistent scenarios $\mathcal{S}_I$ results in the subset $\mathcal{S}^{\mathcal{K}^{\mathbf{j}}} = \{\mathbf{s}_{15}^I, \mathbf{s}_{36}^I, \mathbf{s}_{76}^I, \mathbf{s}_{124}^I\} \subset \mathcal{S}^I$ of consistent scenarios. These scenarios are illustrated by gold color and different markers in Figure 4.6, where the colouring corresponding to scenarios $\mathbf{s}_5^I$, $\mathbf{s}_{88}^I$ and $\mathbf{s}_{91}^I$ has been removed from the morphological field for purposes of clarity.

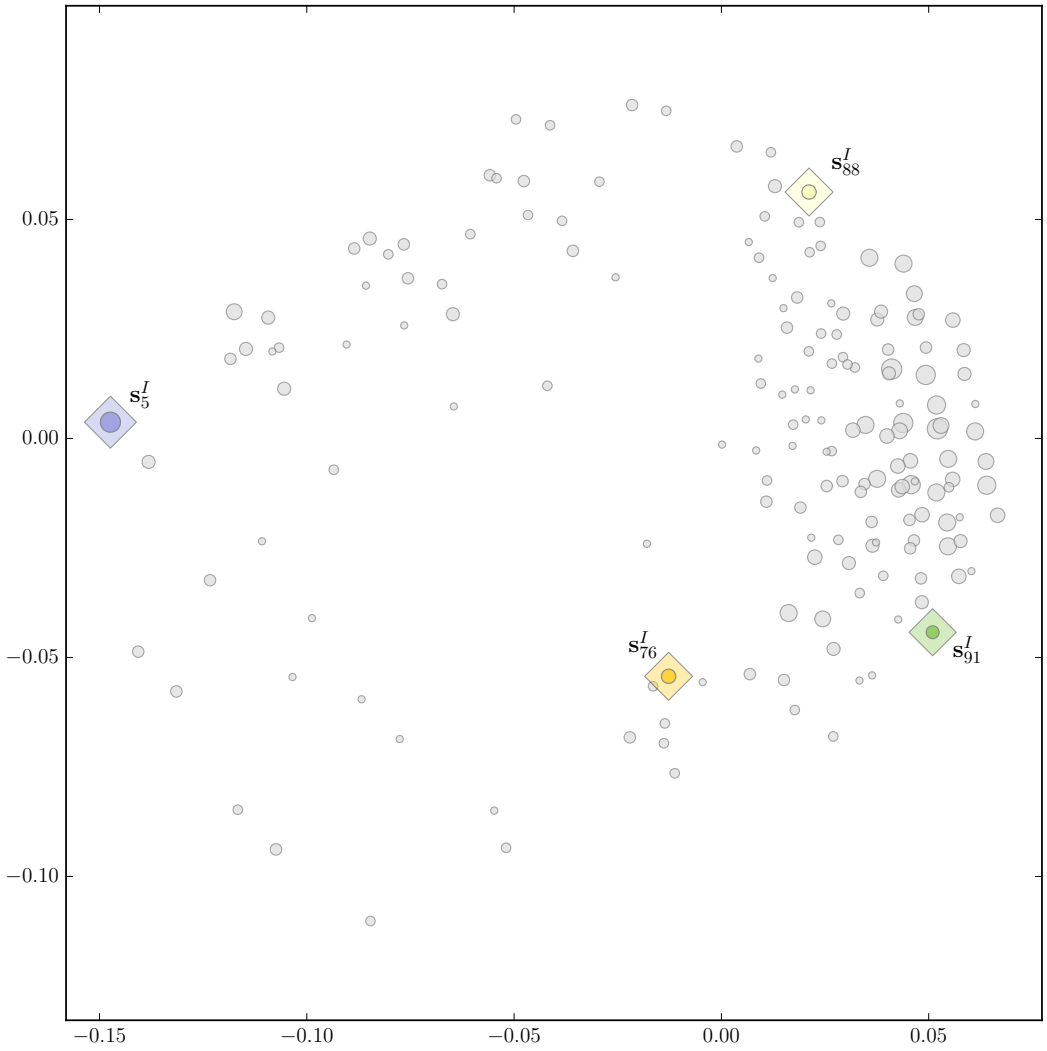
Out of the scenarios marked with gold colour in Figure 4.6, we select $\mathbf{s}_{76}^I = \begin{bmatrix}4 & 3 & 1 & 1 & 2 & 3\end{bmatrix}$ as the fourth scenario, because this scenario shares only one uncertainty factor level with each of the three already selected scenarios. The final set of $M = 4$ internally consistent and mutually dissimilar scenarios for the petrochemical example is illustrated in Figure 4.7. These scenarios and their internal consistencies are

$$\begin{aligned}
\mathbf{s}_5^I &= \begin{bmatrix}2 & 1 & 2 & 1 & 1 & 1\end{bmatrix}, & \alpha(\mathbf{s}_5^I) &= 1.6,\\
\mathbf{s}_{76}^I &= \begin{bmatrix}4 & 3 & 1 & 1 & 2 & 3\end{bmatrix}, & \alpha(\mathbf{s}_{76}^I) &= 1.267,\\
\mathbf{s}_{88}^I &= \begin{bmatrix}3 & 4 & 4 & 3 & 2 & 2\end{bmatrix}, & \alpha(\mathbf{s}_{88}^I) &= 1.267,\\
\mathbf{s}_{91}^I &= \begin{bmatrix}1 & 2 & 3 & 2 & 3 & 3\end{bmatrix}, & \alpha(\mathbf{s}_{91}^I) &= 1.2.
\end{aligned}$$



Energy regulation's focus	**Electricity price**	**Competitive field**	**Activity of switching supplier**	**Digigalization & technology**	**Finnish economy**
Environment & renewable energy	Low, under 30 €/MWh	Traditional: private & municipal	Low, under 8%/year	Digital evolution	Deep recession
Energy security & reliability	Moderate, 30-45 €/MWh	Consolidation	Moderate, 9-14%/year	Fast digitalization	Zero growth
Market-based energy industry	High, over 45 €/MWh	International competitive field	High, over 15%/year	Digital revolution	Strong growth
Citizens: empowerment & protection	Turbulent, 0-200 €/MWh	New players from different industries			

Figure 4.6: UFLM from $\begin{bmatrix}4 & 3 & 1\end{bmatrix} \in \mathcal{K}_1 \times \mathcal{K}_2 \times \mathcal{K}_3$ to the set $\mathcal{S}^I$ of most consistent scenarios, which is visualized with (inverse) Scenario Map.



Energy regulation's focus	Electricity price	Competitive field	Activity of switching supplier	Digigalization & technology	Finnish economy
Environment & renewable energy	Low, under 30 €/MWh	Traditional: private & municipal	Low, under 8%/year	Digital evolution	Deep recession
Energy security & reliability	Moderate, 30-45 €/MWh	Consolidation	Moderate, 9-14%/year	Fast digitalization	Zero growth
Market-based energy industry	High, over 45 €/MWh	International competitive field	High, over 15%/year	Digital revolution	Strong growth
Citizens: empowerment & protection	Turbulent, 0-200 €/MWh	New players from different industries			

Figure 4.7: Set of $M = 4$ internally consistent and mutually dissimilar scenarios visualized with Scenario Map.

Chapter 5

Case study

As a part of this thesis, an interactive software tool, Scenario Builder™ [Capful (2018)], was developed for a Finnish management consultancy Capful that specialises in scenario work and strategy development and implementation. The tool implements the scenario identification and visualization method of this thesis and automates various tasks, such as setting up a morphological field and constructing a draft consistency table. In this chapter, we present the real case where this tool was successfully applied in a project to identify a set of consistent and mutually dissimilar scenarios for the Finnish National Emergency Supply Organization.

National Emergency Supply Organization (NESO) is a network that maintains and develops security of supply in Finland on the basis of public-private partnership initiatives [Huoltovarmuuskeskus (2016)]. Its primary objective is to ensure the conditions necessary for the operations of organisations that are critical to security of supply. NESO consists of the following agencies and bodies:

- The National Emergency Supply Agency is tasked with planning and measures related to developing and maintaining security of supply.
- The National Emergency Supply Council is a body that assesses and reviews the general state of security of supply.
- NESO sectors are organizations responsible for steering, coordinating, and monitoring preparedness in their respective security of supply fields. These organizations consist of public authorities, associations and other significant operators in the security of supply fields.
- NESO pools comprise companies and businesses responsible for operational preparedness in their security of supply fields.

From fall 2017 to spring 2018, the NESO developed scenarios with Capful in a project 'Huoltovarmuuden skenaariot 2030 (Scenarios for maintaining the security of supply in 2030)' to provide insight into the future of security of supply. In this project, NESO was seeking answers to the following two questions:

(i) What do the alternative future scenarios look like with respect to emergency supply work?

(ii) How are the essential operations of the Finnish society secured in the alternative futures described by these scenarios?

