Advancing the use of MACBETH in health settings: A new decision support system to assist moving from qualitative judgments towards numerical value functions

Miguel Joaquim Roque Fernandes

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors: Prof. Mónica Duarte Correia de Oliveira
Eng. João Carlos Nunes Bana e Costa

Examination Committee

Chairperson: Prof. Mário Jorge Costa Gaspar da Silva
Supervisor: Prof. Mónica Duarte Correia de Oliveira
Member of the Committee: Prof. João Carlos Da Cruz Lourenço

July 2024

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I am deeply grateful to my family for their unwavering support and encouragement throughout this journey.

I extend my heartfelt thanks to my friends for their invaluable help and support.

I am profoundly thankful to my supervisors, Prof. Mónica Duarte Oliveira and Eng. João C. Bana e Costa, for their exceptional guidance and mentorship.

I thank Prof. Ricardo Mateus for his support and insightful guidance.

I also express my sincere gratitude to Prof. Carlos A. Bana e Costa for his continuous support and expert guidance during many moments of this journey.

Resumo

A metodologia MACBETH, Measuring Attractiveness by a Categorical Based Evaluation Technique, tem sido amplamente utilizada em contextos complexos para construir modelos de valor quantitativos baseados em juízos qualitativos de diferença de atratividade, elicitados em decisores. Quando os juízos qualitativos são expressos por cada decisor separadamente, sem discussão em grupo para conciliar os seus juízos, os analistas de decisão enfrentam o problema de como derivar uma função de valor coletiva a partir dos juízos individuais dos decisores.

Este trabalho propõe uma técnica original, implementada em Python, para abordar esta questão no âmbito de análise de decisão multicritério com MACBETH. Para cada critério de avaliação, a função de valor coletiva proposta segue a maioria das funções dos decisores e respeita uma propriedade de trade-off constante.

Um segundo tema foca-se na conciliação de funções de valor de diferentes decisores. Em terceiro lugar, este trabalho também examina o impacto da variabilidade nas funções de valor criadas a partir de matrizes de juízos incompletas.

Para cada desafio, foi desenvolvido um módulo num Sistema de Apoio à Decisão online e testado num caso específico.

Palavras-chave: Análise de Decisão Multicritério, Funções Parciais de Valor, MACBETH, Sistemas de Apoio à Decisão, Avaliação das Tecnologias da Saúde, Políticas Públicas de Saúde

Abstract

Standing for Measuring Attractiveness by a Categorical Based Evaluation Technique, MACBETH has been widely used in complex evaluation contexts to construct quantitative value models based on qualitative judgements of difference in value elicited from a panel of evaluators. When the qualitative judgements are expressed by each evaluator separately, with no group discussion to reconcile their judgements, decision analysts face the problem of how to derive a collective value function from the evaluators' individual judgements. This work proposes an original technique, implemented in Python, to address this issue under a MACBETH multicriteria framework. For each evaluation criterion, the proposed collective value function respects the majority rule and assumes a collective constant trade-off attitude. A second research topic focuses on the reconciliation of value functions from different decision makers. Thirdly, this work also examines the impact of variability in value functions created from incomplete matrices of judgments. For each challenge, a module was developed in a Web-based Decision Support System and tested for a case.

Keywords: Multi-criteria Decision Analysis, Multiattribute Value Theory, Partial Value Functions, MACBETH, Decision Support Systems, Health Technology Assessment, Health Policy

Contents

List of Tables

List of Figures

Acronyms

ANP	Analytic Network Process
DM	Decision-Maker
DMG	Decision-Makers Group
DSS	Decision Support System
GDP	Gross Domestic Product
GP	Goal Programming
HB-HTA	Hospital-Based Health Technology Assessment
HTA	Health Technology Assessment
LP-MACBETH	Linear Programming — MACBETH
MACBETH	Measuring Attractiveness by a Categorical-Based Evaluation Technique
MAD	Mean Absolute Deviation
MAUT	Multiattribute Utility Theory
MCDA	Multicriteria Decision Analysis
MVT	Multiattribute Value Theory
R&D	Research and Development
VPS	Virtual Private Server

Chapter 1

Introduction

Decision-making is a fundamental part of our day-to-day lives and is essential for the functioning of organizations and societies. In organizations, decision-making is a key activity with a significant impact on the success of the organization and the well-being of its members. In the healthcare context, decision-making can have a significant impact on the health of the population and the functioning of the healthcare system. It can significantly affect the quality of care provided to patients, the allocation of resources in the healthcare system, and the development of health policies.