The goals of the project were to provide foresight for the upcoming decisions of the Finnish Government on security of supply and to support the strategic and operational decision making of NESO. Moreover, the developed national-level scenarios published in the public report of this scenario project work as the starting point for many company-level scenario development projects of, e.g., NESO pools member companies.

The scenario development project was carried out in four workshops and several other meetings which were attended by numerous representatives from NESO sectors and pools, as well as external subject experts. In the first workshop, the process began by recognizing ten relevant uncertainty factors. Four possible levels were identified for each of these uncertainty factors, brief descriptions of which are presented in Table 5.1. In the actual scenario building process, more holistic developments from year 2018 to year 2030 were associated with each level, detailing relevant events and plot elements for these developments.

In between the first and second workshop, the $10(10-1)/2 \cdot 4^2 = 720$ pairwise consistency values between each pair of levels were initially assessed by professional scenario practitioners at Capful; this took approximately one working day. In the following workshop, these consistency values were discussed, verified, and partly revised with the stakeholders from NESO. The final consistency table is presented in Table 5.2.

Table 5.1: Uncertainty factors and their levels in 'Huoltovarmuuden skenaariot 2030 (Scenarios for maintaining the security of supply in 2030)'.

A.	Globalization & international relationships	1. Slow & steady globalization 2. Fast & marked-driven globalization 3. Blocs & new alliances 4. Protectionism & deglobalization
B.	Geopolitical focus	1. Traditional industrialized countries 2. Marked-driven shift of focus to east 3. Strongly state-led countries increase their role 4. Shift of power to supranational organizations
C.	State of security	1. Territorial disputes 2. Conflicts related to (hybrid)influencing 3. Intensified terrorism 4. Conflicts subside
D.	Foreign affairs of Finland	1. Non-aligned & neutral Finland 2. Allied within the European level 3. Allied over the European level & NATO membership 4. Bilateral defence cooperation pacts
E.	Natural resources & migrations	1. Strong regional scarcities of resources 2. Unequal distribution of resources 3. Increased resource efficiency & distribution 4. Decreasing extreme climate phenomena
F.	Data accessibility & information security	1. Open & public data 2. Cyber risks & cybersecurity 3. Data monopolies 4. Post-truth era
G.	Global economy	1. Equally distributed economic growth 2. Polarized economic growth 3. Economic stagnation 4. Financial crises & turbulence
H.	Intelligent systems & machines	1. Digital evolution - Machines help humans 2. Digital leap - Machines alongside humans 3. Digital revolution - Machines making the decisions 4. Digital stagnation
I.	Industrial structure & jobs	1. Traditional 2. Everything-as-a-service 3. New wave of globalization 4. Era of robotization
J.	Development of EU	1. Slow decline of the EU 2. Harmonization & strong interdependence 3. Internal blocs within Europe 4. Decay of the EU

Table 5.2: Consistency values between each pair of factor-specific levels.

		B				C				D				E				F				G				H				I				J			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
A	1	3	2	-3	-1	-3	-2	1	3	1	2	1	-1	1	-3	3	1	2	-2	-2	-1	3	-1	1	-2	2	2	1	-1	2	2	1	0	-1	3	-1	-3
	2	1	3	-2	3	-1	1	-1	1	0	1	0	-2	1	-1	1	0	1	-1	2	0	1	3	-1	2	1	2	2	-1	-1	2	3	2	1	1	2	1
	3	1	1	2	-2	3	3	1	-1	-1	2	2	1	2	2	-1	-1	0	2	0	1	-1	1	2	3	1	0	0	1	2	1	-2	0	1	-1	3	1
	4	1	-1	2	-3	3	3	3	-2	2	-1	-1	3	2	3	-3	-1	-2	3	1	3	-2	-1	3	2	1	0	0	2	2	1	-3	0	3	-2	1	3
B	1					-1	2	1	2	-2	3	2	-3	2	-1	2	1	1	-1	-1	1	2	2	2	1	2	2	1	0	2	3	-1	1	-2	3	2	-2
	2					-1	2	0	2	2	1	-1	0	1	2	1	1	1	0	1	1	2	2	-1	1	1	2	2	-1	-1	2	2	1	2	-1	2	2
	3					2	3	1	-2	-2	2	2	1	1	2	-2	-1	-2	2	2	3	-2	1	2	2	1	0	-1	1	1	0	-2	-1	1	1	0	-1
	4					-2	2	1	0	1	0	-1	-1	0	1	-1	0	2	0	2	1	1	2	-1	2	-1	2	2	-2	-2	2	3	1	-1	3	1	-1
C	1									0	2	3	1	3	2	-2	-2	-2	2	1	2	-3	1	1	2	1	0	0	2	0	0	-1	0	1	2	2	1
	2									0	3	2	1	1	1	-2	-2	-2	3	2	2	-2	1	1	1	1	0	0	2	0	0	-1	2	1	2	2	1
	3									-2	2	1	0	2	2	-2	-2	0	2	0	1	-2	2	1	1	1	0	0	2	0	0	-1	0	1	1	2	2
	4									1	2	0	-1	-2	-2	3	3	3	1	-1	-2	3	-1	0	-2	1	1	0	-3	1	1	2	0	1	3	-1	-3
D	1													1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	2	0	2	-3	-1	3
	2													2	2	2	1	2	0	1	0	2	1	1	0	0	0	0	0	0	0	2	0	-2	3	1	-3
	3													1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1
	4													1	2	-2	-2	-1	2	0	0	-2	1	0	1	0	0	0	1	0	0	2	0	1	-2	-1	2
E	1																	0	1	1	1	-3	2	2	2	1	0	0	2	1	1	2	0	1	1	3	1
	2																	-1	2	2	2	-3	3	3	2	1	0	-1	2	2	1	1	0	1	1	2	1
	3																	2	-1	-1	-2	3	-1	-2	-2	1	3	3	-3	0	2	1	1	-1	1	-1	-1
	4																	1	-1	0	0	3	1	-2	-2	2	3	3	-2	0	1	0	1	-1	1	-1	-1
F	1																					2	-1	-1	-1	1	3	2	-3	1	3	1	1	-2	3	-2	-3
	2																					-1	2	3	2	1	0	-1	2	2	-1	3	1	2	-2	2	2
	3																					-2	2	1	1	1	1	2	1	1	2	1	1	2	-2	3	1
	4																					-1	1	1	1	1	0	0	2	1	0	0	0	1	-1	1	1
G	1																									1	2	-2	-2	1	2	1	1	1	2	-2	-2
	2																									0	1	3	-1	1	2	2	3	1	-2	3	2
	3																									1	-1	1	1	2	-1	-1	0	2	-1	3	1
	4																									1	0	2	2	1	-1	2	2	2	-1	3	1
H	1																													3	1	1	-1	0	0	0	0
	2																													-1	2	2	2	0	0	0	0
	3																													-3	2	2	3	0	0	0	0
	4																													3	-3	-1	-3	1	-1	0	2
I	1																																	1	2	0	1
	2																																	0	0	-2	0
	3																																	2	1	0	2
	4																																	0	0	0	0

Before and during the second workshop, the set of $I = 3000$ most consistent scenarios was identified by using techniques described in Chapter 3. The number of $I = 3000$ was selected to be sufficiently large to represent most

of the many uncertainty factor levels of Table 5.1, but small enough to be comprehensively manageable. Figure 5.1 shows the two-dimensional Scenario Map representation of these scenarios, created by using techniques described in Chapter 4. Here, the first (horizontal axis) and second (vertical axis) principal dimensions explain 21.88% and 8.86% of the inertia of the $K - J = 40 - 10 = 30$-dimensional scenario cloud, respectively.

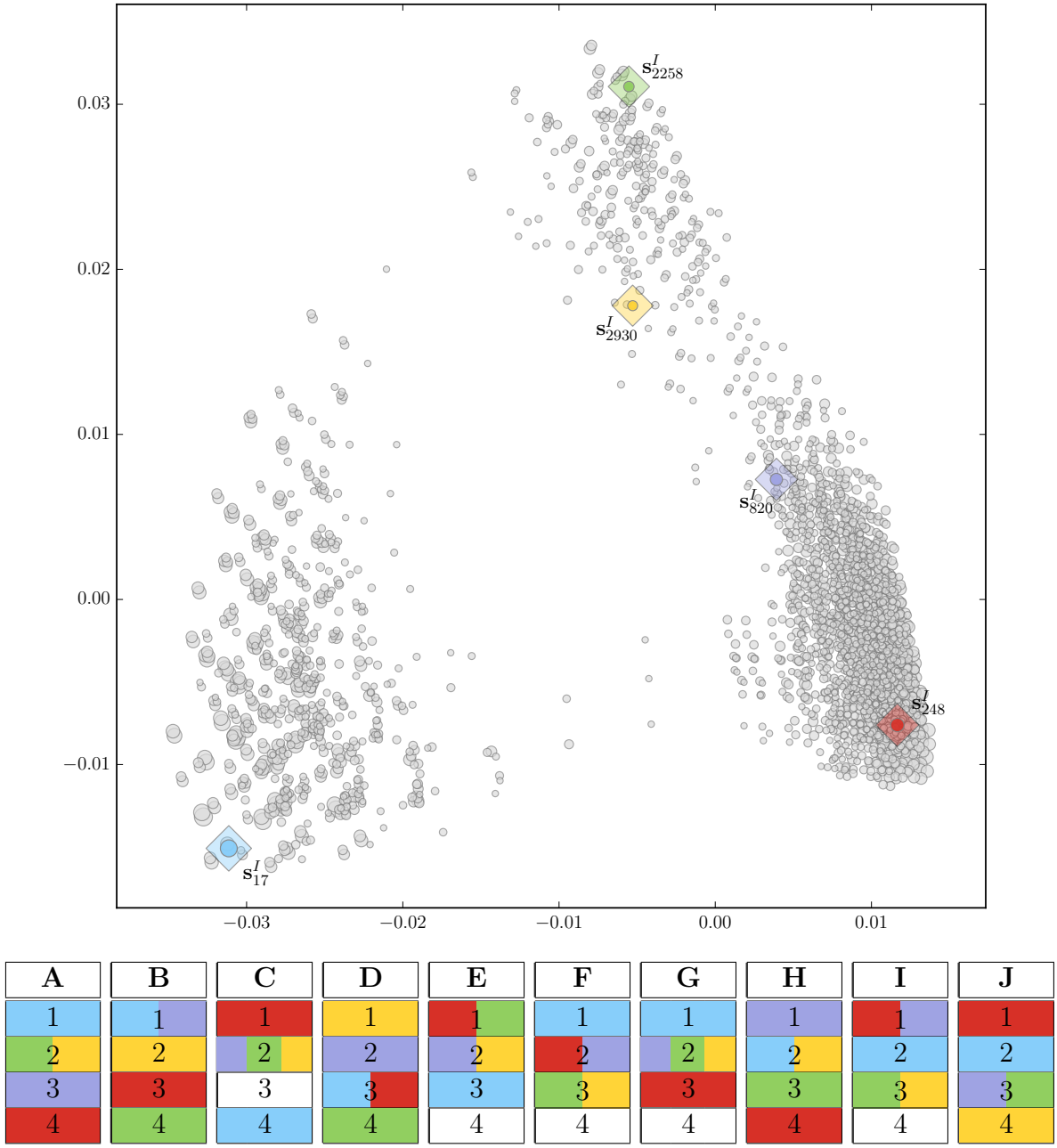


Figure 5.1: $M = 5$ internally consistent and mutually dissimilar scenarios of NESO case visualized with Scenario Map.

The Scenario Map and UFLM were used to iteratively find $M = 5$ mutually dissimilar and internally consistent scenarios that together would cover the majority of possible uncertainty factor levels. As a starting point, the following scenarios were used, illustrated by color codes in the Scenario Map in Figure 5.1:

$$\begin{aligned}
\mathbf{s}^I_{17} &= \begin{bmatrix}1 & 1 & 4 & 3 & 3 & 1 & 1 & 2 & 2 & 2\end{bmatrix} \text{(blue)}, & \alpha(\mathbf{s}^I_{17}) &= 1.867,\\
\mathbf{s}^I_{248} &= \begin{bmatrix}4 & 3 & 1 & 3 & 1 & 2 & 3 & 4 & 1 & 1\end{bmatrix} \text{(red)}, & \alpha(\mathbf{s}^I_{248}) &= 1.578,\\
\mathbf{s}^I_{820} &= \begin{bmatrix}3 & 1 & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 3\end{bmatrix} \text{(purple)}, & \alpha(\mathbf{s}^I_{820}) &= 1.489,\\
\mathbf{s}^I_{2258} &= \begin{bmatrix}2 & 4 & 2 & 4 & 1 & 3 & 2 & 3 & 3 & 3\end{bmatrix} \text{(green)}, & \alpha(\mathbf{s}^I_{2414}) &= 1.378,\\
\mathbf{s}^I_{2930} &= \begin{bmatrix}2 & 2 & 2 & 1 & 2 & 3 & 2 & 2 & 3 & 4\end{bmatrix} \text{(gold)}, & \alpha(\mathbf{s}^I_{2930}) &= 1.356.
\end{aligned}$$

Some uncertainty factor levels are not present in any of these initial scenarios, namely, levels C.3, E.4, F.4, G.4, and I.4. To include these factor levels in the analysis, some of the scenarios were combined with other scenarios that had nearly identical level configurations and markers close to the initial scenarios in the two-dimensional plane. In particular, scenarios $\mathbf{s}^I_{17}$ and $\mathbf{s}^I_{248}$ were combined with scenarios $\mathbf{s}^I_{384} = \begin{bmatrix}1 & 1 & 4 & 2 & 4 & 1 & 3 & 2 & 2 & 2\end{bmatrix}$ and $\mathbf{s}^I_{2386} = \begin{bmatrix}3 & 3 & 3 & 4 & 2 & 4 & 3 & 4 & 1 & 4\end{bmatrix}$, respectively, as illustrated by their blue and red color coding in Figure 5.2. In practice, a combination scenario $(\mathbf{s}^I_{i_1}, \mathbf{s}^I_{i_2})$ would describe a future in which some uncertainty factors would have multiple levels; e.g., in the red combination scenario, Global & international relationships (factor A, cf. Table 5.1) would be characterized both by Blocs & new alliances (level 3) and Protectionism & deglobalization (level 4).

To include the remaining factor levels G.4 and I.4, these levels were added to existing scenarios such that the added levels would be consistent with the levels already contained in the scenarios. To do this, Scenario Map and UFLM were utilized. For example, based on the UFLMs visualized in Figure 5.3, scenarios with level 3 (New wave of globalization) on factor I (Industrial structure & jobs) are close to scenarios with level 4 (Era of robotization) on the same factor. Hence, it was decided that level I.4 would be added to scenarios $\mathbf{s}^I_{2258}$ (green) and $\mathbf{s}^I_{2930}$ (gold) which already contained level I.3. The final set of $M = 5$ scenarios is shown in Table 5.3.