Decision analysis is a sociotechnical process that generates insight into the consequences of decisions and helps decision-makers to make better and more informed decisions (Phillips et al., 1990; Phillips, 2005). Decision support methods are tools and techniques used in decision analysis.

Multi-criteria Decision Analysis (MCDA) is a “collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter” (Belton and Stewart, 2002).

The decision-making process can involve a facilitator/consultant who helps the decision-makers in the process of decision-making and a decision support system that assists in the analysis of the decision problem.

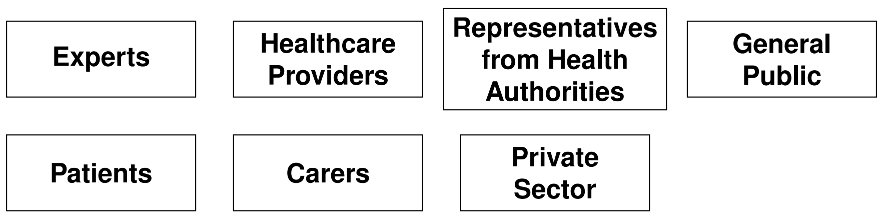


Figure 1.1: Stakeholders in healthcare decision-making, see Gongora-Salazar et al., 2022.

In the context of healthcare, decisions are ubiquitous and many times involve a plethora of stakeholders. Figure 1.1 shows the stakeholders involved in healthcare decision-making, which include patients, healthcare providers, payers, researchers, families and governments (Gongora-Salazar et al., 2022).

MCDA is used to support decision-making in several contexts in healthcare, including priority setting, clinical decision making, regulatory decisions, planning and R&D (Gongora-Salazar et al., 2022). Refer to figure 1.2 for an illustration of the different contexts of healthcare decision-making and examples of decisions made in each context.

MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) is a well-established decision-aid approach to multicriteria value measurement. It is designed to measure the attractiveness or value of options through non-numerical pairwise comparison questioning (Bana e Costa and Vansnick, 1994). MACBETH has been widely applied in various areas including healthcare, energy, etc. (Hummel et al., 2017; Costa et al., 2019).