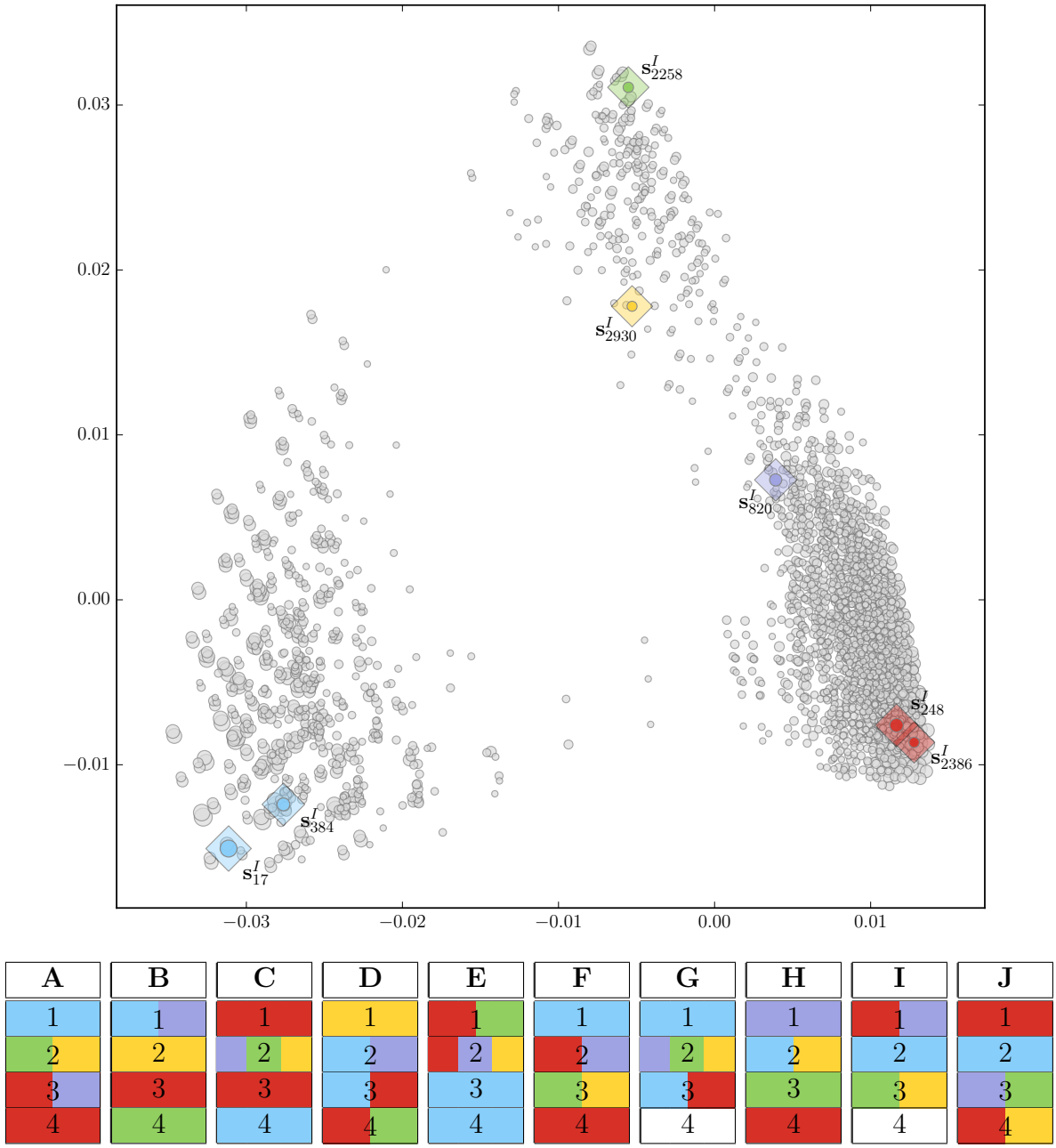
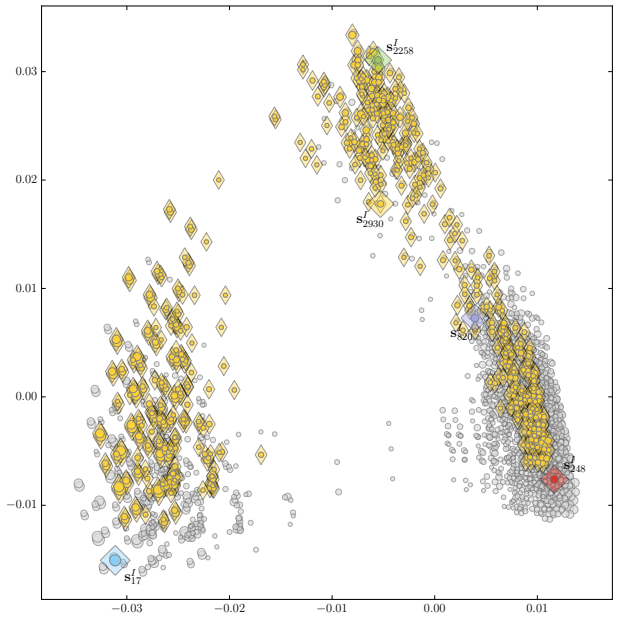
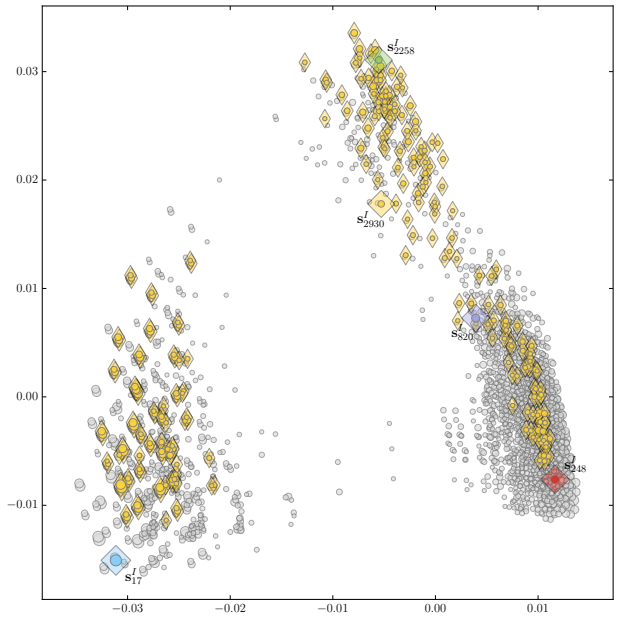


Figure 5.2: Illustration of combination scenarios $(\mathbf{s}^I_{17}, \mathbf{s}^I_{248})$ and $(s^I_{384}, \mathbf{s}^I_{2386})$.



A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4

(a) ULFM from [3] $\in \mathcal{K}_9$ to the set $\mathcal{S}^I$ of most consistent scenarios.



A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4

(b) ULFM from [4] $\in \mathcal{K}_9$ to the set $\mathcal{S}^I$ of most consistent scenarios.

Figure 5.3: UFLM for two levels 3 and 4 of factor I.

Table 5.3: Final scenarios of 'Huoltovarmuuden skenaariot 2030 (Scenarios for maintaining the security of supply in 2030)'.

A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4

This set of five scenarios was refined, concretized, focused, and further worked on in the third workshop. As a result of this workshop, five plausible stories of the future were created. These stories comprise, e.g., (i) three-phase

timelines of logical events occurring in years 2018-2021, 2022-2026, and 2026-2030, amplified with effective visualizations, and (ii) influence diagrams crystallizing the interdependencies of the driving forces and focal events of each scenario. The summary scenario narratives of the five final scenarios are presented in Table 5.4. More extensive depictions of the scenarios can be found in the public report of the project [Huoltovarmuuskeskus (2018)]. In the fourth workshop, the scenarios were further specified by explicating their consequences for NESO sectors and pools in more detail. In the final workshop, actions and policies for emergence supply work in each scenario were analysed as well.

Table 5.4: NESO scenarios.

1. Global interdependence	
The financial crisis of developing countries reveals the weaknesses of the leading developing countries, which strengthens the position of the west as safe haven. The high costs of conflicts and catastrophes compel countries to cooperate more closely. The role of international institutes becomes crucial. Applications of blockchain technology create new operating models and enhance trust. Global conflicts and migrations become manageable, and the internal cooperation of EU will tighten. The role of EU countries in Nato strengthens and Finland joins Nato as well. International regulation and strict climate	and environmental politics increase as the Western countries promote international agreements. However, global economic growth slows down when the social problems of China grow. The globalization process slows down and becomes more controlled. Regulatory frameworks for digital trust platforms and transactions of sharing economy are developed. Highly educated Asians migrate to Western countries in the pursuit of higher standard of living, clean air, and democratic society. The role of EU in international politics increases.
Scenario 2: Armed power politics	
Increasingly severe climatic conditions and expanding military conflicts create unprecedently vast international migration from Eastern Africa and Middle East towards Europe all the way up to Finland. As a result of nationalist and protectionist mindset, states close themselves and detach from international agreements. The conflicts culminate in Middle East, North Korea, and Eastern Ukraine. In the era of armed power politics, the world drifts into conflicts and threats to sovereignty of states grow. Finland joins Nato as defence cooperation agreements divide nations	into groups. Massive migratory movements increase clashes of cultures, generating platforms for extensive organised terrorism. EU is left behind in the interests of defensive alliances and individual states. Maritime transport in the Baltic Sea is disrupted, which creates problems in trade and energy delivery in the area. The situation of the security politics and cyber risks reduce cooperation and sharing of information between countries. The Baltic Sea region, the Arctic, and other critical logistic routes work as arenas for power politics. The armament of space begins.

Scenario 3: Blocs and hybrid influencing
Politics based on national interests inhibit international cooperation and the strengths on international agreements decline. The line between war and peace is blurred and cyber attacks increase. When severe hybrid operations are revealed, nations awake to the real risks of hybrid influencing. The world divides into blocs and rivalling alliances. Differences in social, economical, and ethical values between the blocs rise strongly. Cybersecurity actions amplify and critical ICT systems of rivalling blocs and states are differentiated. EU focuses in enhancing internal trade and forming a hybrid defence alliance. As a result of information influencing, the people's trust in institutions is weakened and the cohesive factors in societies are blurred. The prevalent trust of the early 21st century in states, companies, information, and information systems is regarded as naive. Blocs and states develop own internal systems to control information, prevent external influence, and improve security.
Scenario 4: Technological world order
Technological development accelerates. Robotics and artificial intelligence radically change operating models and labour requirements. Experts succeed, but the unemployment of low-educated people rises. Increasing networking and economies of scale compel systems to become global. Global tech giants reinforce their dominant positions as the owners of data and information. Transitions occur from national solutions to supranational organizations and from public to private services. Increasing proportion of public services fall behind in the rapid technological development. Responsibility for livelihood and welfare shifts even more from societies to individuals. Although industries centralize to large operators, technologies and platform economies enable the scattering of work and production. The role of cities compared to states grows, and global geopolitical focal points centralize to the technological hubs of the world. Ubiquitous integration of digitalization, automation, and robotics enables superiority of artificial intelligence to human reasoning ability. The capital created by technology centralizes, thus increasing inequality between people and regions.
Scenario 5: Dominance of the East
The value of the natural resources and rare raw-materials possessed by China rise, when technologies increase their consumption. China and emerging Asian economies improve their competitive advantage. The West is shaken by stock market crash and burst of real estate bubble emanating from the US. Protectionism, growing debt, and internal political problems weaken many Western countries. China, Russia, and Islamic States find the common will to break the dominance of the West. Traditional democracy is considered to be ineffective, stiff and even dangerous. Western values and social models no longer form the basis of international actions. The integrity and power of EU is weakened. Finland identifies itself as a neutral gateway between the East and the West. Asian investments increase in European regions, including strategically important locations. Chinese-funded infrastructure projects unites Eurasia, and the Northern Sea Route becomes more significant logistically. The center of economy, business, and politics moves to the East. The globalization development in the world of Eastern dominance is a combination of market economy, technological development, and political control by the elite.