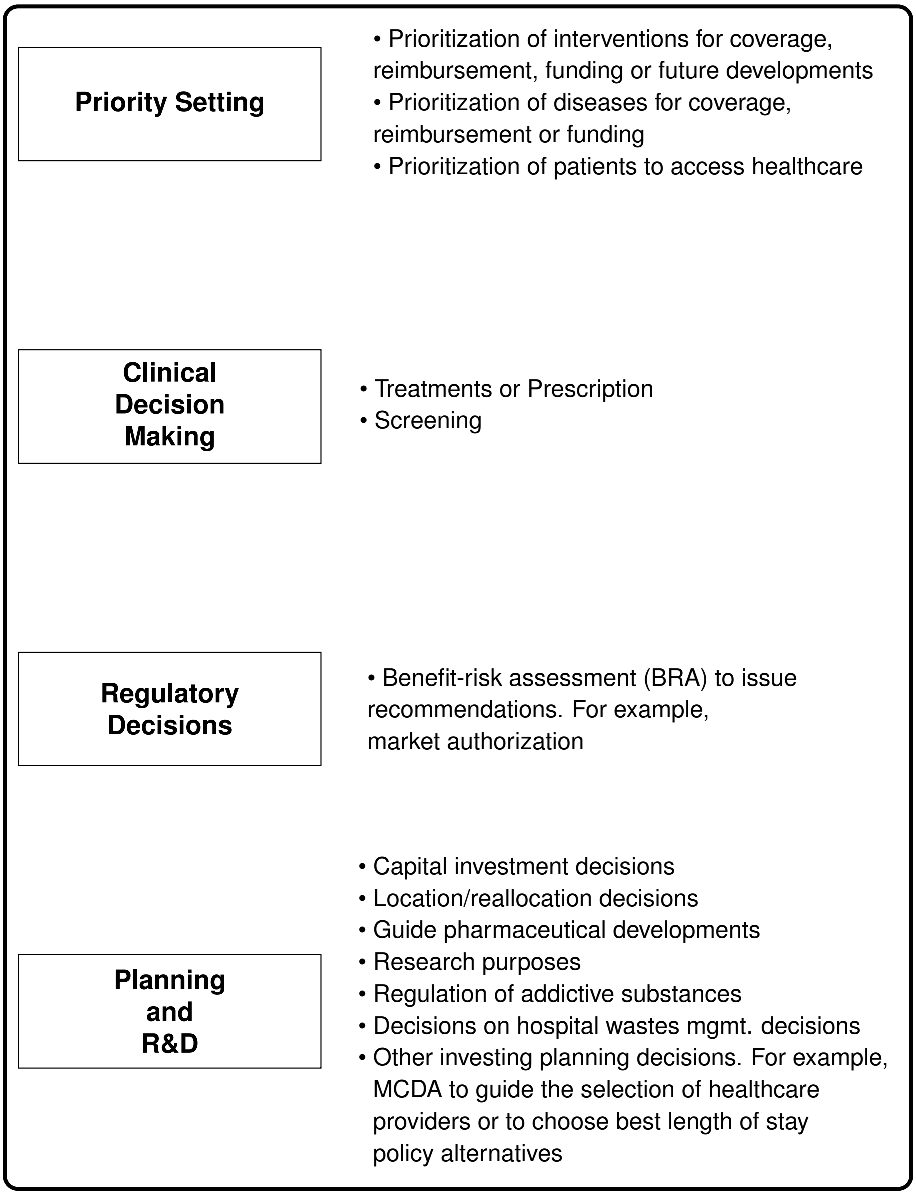


Figure 1.2: Contexts of healthcare decision-making and examples of decisions made in each context, see Gongora-Salazar et al., 2022.

1.1 Motivation

MCDA is widely used in health technology assessment and health policy (Oliveira et al., 2019). The complexity and multi-dimensionality inherent in healthcare decision-making present a significant challenge to decision-makers and policy-makers which requires the use of decision support methods such as MCDA to help make better and more informed decisions. Moreover, there has been a rising demand coupled with a scarcity of resources in healthcare systems around the world, making decisions about the allocation of resources in healthcare systems increasingly important (Angelis and Kanavos, 2017).

Interest in MACBETH has been growing in recent years (Ferreira and Santos, 2021). Yet, there is a need to promote rigour in the application of MCDA methods, especially in health settings (Oliveira et al., 2019). In particular, after the construction of several different value functions for a given criterion for each decision-maker or group of decision-makers, there is a need to aggregate these value functions into a single value function that represents the preferences of the group of decision-makers. This is a crucial step in the process and one that evokes several issues.

There are different ways to gather the preferences of stakeholders. Decision conferences promote direct communication between stakeholders aiming to define the way forward (Phillips and Bana e Costa, 2007). Other alternatives to gather the views or perspectives of stakeholders include Delphi, which can be done online.

Decision-makers in healthcare settings often have a limited understanding of the methods being used. In particular during the construction of partial value functions to measure the (partial) value of each technology or policy option on each criterion, by converting the respective performance or impact of the option into a (partial) value score. Distinguishing performance from value is crucial and requires the elicitation of value judgments and expertise with adequate techniques (Parnell, 2012).

1.2 Objective of the Dissertation

The objective of the present dissertation is to develop an intuitive tool to help decision-makers build partial value functions in health settings. This work aims to build a decision support system that assists decision-makers in moving from MACBETH qualitative judgments to curve shapes and functions, thereby fostering good practice in the area of Multi-Criteria Decision Analysis (MCDA) in Health Technology Assessment and Health Policy.

The specific objectives of the dissertation are:

- Exploring techniques for building value functions, and identifying the most appropriate methods in MCDA in a healthcare setting.
- Designing a decision support system that helps decision-makers convert MACBETH qualitative judgments into curve shapes and functions in health contexts, and integrating the selected methods for building value functions.

- Implementing the decision support system into a software program, in a web platform, to make it user-friendly and accessible to decision-makers.
- Testing the decision support system in a case study to evaluate its effectiveness in supporting decision-making in health settings.
- Analyzing and discussing the results of the case study, and providing recommendations for further improvements of the decision support system.

The expected deliverables of this work are:

- A introduction on MCDA modelling in Health Technology Assessment and Health Policy contexts, with a focus on the challenges and methodological issues related to the construction of partial value functions.
- An overview of existing techniques for building value functions, and an identification of the most appropriate methods for the health setting.
- A decision support system that helps decision-makers convert MACBETH qualitative judgments into curve shapes and functions in health contexts, integrated with the selected methods for building value functions.
- A software program (in a web platform) that implements the decision support system and is user-friendly and accessible to decision-makers.
- A case study that tests the effectiveness of the decision support system in supporting decision-making in health settings, and a report analyzing and discussing the results.
- Recommendations for further improvements of the decision support system based on the case study results.

Chapter 2

Background

In this chapter, a brief overview of Multicriteria Decision Analysis (MCDA) is provided, outlining the MCDA process and explaining its various steps. An introduction to additive models is also included, followed by a description of different MCDA methods. The MACBETH approach is presented in greater detail, emphasizing its unique features and applications.

Additionally, various methods for constructing partial value functions are discussed, highlighting the different shapes these functions can take. The chapter also addresses common challenges in the MCDA process and explores various approaches to resolving these issues. These challenges and solutions form the basis and motivation for the research conducted in this thesis.

2.1 Health Technology Assessment (HTA)

Health Technology Assessment (HTA) is a multidisciplinary field of knowledge that uses various methodologies to collect and process evidence. The purpose of HTA is to support evaluators in understanding the relative value of health technologies or health interventions in different dimensions such as social, economic, organizational, and ethical. It assesses the direct and indirect consequences of health technology use by theoretical and practice-oriented research. HTA tools have been emerging to address the challenges in the field, such as the approval and adoption of new health technologies with limited evidence on safety and effectiveness, the need for involvement of health stakeholders and the implementation of HTA findings, and issues related to the deployment of existing evaluation methods and processes (Oliveira et al., 2019).

In recent years, the significance of HTA has grown. There are various factors, such as the increase in adoption of HTA by EU countries to address challenges of policy-making that harness the benefits of technology and innovation with limited budgets and complex expectations through a comprehensive approach for priority setting and resource allocation (Sorenson et al., 2008), and the advent of precision medicine and genomic technologies (Afonso et al., 2023). These advancements have transformed the landscape of healthcare by enabling more personalized and effective treatments, especially in fields like oncology. The complexity and cost associated with these innovative technologies necessitate robust

HTA processes to ensure their value for money and to guide informed decision-making in healthcare institutions (Afonso et al., 2023).

2.2 Multicriteria Decision Analysis (MCDA)

MCDA is a method used to evaluate and compare alternatives based on multiple criteria. In the process of Multi-Criteria Decision Analysis (MCDA) modelling, several phases are involved. The first phase is *Problem Definition*, where the decision problem is clearly articulated. This is followed by *Criteria Identification*, where the relevant criteria for evaluating the alternatives are identified. The next phase is *Alternatives Generation*, where a set of potential alternatives is generated. In the subsequent phase of *Weights Assignment*, weights are assigned to the criteria to reflect their relative importance. The *Performance Evaluation* phase involves assessing the performance or value of the alternatives against each criterion. The results from the evaluation are then aggregated in the *Aggregation* phase. The next steps involve *Ranking and Sensitivity Analysis*, where the alternatives are ranked based on their aggregated performance, and sensitivity analysis is conducted to examine the robustness of the results. Finally, in the *Decision Making* phase, the final decision is made based on the results given by the model and considering the sensitivity and robustness analyses. The entire MCDA process can be divided into three overarching phases: *Structuring*, *Evaluation*, and *Model Validation* (Angelis and Kanavos, 2016) It is vital to note that the MCDA process is iterative, and decision-makers may revisit and refine the steps, even by going back several steps if necessary, to ensure a comprehensive and robust analysis.

In a context with multiple stakeholders, an approach called MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) may be used to gather judgements from each stakeholder or group of stakeholders. This scenario often occurs when a Delphi survey is conducted before a decision conference to obtain the MACBETH judgements from each stakeholder.

However, a challenge arises when combining the MACBETH judgements from different stakeholders or stakeholder groups to construct a unified value function. There is still work to be done on the subject of how to effectively merge MACBETH judgements from diverse stakeholders or groups and create a single value function.

Additionally, within the field of decision support systems, several techniques and concepts are commonly utilized, some which involve assessing the partial value of different alternatives within a decision-making process.

Curve fitting is another concept used in decision support systems, which refers to the process of finding a mathematical function that best fits a set of data points. It is often employed to develop models or predict outcomes based on existing data.

Lastly, a Decision Support System (DSS) is an interactive computer-based tool that aids decision-making processes. It provides users with relevant information, analysis, and modelling tools to facilitate decision-making tasks.

According to Beiderbeck et al., 2021, the Delphi technique is described as a scientific method used to arrange and oversee structured group communication processes. Its primary objective is to produce insights into present or future challenges.