Chapter 6

Discussion

In this chapter, we discuss advantages and limitations of the scenario identification and visualization method of this thesis in Sections 6.1 and 6.2, respectively. The advantages and limitations are considered with respect to the four objectives of scenario development in Table 2.1: explorativeness, trustworthiness, efficiency, and accessibility. Section 6.3 then presents approaches to extend the method.

6.1 Advantages

The visualization of the set of consistent scenarios with Scenario Map helps discover unprecedented combinations of uncertainty factor levels contained in this set, and hence our method fulfils the explorativeness aspect of Table 2.1. In fact, the dimensionality reduction technique (MCA) used for the visualization of consistent scenarios is, by one of its many definitions, an exploratory data analysis technique [Husson et al. (2017)]. By using an exploratory data analysis technique in the selection of dissimilar consistent scenarios (cf. mathematical optimization), our method does not strictly impose selections of any particular scenarios. Instead, our method visualizes and reveals the underlying structure of a relatively large filtered set of I scenarios which contains only the most consistent scenarios that can be deemed plausible. Then, the selection of the skeletal scenarios from which the final scenarios are created is retained for judgement of the scenario planning intervention participants. As such, the proposed method for selecting scenarios based on their dissimilarity in the Scenario Map can be considered as a *weakly formative inductive scenario identification method*.

The scenario identification method of this thesis is very transparent, thus fostering the trustworthiness aspect of Table 2.1. The transparency of the method is supported by advancing from small, easily acceptable deductions, which is characteristic to inductive scenario development processes. The framework for these deductions is well defined and structured, proceeding from (i) the morphological field to (ii) consistency assessments of pairs of uncertainty factor levels, and then from the consistency table to (iii) the set of the I most consistent scenarios, which is visualized with (iv) dimensionality reduction techniques, finally (v) arriving at the set of M mutually dissimilar consistent scenarios based on examining distances. Any scenarios developed with our method can be traced back to the elementary inferences they were based on in the morphological field and the consistency table.

Given the consistency assessments of pairs of uncertainty factors levels, the following deductions of the scenario identification method of this thesis can be made very efficiently. The algorithms developed in Section 3.2 are sufficiently efficient to generate the set of consistent scenarios in a workshop setting, even if there are many uncertainty factors and their levels. Moreover, the comprehensive visualization the Scenario Map makes it possible to explore this set of consistent scenarios in a timely manner.

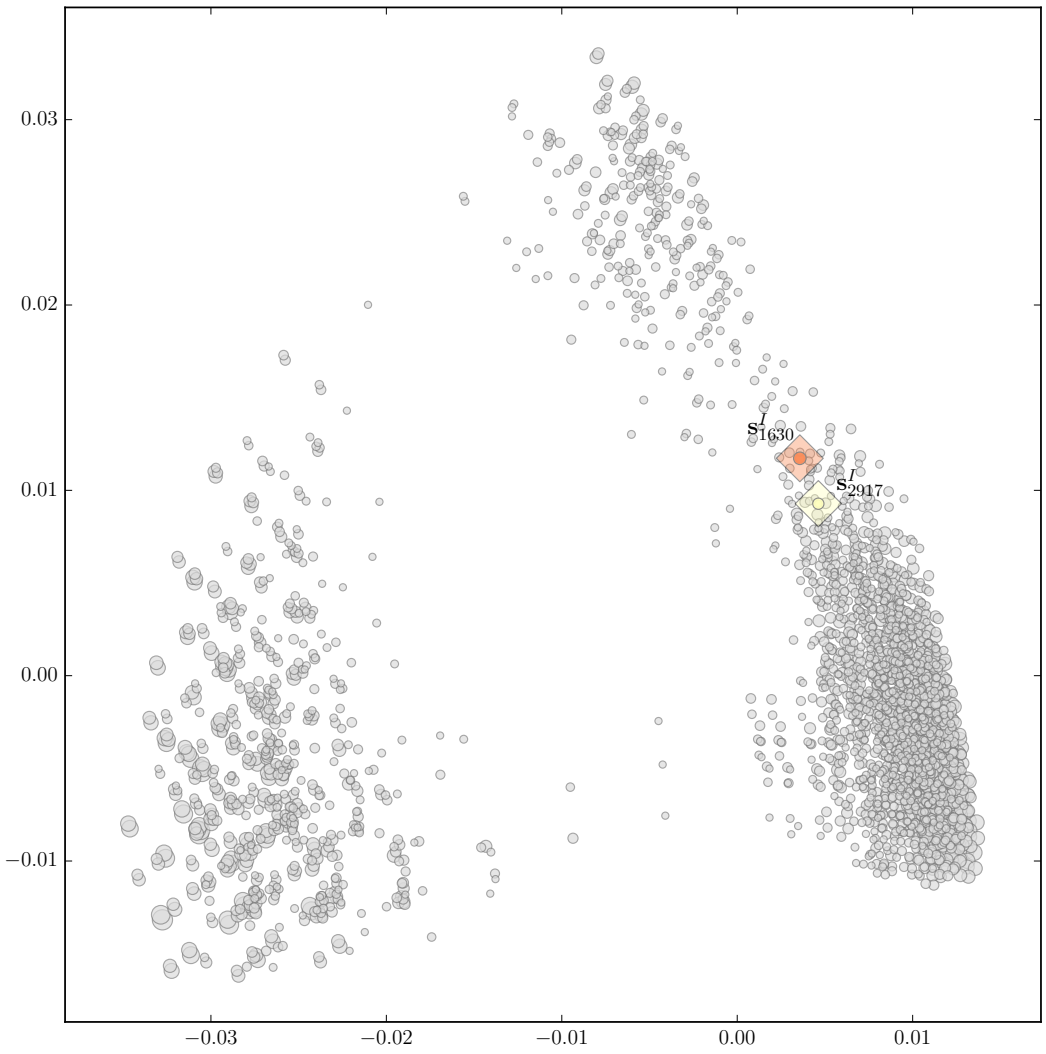
One of the key advantages of our method is its accessibility, conversely to some purely formative scenario planning methods that require a strong mathematical background to use. Although the implementation part of our method makes no difference, the usage of our method with an interactive software tool (such as Scenario Builder™, the interactive software implementation of Capful) does not impose strong background requirements for usage. The overall consistencies of the scenarios are based on simple arithmetic means of pairwise consistencies of factor levels, not requiring, e.g., vast understanding of probability calculus to comprehend. The dissimilarity assessment is based on the observation of the distances of the principal coordinates of consistent scenarios, thus utilizing one of the fundamental human perceptions, the Gestalt Law of Proximity [see, e.g., Ware (2013)]. Then, the Scenario Map that links the principal coordinates to the actual skeletal scenarios in the morphological field through color coding utilizes another fundamental human perception, the Gestalt Law of Similarity. The distinct markers used in Scenario Map and UFLM also conform to this Gestalt Law.

6.2 Limitations

We discuss the three most important limitations of the scenario identification and visualization method of this thesis in this section. First, the whole analysis is based on filling the consistency table of pairwise consistencies with numerical values. These numerical values are often ambiguous, and eliciting strict numerical value indicators of future events entails many challenges [see, e.g, O'Hagan et al. (2006)]. Moreover, filling a large consistency table can be very time consuming, which can be considered as a limitation in the efficiency of our method (third aspect in Table 2.1). For example, the consistency table of NESO case took roughly one working day to fill, and hence this task would not have been possible to be carried out by one person during a workshop. Nevertheless, some means can be taken to reduce the time spent by workshop participants on this task. In the consistency table, the pairwise consistency tables only need to be internally coherent. Consequently, a large consistency table can be filled in a workshop setting by dividing the workshop participants into groups, each of which fills one or few of the pairwise consistency tables. These groups can be organized in a way that the group members have an adequate background knowledge for making inferences on the particular uncertainty factors that each pairwise consistency table considers. If the workshop participants hesitate making the consistency assessments, the approach used in the NESO case can be taken, that is, letting the scenario planning professionals fill the table in between two workshops. Then, willing stakeholders can contribute in the consistency values by verifying and revising the values afterwards. Due to the efficiency of the algorithms developed in Section 3.2, the Scenario Map can be generated in a workshop setting once the consistency values have been revised.

The second limitation concerns the amount of inertia in the original space that can be illustrated with a 2D scatter plot. The visualization of Scenario Map has been designed specifically to alleviate this limitation, but it can still be very tempting to consider dissimilarity only through the scatter plot. Instead, the actual consistent scenarios that each principal coordinate corresponds to should always be kept clearly in mind. For example, it is possible that two mutually dissimilar scenarios are relatively close in the plot of the first two principal coordinates, e.g, if the difference of these scenarios is mostly depicted in the third principal dimension. Two such scenarios in the NESO case are presented in Figure 6.1: these two scenarios have very similar first two principal coordinates but they only share three uncertainty factor levels. These two scenarios are clearly very dissimilar, but their dissimilarity

would become visible only by exploring more principal dimensions. Thus, even though the first two principal coordinates give the best representation of the consistent scenarios in two dimensions, awareness of the inherent limitation of considering only two dimensions of a high-dimensional scenario cloud is necessary when utilizing the methods presented in this thesis.