Multi-Attribute Value Theory (MAVT) is a decision-making framework that provides a systematic approach for evaluating and comparing alternatives based on multiple attributes or criteria. It recognizes that decision-making involves considering multiple factors, each with its importance and value. MAVT allows decision-makers to quantify and analyse these attributes, assigning weights to reflect their relative importance and assessing the performance or value of alternatives against each attribute. By considering multiple criteria simultaneously, MAVT enables a more comprehensive and structured decision-making process. It helps decision-makers gain insights into trade-offs, identify the most desirable alternatives, and make informed choices that align with their objectives. MAVT is important because it provides a rigorous and transparent methodology to handle complex decision problems, especially those involving diverse criteria and conflicting objectives. It promotes a systematic evaluation of alternatives, leading to more robust and well-informed decisions in various domains such as business, engineering, healthcare, and public policy (Angelis and Kanavos, 2016; Belton and Stewart, 2002).

To access the partial value of a technology or policy option on a criterion, the performance of the option on the criterion is converted into a partial value score (Belton and Stewart, 2002). There are several methods to build partial value functions (Parnell, 2012). Although such evaluations in theory fully define the value function, they can be quite demanding on the individuals who must provide the evaluations (Belton and Stewart, 2002). Especially in healthcare settings where decision-makers often have a limited understanding of the methods being used, as mentioned in section (1.1).

2.2.1 Additive Models

The MCDA methodology often relies on additive models to calculate the total value of each option. In these models, the partial values of the options on each criterion are summed to obtain the overall value. This approach provides a clear and straightforward way to aggregate the impacts of different criteria.

The total value of an option, denoted as $V(a)$, can be calculated using the following equation:

$$V(a) = \sum_{i=1}^{n} w_i v_i(a) = w_1 v_1(a) + w_2 v_2(a) + \cdots + w_n v_n(a) \quad w \geq 0, \sum_{i=1}^{n} w_i = 1, v_i(a) \geq 0, \forall i \in 1, \ldots, n \tag{2.1}$$

In this equation, w_i represents the weight assigned to criterion i, reflecting its relative importance. The partial value of option a on criterion i is denoted as $v_i(a)$. Both the weights and partial values are non-negative. The weights should sum up to 1, indicating their proportional contribution to the total value (Belton and Stewart, 2002).

By using this additive model, the MCDA process can assign a value between 0 and 100 to each option. The value represents the attractiveness or desirability of the option, taking into account the different criteria and their impacts. A higher value indicates a more favourable option, while a lower value suggests a less desirable one.

It is important to note that the additive model assumes certain conditions for its applicability. The criteria should be exhaustive, meaning that they cover all relevant aspects of the decision problem. They should also be non-redundant, meaning that no criterion duplicates or overlaps with others. Additionally, the criteria should be preference-independent, meaning that the value assigned to an improvement in one criterion does not depend on the performance of other criteria.

The use of additive models in MCDA has been widely studied and applied in various fields, including transport networks, immigration, education, investment, environment, energy, defense, and health settings. It provides a transparent and systematic approach to decision-making by considering multiple objectives and their relative importance.

2.2.2 Multiattribute Utility Theory (MAUT)

Multiattribute Utility Theory (MAUT) is a decision-making framework that can be used to evaluate and compare options based on multiple attributes. It is a generalization of the expected utility theory, which is a decision-making framework that evaluates options based on a single attribute. Its objective is to model and represent the decision maker's preferential system into a utility or value function. This utility function is defined on the criteria space. MAUT allows decision-makers to consider their preferences in the form of multiple attribute utility functions and is mostly used to capture the uncertainty relating to the outcomes of alternatives rather than the uncertainty relating to attribute values (Greco et al., 2016). MAUT can be used to model complex decision problems involving multiple attributes and options, enabling a systematic and rigorous evaluation process.

According to Belton and Stewart, 2002 and Greco et al., 2016, several value function methods can be used to build partial value functions. Let us now focus on those that respect Multiattribute Value Theory (MAVT):

- Definition of a partial value function: This method relates value to performance in terms of a measurable attribute reflecting the criterion of interest. For example, the number of flights per week could be a measurable attribute in a certain context.
- Construction of a qualitative value scale: In this method, the performance of alternatives can be assessed by reference to descriptive pointers, or word models (to which appropriate values are assigned).
- Direct rating of the alternatives: In this case, no attempt is made to define a scale that characterises performance independently of the alternatives being evaluated. The decision maker simply specifies a number or identifies the position on a visual analogue scale, which reflects the value of an alternative in relation to the specified reference points.
- Direct rating techniques: In which values of v_i are directly assessed with reference to two arbitrarily chosen points.
- Procedures based on bisection: The decision-maker is asked to assess a point that is "halfway" in terms of preference between two reference points.
- Procedures trying to build standard sequences: On each attribute in terms of preference.