A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4

Figure 6.1: Two scenarios of the NESO case that are close together in the scatter plot of principal coordinates but differ in the morphological field.

The third challenge of our method is that it offers no formal guidance about how to select the number I of the most consistent scenarios to include in the Scenario Map. This number should be large enough to ensure that the final scenarios that are selected among the I most consistent scenarios are able to cover most of the uncertainty factor levels. However, if I is too large, then the scatter plot of the principal coordinates of consistent scenarios can become very cluttered. Moreover, for very large values of I, the set of most consistent scenarios may become incomprehensive due to its size, even with the effective visualizations this thesis presents. In practice, a suitable value of I is to be found through trial and error.

6.3 Extensions

One approach to retaining the manageability of the set of the I most consistent scenarios for even larger values of I is to illustrate the principal coordinates of these scenarios not only by markers, but also with a Gaussian Mixture Model (GMM) density plot [Barber (2015)]. This extension, although not fully reported in this thesis, was in fact implemented in the Scenario Builder™ [Capful (2018)]. With the help of the GMM density plot, not all consistent scenarios need to be visible in the scatter plot, because the GMM plot depicts the general structure of the scenario cloud. The Scenario Map can then be explored by studying partitions of the set of the most consistent scenarios, e.g., sets $\mathcal{S}^{I_p^+}$ and $\mathcal{S}^{I_p^-}$ which contain the more consistent portion of $p \cdot I$ scenarios and less consistent portion of $(1-p) \cdot I$ scenarios, $p \in (0, 1]$.[1]. In Scenario Builder™, this approach is implemented such that the most consistent portion $\mathcal{S}^{I_p^+}$ of $p \cdot I$ scenarios can be explored with any $p \in (0, 1]$, while retaining the GMM density plot of the complete set $\mathcal{S}^I$. Moreover, the scenarios that are selected for further development will remain in the visualization, even if other scenarios of similar consistency are hidden. Figure 6.2 illustrates the GMM density plot for the NESO case with $I = 3000$ and $p = 1/6$, whereby only $p \cdot I = 500$ scenarios are covered by the scatter plot (cf. Figure 5.1). With the interactive software, the first few scenarios can be selected from a manageable set $\mathcal{S}^{I_p^+}$ of I_p^+ most consistent scenarios, after which the final set of scenarios covering most of the uncertainty factor levels can be obtained with the help of UFLM from the complete set $\mathcal{S}^I$ in the scatter plot.

[1] I.e., $\mathcal{S}^{I_p^+}, \mathcal{S}^{I_p^-} \subseteq \mathcal{S}^I$ such that $\mathcal{S}^{I_p^+} \cup \mathcal{S}^{I_p^-} = \mathcal{S}^I, \mathcal{S}^{I_p^+} \cap \mathcal{S}^{I_p^-} = \emptyset, |\mathcal{S}^{I_p^+}| = p \cdot I, |\mathcal{S}^{I_p^+}| = (1-p) \cdot I, \alpha(\mathbf{s}^{I_p^+}) \geq \alpha(\mathbf{s}^{I_p^-}) \forall \mathbf{s}^{I_p^+} \in \mathcal{S}^{I_p^+}, \mathbf{s}^{I_p^-} \in \mathcal{S}^{I_p^-}$

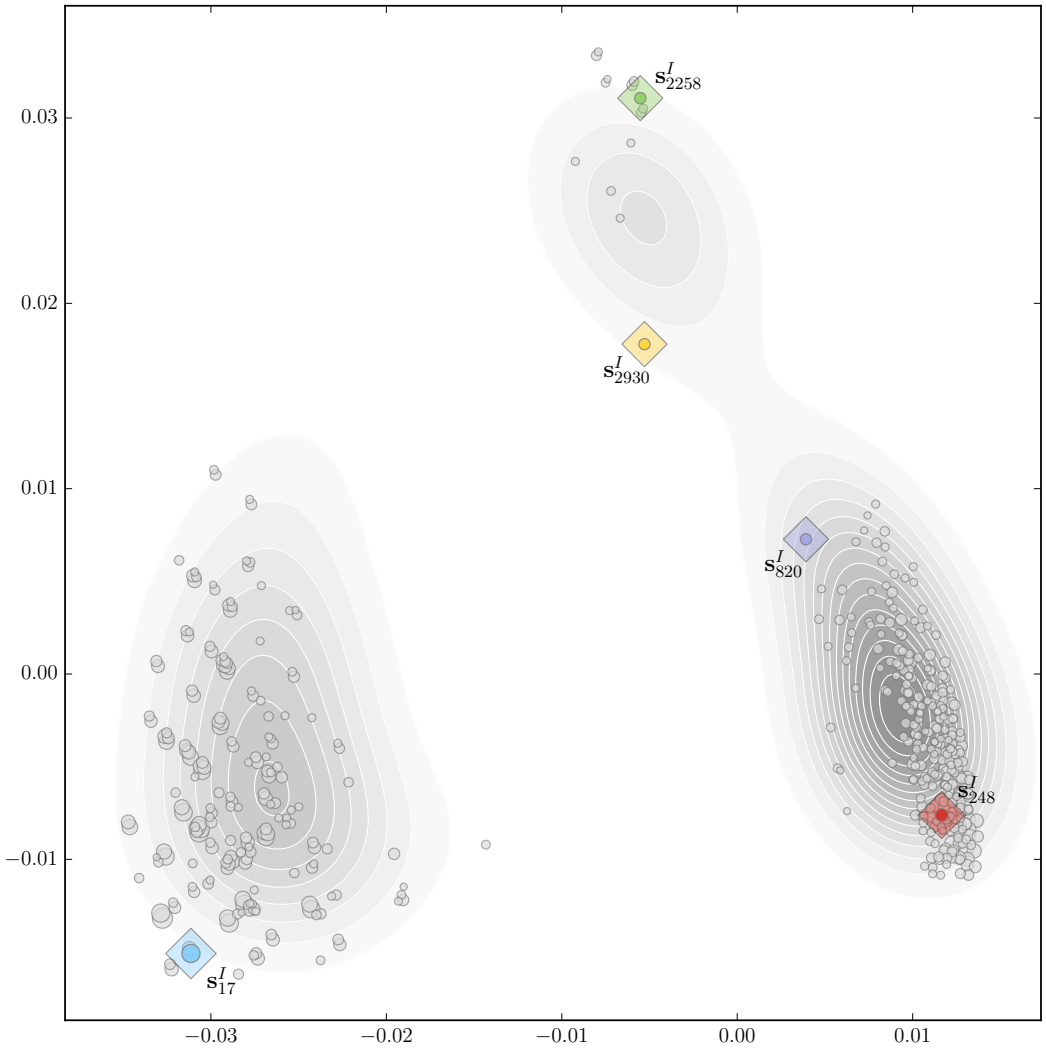
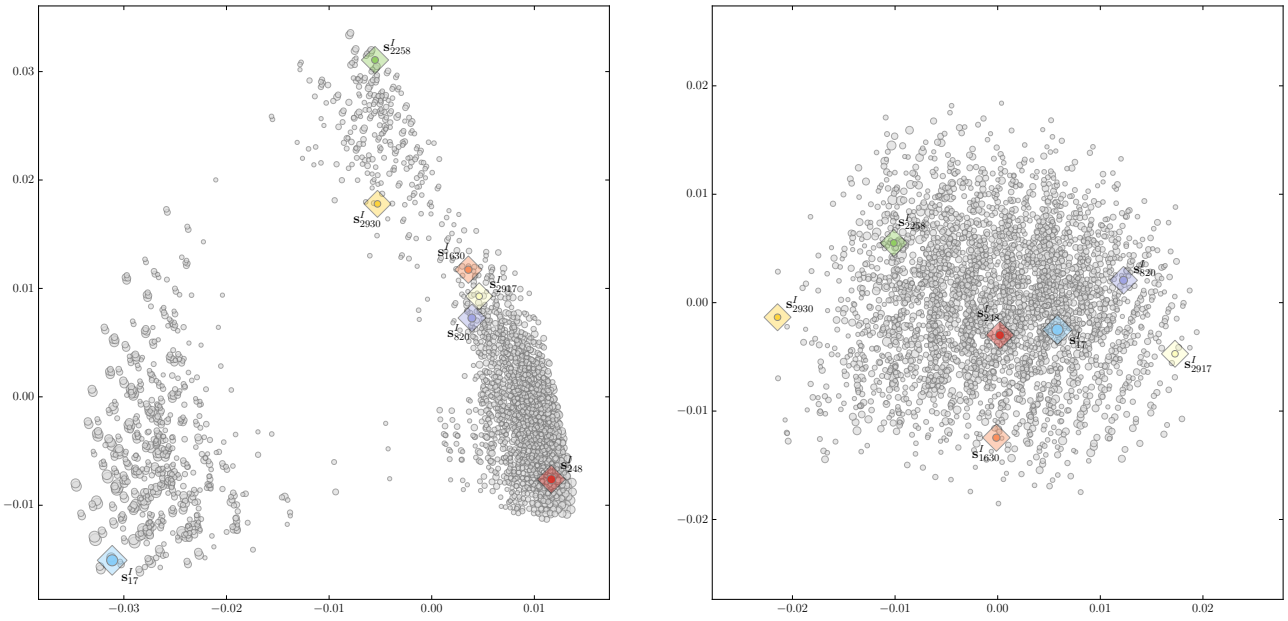


Figure 6.2: The NESO case GMM density plot with $I = 3000$ and $p = 1/6$.

To alleviate the limitations of the amount of inertia that can be explained, one obvious approach is to analyse more than two principal dimensions. A typical approach to visualize more than two dimensions using scatter plots is to use a scatter plot matrix. In scatter plot matrix, data of ρ dimensions can be visualized by creating ρ^2 scatter plots in a square table, such that the scatter plot in the r_1th column and r_2th row displays the r_1th dimension of the data on the horizontal axis and r_2th dimension on the vertical axis.

Nevertheless, because the principal dimensions by definition represent the high-dimensional point cloud in the order of descending inertia, their quality

of representation quickly deteriorates. To illustrate this effect, we present the first four principal dimensions of the NESO case in Figure 6.3 using two scatter plots. The $M = 5$ consistent scenarios are illustrated with the same color codings in the left subfigure of the first two principal principal dimensions and in the right subfigure of third (horizontal axis) and fourth (vertical axis) principal dimensions. The explained inertias of these dimensions are 21.88%, 8.86%, 5.88%, and 4.06%, respectively. Accordingly, the projected scenario cloud in principal dimensions 3 and 4 is more cluttered, thus making inferences of dissimilarity more difficult. In addition to the $M = 5$ final scenarios, the two scenarios that were close together in the scatter plot of Figure 6.1 are also illustrated in Figure 6.3, marked by orange and light yellow color codings. Even though the scenario cloud in the third and fourth principal dimension tends to be cluttered, these scenarios do differ with respect to these latter principal dimensions more than in the first two.



(a) NESO scenarios in principal dimensions 1 and 2. (b) NESO scenarios in principal dimensions 3 and 4.

A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4

Figure 6.3: Visualization of the first $\rho = 4$ principal dimensions of the NESO case scenarios using Scenario Map.

Because the method presented in this thesis is, first and foremost, a visualization method, another straightforward extension would be to augment this method by (purely) formative methods, such as those in Tietje (2005). Technically, these methods identify the set of $M = 5$ mutually dissimilar scenarios by solving the multi-objective integer optimization problem

$$\max_{\mathbf{S}^M \in \mathbb{N}^{M \times J}} \left[\boldsymbol{\alpha}(\mathbf{S}^M) \quad \mathbf{D}(\mathbf{S}^M) \right], \tag{6.1}$$

where $\mathbf{S}^M$ is a matrix with the M consistent and mutually dissimilar scenarios as rows, $\boldsymbol{\alpha}(\mathbf{S}^M)$ is the M-vector of overall consistencies of scenarios $\mathbf{S}^M$ and $\mathbf{D}(\mathbf{S}^M)$ is a $(M-1)\times(M-1)$ triangular matrix of the distances between pairs of scenarios in $\mathbf{S}^M$. Then, the solution to this optimization problem can be visualized with the methods presented in this thesis. On the other hand, the visualization method of this thesis can be used to augment virtually any type of formative scenario planning methods that solve the optimization problem of Equation (6.1). Viewing the dissimilarity of the results of such formative scenario planning methods in the comprehensive Scenario Map clearly shows the quality of the result, and Scenario Map can even be used as a validation method for such approaches. Nevertheless, as presented in this thesis, our method works as a stand-alone scenario identification method as well.

Finally, the visualization method presented in this thesis could be extended to illustrate not only the consistent scenarios, but also the uncertainty factor levels of which these scenarios consist. In fact, MCA is often used exactly for this purpose, although our method uses MCA solely to project the scenario cloud onto a two-dimensional plane. In our method, UFLM can be used to inspect the general structure of the uncertainty factor levels implicitly. However, illustrating these levels explicitly as lines or arrows would provide multiple (unorthogonal) scenario axes. Such a visualization could be considered as a link between the inductive scenario planning methods and the deductive 2x2 scenario matrix approach.

Chapter 7

Conclusions

To succeed in an unpredictable environment, companies must develop methods and strategies to understand uncertain and complex future [Vilkkumaa et al. (2018)], for which developing multiple plausible scenarios provides a robust and holistic framework [Schoemaker (1993)]. Traditional scenario development methods make tradeoffs between the explorativeness, trustworthiness, efficiency, and accessibility of the method [McBride et al. (2017)]. For example, the deductive 2x2 scenario matrix technique [Schwartz (1991)] can foster building four divergent scenarios efficiently, while restricting the exploration of the future possibility space [Wright et al. (2013), Lord et al. (2016)]. On the other hand, unstructured inductive scenario building methods are effective in their explorativeness, but may overwhelmingly depend on the creativity and imagination of the participants of the scenario intervention to the point where there is the risk of not developing scenarios in a timely manner [McBride et al. (2017)].

In this thesis, we have developed a structured method for identifying and visualizing scenarios for characterizing multiple mutually dissimilar but plausible futures. Our method is inductive in that the holistic depictions of uncertain and complex futures are built with a specific-to-general approach [McBride et al. (2017)]. Furthermore, the procedure is formative because these depictions are based on vectors of factor-specific levels explored using mathematical formulations and algorithms. Yet, the method is only weakly formative in that the visualization method of this thesis allows scenario planning practitioners and scenario development stakeholders to become involved in the scenario identification, regardless of their backgrounds in mathematical fields.

This weakly formative inductive method satisfies most of the important aspects of scenario development processes. Explorative scenarios can be developed, because the procedure does not impose a limit on the number of uncertainty factors, and because even a large cloud of consistent scenarios can be studied such that unexpected plausible combinations of factor levels are allowed to emerge. All the steps taken in our method are well reasoned and transparent, thus fostering the trustworthiness of the developed scenarios. Because the selection of dissimilar scenarios is based on the direct observation of distance, the method is very accessible as well. Although the scenario identification process is based on a possibly time consuming task of assessing all the internal consistencies between pairs of factor-specific levels, the algorithms developed in this thesis for generating consistent scenarios help efficiently produce the visualization of consistent scenarios for their interactive exploration.

The methods presented in this thesis were used to guide the scenario building project of the Finnish National Emergency Supply Organization, and hence the developed method supports the investigation of uncertain and complex future in practice. The results show that the method is reliable and produces explorative and divergent scenarios, which are viewed as trustworthy by the stakeholders of the scenario development exercise. During informal discussions in the project workshops, positive feedback from the project stakeholders included, among others, comments on the appeal of pairwise consistency assessments, because "pairwise comparisons are the most effective means of human reasoning". Moreover, the stakeholders appreciated that the view on the future is not limited by using axes that only consider two uncertainty factors.

Bibliography

Alpaydin, E. (2014). *Introduction to Machine Learning*. Cambridge, Massachussets: The MIT Press, 3rd edition.

Ayres, R. U. (1969). *Technological Forecasting and Long-range Planning*. New York: McGraw-Hill, 1st edition.

Barber, D. (2015). *Bayesian Reasoning and Machine Learning*. Cambridge: Cambridge University Press, 7th edition.

Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396.

Bowman, G. (2016). The practice of scenario planning: an analysis of inter-and-intra-organizational strategizing. *British Journal of Management*, 27(1):77–96.

Bowman, G., MacKay, R. B., Masrani, S., and McKiernan, P. (2013). Storytelling and the scenario process: Understanding success and failure. *Technological Forecasting & Social Change*, 80(4):735–748.

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

Brauers, J. and Weber, M. (1988). A new method of scenario analysis for strategic planning. *Journal of Forecasting*, 7(1):31–47.

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

Capful (2018). *Scenario Builder*. `http://capful.fi/en/services/corporate-clients/scenario-builder/`. Accessed: 2018-05-09.

Carlsen, H., Eriksson, E. A., Dreborg, K. H., Johansson, B., and Bodin, O. (2016). Systematic exploration of scenario spaces. *Foresight*, 18(1):59–75.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2017). *Introduction to Algorithms*. Cambridge, Massachussets: MIT Press, 3rd edition.

Godet, M. (2000). How to be rigorous with scenario planning? *Foresight*, 2(1):5–9.

Greenacre, M. (2017). *Correspondence Analysis in Practice*. Boca Raton: Chapman & Hall/CRC, 3rd edition.

Greenacre, M. and Blasius, J. (2006). *Multiple Correspondence Analysis and Related Methods*. Boca Raton, Florida: Chapman & Hall/CRC, 1st edition.

Hopcroft, J. and Kannan, R. (2012). *Computer Science Theory for the Information Age*. Shangahi: Shanghai Jiaotong University Press, 1st edition.