2.2.3 Multi-criteria Decision Analysis Methods

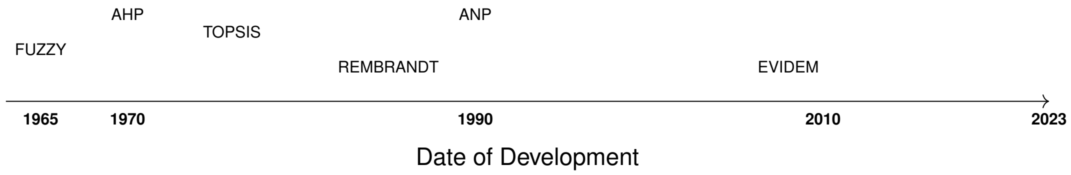


Figure 2.1: Techniques for Multi-Criteria Decision Analysis (Greco et al., 2016)

Among the various methodologies for Multi-Criteria Decision Analysis (MCDA), outranking methods, including ELECTRE (ELimination Et Choix Traduisant la REalité), PROMETHEE (Preference Ranking Organization METHod for Enrichment Evaluations), QUALIFLEX, REGIME, ORESTE, ARGUS, EVAMIX, TACTIC, MELCHIOR, MAPPAC, PRAGMA, IDRA, and PACMAN, aim to compare alternatives by identifying and prioritizing the most preferred options according to a set of criteria (Greco et al., 2016). These methods take into account the relative importance or preference of each criterion, allowing for a comprehensive evaluation of the alternatives.

Analytic Hierarchy Process (AHP) and Analytic Network Process (ANP) are decision-making techniques that involve structuring the problem into a hierarchy or network of criteria and alternatives and then obtaining priority weights for each element through pair-wise comparisons (Saaty, 1977, 2005). REMBRANDT is a known alternative to the Analytic Hierarchy Process (AHP). It uses a direct rating system on a logarithmic scale to replace the 1 – 9 scale of AHP. Instead of the eigenvector-based synthesis approach used in AHP, REMBRANDT uses a method based on the geometric mean to identify estimated weights and scores from pairwise comparison matrices (for Communities and Local Government: London, 2019). These methods provide a systematic and transparent approach to capture preferences in complex decision situations, allowing for the incorporation of both qualitative and quantitative information.

Goal Programming (GP) is an optimization method that seeks to minimize the deviations between the desired goals and the actual outcomes of the alternatives (Charnes, 1961). By incorporating multiple objectives into a single mathematical programming model, GP provides a powerful tool for decision-makers to balance conflicting criteria and find the most effective solution in resource allocation or planning contexts.

Fuzzy Approaches, such as Fuzzy Multi-criteria Optimization, address the inherent uncertainty and vagueness often present in real-world decision problems. By employing fuzzy set theory and allowing for the use of linguistic variables, these approaches enable the modelling of imprecise or ambiguous information, leading to more realistic and flexible decision support systems.

Multiattribute Value Theory (MAVT) is a family of methods that encompass various elicitation techniques for structuring and quantifying values or preferences in multi-criteria decision-making (Keeney and Raiffa, 1976). Based on establishing partial value functions for each criterion, these techniques, including Direct-rating, Rating Differences, Bisection, and MACBETH, allow for the construction of preference scales that aid in evaluating and comparing the performance levels of alternatives (Bana e Costa and Oliveira, 2012; Winterfeldt and Edwards, 1986).

All of these methods have their advantages and disadvantages. However, there are some good reasons to favor MAVT methods, such as MACBETH. First, MAVT methods have a strong theoretical basis and are well understood (Oliveira et al., 2019). MAVT methods respect the axioms of utility theory as they are based on the assumptions of complete and transitive preferences (Angelis and Kanavos, 2016). On the other hand, outranking methods do not adhere to the axioms of utility. They were proposed by B. Roy as an alternative to utility-based methods, focusing on building realistic enrichments of the dominance relation instead of relying on utility functions (Greco et al., 2016).

The MAVT methods framework includes defining objectives, selecting criteria, scoring options, and assigning weights to the selected criteria, making it a versatile and reliable approach for decision-making processes (Angelis and Kanavos, 2016).

In Health Technology Assessment (HTA), the predominant methods used include participative methods (85% of studies), face-to-face approaches for model-building such as decision conferences and workshops (70%), web-based formats (15%), questionnaires/surveys (22%), interviews (4%), and Delphi processes (4%) (Oliveira et al., 2019).