Huoltovarmuuskeskus (2016). *National Emergency Supply Organization (NESO)*. `https://www.nesa.fi/organisation/`. Accessed: 2018-05-09.

Huoltovarmuuskeskus (2018). *Huoltovarmuuden skenaariot 2030*. Accessed 2018-05-09 via: `https://www.huoltovarmuuskeskus.fi/`.

Husson, F., Le, S., and Pages, J. (2017). *Exploratory Multivariate Analysis by Example Using R*. Boca Raton, Florida: Chapman & Hall/CRC, 2nd edition.

Johansen, I. (2018). Scenario modelling with morphological analysis. *Technological Forecasting & Social Change*, 126(1):116–125.

Kaplansky, I. (1972). *Set theory and metric spaces*. Boston: Allyn and Bacon, 1st edition.

Kohavi, R. and John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324.

Leskovec, J., Rajaraman, A., and Ullman, J. D. (2014). *Mining of Massive Datasets*. Cambridge: Cambridge University Press, 1st edition.

Lord, S., Helfgott, A., and Vervoort, J. M. (2016). Choosing diverse sets of plausible scenarios in multidimensional exploratory futures techniques. *Futures*, 77(1):11–27.

Martelli, A. (2001). Scenario building and scenario planning: State of the art and prospects of evolution. *Futures Research Quarterly Summer*, 17(2):57–74.

McBride, M. F., Lambert, K. F., Huff, E. S., Theoharides, K. A., Field, P., and Thompson, J. R. (2017). Increasing the effectiveness of participatory scenario developement through codesign. *Ecology and Society*, 22(3).

Musser, D. R. (1997). Introspective sorting and selection algorithms. *Software: Practice and Experience*, 27(8):983–993.

Ogilvy, J. and Schwartz, P. (1998). *Learning From the Future*, chapter Plotting Your Scenarios, pages 57–80. New York: Wiley.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Sourthern Gate, Chichester: John Wiley & Sons, 1st edition.

Pillkahn, U. (2008). *Using Trends and Scenarios as Tools for Strategy Development*. Erlangen: Publicis Corporate Publishing, 1st edition.

Rhyne, R. (1971). *Projecting Whole-Body Future Patterns - The Field Anomaly Relaxation (FAR) Method*. Menlo Park, California: Stanford Research Institute, 1st edition.

Ritchey, T. (2006). Problem structuring using computer-aided morphological analysis. *Journal of the Operational Research Society*, 57(1):792–801.

Ritchey, T. (2011). Modeling alternative futures with general morphological analysis. *World Future Review*, 3(1):83–94.

Schoemaker, P. J. (1993). Multiple scenario development: Its conceptual and behavioural foundation. *Strategic Management Journal*, 14(3):193–213.

Scholz, R. W., Bösch, S., Carlucci, L., and Oswald, J. (1999). *Nachhaltige Regionalentwicklung: Chancen der Region Klettgau*. Zürich: Rüegger, 1st edition.

Scholz, R. W. and Tietje, O. (2001). *Embedded Case Study Methods: Integrating Quantitative and Qualitative Knowledge*. Thousand Oaks: Sage Publications, 1st edition.

Schwartz, P. (1991). *The Art of the Long View: Planning for the Future in an Uncertain World*. New York: Doubleday, 1st edition.

Strang, G. (2016). *Linear Algebra and its Applications*. Wellesley: Wellesley-Cambridge Press, 5th edition.

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.

Tietje, O. (2005). Identification of s small reliable and efficient of consistent scenarios. *European Journal of Operational Research*, 162(1):418–432.

van der Heijden, K. (1996). *Scenarios: The Art of Strategic Conversation*. Hoboken, N.J.: Wiley, 1st edition.

van Vliet, M., Kok, K., Veldkamp, A., and Sarkki, S. (2012). Structure in creativity: an exploratory study to analyse the effects of structuring tools on scenario workshop results. *Futures*, 44(8):746–760.

van Vught, F. A. (1978). Pitfalls of forecasting: Fundamental problems for the methodology of forecasting from the philosophy of science. *Futures*, 19(2):184–196.

van't Klooster, S. A. and van Asselt, M. B. A. (2006). Practicing the scenario-axes technique. *Futures*, 38(1):15–30.

Vilkkumaa, E., Liesiö, J., Salo, A., and Ilmola-Sheppard, L. (2018). Scenario-based portfolio model for building robust and proactive strategies. *European Journal of Operational Reserach*, 266(1):205–220.

Volkery, A. and Ribeiro, T. (2009). Scenario planning in public policy. *Technological Forecasting and Social Change*, 76(9):1198–1207.

Ware, C. (2013). *Information Visualization: Perception for Design*. Amsterdam: Elsevier, 3rd edition.

Wilkinson, A. and Kupers, R. (2014). *The essence of scenarios: learning from the Shell experience*. Amsterdam: Amsterdam University Press, 1st edition.

Wright, A. (2005). The role of scenario as prospective sensemaking devices. *Management Decision*, 43(1):86–101.

Wright, G., Bradfield, R., and Cairns, G. (2013). Does the intuitive logics method - and its recent enhancements - produce "effective" scenarios. *Techological Foresights and Societal Change*, 80(1):631–642.

Zwicky, F. (1948). Morphological astronomy. *Observatory*, 68(845):121–143.

Zwicky, F. (1969). *Discovery, Invention, Research - Through the Morhpological Approach*. Toronto: The Macmillan Company, 1st edition.

Appendix A

Naive algorithms

A naive way to generate the set of all possible scenarios $\mathcal{S}$ (Definition 3) utilizes nested for-loops, as presented in Algorithm 5.

Algorithm 5 Naive enumeration of the set of all possible scenarios $\mathcal{S}$.

Input: $\mathbf{K}, j$ ▷ Numbers of levels of the uncertainty factors, current recursion depth
Output: $\mathcal{S}^j$ ▷ Set all of possible scenarios in current recursion

```
1: function ScenarioSet(K, j)
2:     if j ≠ J then
3:         S^{j+1} ← ScenarioSet(K, j + 1)                ▷ Recursion
4:         S^j ← { }                                      ▷ Initialize the set of scenarios
5:         for s^{j+1} in S^{j+1} do
6:             for k_j = 1, 2, ..., K_j do
7:                 s^j ← [k_j  s^{j+1}]
8:                 S^j ← S^j ∪ {s^j}
9:             end for
10:        end for
11:    else
12:        S^J ← { }                                      ▷ Initialize S
13:        for k_J = 1, 2, ..., K_J do
14:            s^J ← [k_J]
15:            S^J ← S^J ∪ {s^J}
16:        end for
17:    end if
18:    return S^j
19: end function
```

Algorithm 6 presents a simple lookup algorithm for the evaluation of the overall consistency of a scenario $\mathbf{s} \in \mathcal{S}$. This algorithm can be successively called to evaluate the consistencies of all possible scenarios, once we have generated these scenarios $\mathcal{S}$, as presented in Algorithm 7.

Algorithm 6 Naive evaluation of overall consistency of one scenario.

Input: $\mathbf{C}, \mathbf{s}$ ▷ Consistency table, scenario vector
Output: α ▷ Overall consistency value

function ConsistencyValue($\mathbf{C}, \mathbf{s}$)
  $\alpha(\mathbf{s}) \leftarrow 0$ ▷ Initialize overall consistency
  for $j_1 = 1, 2, \ldots, J-1$ **do**
    for $j_2 = j_1 + 1, j_1 + 2 \ldots, J$ **do**
      $\alpha(\mathbf{s}) \leftarrow \alpha(\mathbf{s}) + c^{j_1 j_2}(k_{j_1}, k_{j_2})$ ▷ Accumulate overall consistency
    end for
  end for
  $\alpha(\mathbf{s}) \leftarrow \alpha(\mathbf{s}) / [J(J-1)/2]$ ▷ Scale α
  return $\alpha(\mathbf{s})$
end function

Algorithm 7 Naive evaluation of all scenarios and their overall consistencies.

Input: $\mathbf{C}, \mathbf{K}$ ▷ Consistency table, numbers of levels
Output: $\mathcal{S}, \boldsymbol{\alpha}$ ▷ All possible scenarios and their overall consistencies

function ScenariosAndConsistencies($\mathbf{C}, \mathbf{K}, \mathcal{S}$)
  $\mathcal{S} \leftarrow$ ScenarioSet($\mathbf{K}, 1$)
  $\boldsymbol{\alpha} \leftarrow [\ ]$ ▷ Initialize α
  for s in $\mathcal{S}$ **do**
    $\alpha(\mathbf{s}) \leftarrow$ ConsistencyValue($\mathbf{C}, \mathbf{s}$)
    $\boldsymbol{\alpha} \leftarrow \begin{bmatrix} \boldsymbol{\alpha}^{\mathrm{T}} & \alpha(\mathbf{s}) \end{bmatrix}^{\mathrm{T}}$ ▷ Store α
  end for
  return $\mathcal{S}, \boldsymbol{\alpha}$
end function