2.3 MCDA in HTA

Multicriteria Decision Analysis (MCDA) is increasingly recognized as a valuable tool within Health Technology Assessment (HTA). MCDA aids in evaluating health technologies by allowing for the comparison of multiple criteria, thereby enhancing transparency, efficiency, and objectivity in decision-making processes. The integration of MCDA into HTA frameworks addresses several challenges, such as incorporating diverse stakeholder perspectives and managing the complexity of evaluating health technologies with multiple, often conflicting, criteria.

In hospital-based HTA, the application of MCDA has been particularly beneficial. For example, a recent study utilized MCDA to evaluate genomic testing strategies for acute myeloid leukemia (AML) at the Instituto Português de Oncologia de Lisboa Francisco Gentil (IPO Lisboa). This approach involved mapping clinical pathways, assessing multicriteria value, and evaluating the cost of each strategy. The study highlighted the importance of stakeholder involvement and the use of comprehensive modelling techniques to produce relevant recommendations for decision-makers. This method not only facilitated a thorough assessment of the genomic testing strategies but also contributed to the harmonization of evaluation methods and tools in hospital settings (Afonso et al., 2023).

Furthermore, the use of MCDA in HTA can streamline the decision-making process by providing a clear framework for comparing the value-for-money of different health technologies. This is particularly important in the context of genomic technologies, where the complexity and rapid evolution of the field necessitate robust and adaptable evaluation methods. The incorporation of MCDA in HTA processes supports the development of more nuanced and context-specific assessments, ultimately leading to better-informed healthcare decisions (Afonso et al., 2023).

In addition to its application in hospital-based HTA, MCDA has been used to develop composite indices for population health assessment. The EURO-HEALTHY project, a collaborative initiative involving multiple European countries, aimed to create a Population Health Index (PHI) to evaluate and compare the health status of different regions. By combining qualitative value judgements from diverse stakeholders, the project developed a comprehensive index that captured various dimensions of population health, including social determinants, lifestyle factors, and healthcare services. The PHI model provided a holistic view of population health and enabled policymakers to identify areas for improvement and prioritize interventions based on the index results. This collaborative approach to index development exemplifies the value of MCDA in synthesizing complex data and stakeholder input to support evidence-based decision-making in public health (Bana e Costa et al., 2023b).

2.4 MACBETH

MACBETH (Measuring Attractiveness by a Categorical-Based Evaluation Technique) is a multi-criteria decision-aiding method created by Prof. Carlos Bana e Costa and Prof. Jean-Claude Vansnick (Bana e Costa and Vansnick, 1994). MACBETH uses pairwise comparisons between alternatives and criteria to build a set of ratio scales for the criteria and to calculate the weights of the criteria. The method then combines the weights with the partial value scores of the alternatives on each criterion to obtain a total value score for each alternative (Bana e Costa and Vansnick, 1994).

2.5 Research Objectives

During the following chapters, the work will focus on three main objectives:

1. Research how to fit a single value function defined by a set of points to a continuous function.
2. Research how to aggregate multiple value functions from different decision-makers into a single value function.
3. Research the variability of value functions created by completions of an incomplete matrix of MACBETH judgements.

1. Research how to fit a single value function defined by a set of points to a continuous function

Regarding fitting a single value function defined by a set of points to a continuous function, the goal was to develop a robust methodology to transform from discrete value functions into a continuous value function. This process is critical in applications where a single expression of value is needed for consistency and precision in decision-making.

To achieve this, the following steps were undertaken:

- Selection of Function Type: Identify the appropriate type of function to fit the data. Given the properties of the data, a special family of exponential functions was selected due to its ability to model a wide range of behaviors and shapes while having the constant trade-off attitude property.

- Curve Fitting Techniques: Employ non-linear regression models to fit the exponential curve to the collected MACBETH scores.

- Validation and Refinement: Validate the fitted curve by comparing it with traditional piecewise models. This step ensured that the continuous function not only adhered to the delta property but also provided a good representation of the value scores compared to piecewise representations.

- Implementation in Decision Support Systems (DSS): Integrate the tool of fitting continuous value function into the proposed web DSS to enhance decision-making processes. This involved creating user-friendly interfaces where users could input data and visualize the fitted curve.

2. Research how to aggregate multiple value functions from different decision-makers into a single value function

The second objective focused on reconciling value functions from different decision-makers. In collaborative decision-making scenarios, it is common for individuals to have differing opinions and judgements, resulting in multiple value functions that need to be aggregated into a single, consensus value function.

To address this, the following methodology was developed:

- Aggregation Techniques: Investigate various aggregation techniques to combine multiple value functions into a single function.

3. Research the variability of value functions created by completions of an incomplete matrix of MACBETH judgements

The third research objective examined the variability and uncertainty associated with value functions derived from incomplete matrices of MACBETH judgements. In practical applications, it is often challenging to obtain complete data, resulting in incomplete judgement matrices that can affect the stability of the derived value functions.

The following steps were undertaken to address this challenge:

- Initial exploration: Generate multiple completions of incomplete matrices. This helped in understanding the range of possible outcomes and the variability associated with different completions.

- Implementation of LP-MACBETH: Implement LP-MACBETH on Python to be able to create value functions from matrices of judgements.

- Uncertainty Quantification: Research possible statistical metrics to quantify the uncertainty and instability of the value functions derived from incomplete data. Metrics such as variance and mean absolute deviation were explored to provide insights into the reliability and robustness of the value functions.

- Implementation in DSS: Integrate these metrics into the proposed web DSS to provide users with insights into the reliability and robustness of the value functions. This allowed decision-makers to account for uncertainty and make more informed decisions despite incomplete data.

Chapter 3

Literature Review

3.1 MACBETH

The MACBETH method is based on Multiattribute Value Theory (MAVT) and provides a flexible and robust framework for decision-making in complex and uncertain environments (Bana e Costa and Vansnick, 1994). Partial value functions are used to convert the performance of alternatives on each criterion into partial value scores (Belton and Stewart, 2002). These functions can take various forms, including linear, logarithmic, and S-shaped curves (Bana e Costa et al., 2023a). The construction of partial value functions can be based on expert judgements or empirical data, such as clinical outcomes or patient preferences (Beyer and Fasolo). MACBETH has been applied in various health settings, such as drug development, health technology assessment, and resource allocation. It has been shown to provide transparent and coherent decision-making processes that take into account multiple and conflicting criteria, such as efficacy, safety, cost-effectiveness, and ethical considerations. However, there are some challenges associated with the use of MACBETH in health settings, such as the need for a clear and explicit definition of the decision problem, the selection of appropriate criteria, and the elicitation of reliable and valid judgements from decision-makers and stakeholders. Additionally, the interpretation and communication of the results of MACBETH can be challenging, especially when dealing with complex and uncertain decision problems (Bana e Costa et al., 2023a). Despite these challenges, MACBETH offers a structured and systematic approach to decision-making that can help healthcare organizations and policymakers make informed and evidence-based decisions in a wide range of health-related contexts.

3.1.1 The MACBETH Approach

The matrix of MACBETH judgements is a matrix where the rows and columns represent the alternatives, and the cells represent the difference in attractiveness between the alternatives.

Inconsistent judgements can arise when the decision-maker provides contradictory judgements in the matrix. Matrices containing inconsistent judgements are inconsistent matrices and cannot be used to create a value function using LP-MACBETH.

3.1.2 questioning Protocol

To build a matrix of MACBETH judgements, the decision-maker is asked to compare the attractiveness of two alternatives at a time and give the difference in attractiveness that fits one of the seven semantic categories: no difference, very weak, weak, moderate, strong, very strong, or extreme.

In practice, this is normally done in person or through a questionnaire.

One of the most common ways to do this is by starting by asking the pairwise judgements corresponding to the matrix cells in the last column and following with the matrix cells in the diagonal.

For example, consider the following matrix of judgements and the corresponding questioning protocol example:

	11	10	9	8	7
11	no	very weak	weak	strong	extreme
10		no	weak	moderate	v. strong
9			no	weak	strong
8				no	moderate
7					no

Consistent judgements

Figure 3.1: Example of a MACBETH matrix of judgements.

In this example, the criterion is the printing speed of a printer, and the alternatives are rated on a scale from 7 to 11.

To build the matrix of MACBETH judgements, the facilitator starts by asking the decision-maker to compare the attractiveness of the elements in the last column. The questioning protocol would be as follows. The protocol is presented in question-answer format, where the facilitator asks the question, and the decision-maker provides the answer. The facilitator then records the answer in the corresponding cell of the matrix.

Questioning Protocol Example for Last Column:

- Facilitator: What is the difference in attractiveness between a printing speed of 11 and 7 pages per minute?
  - Decision-maker: The difference in attractiveness between a printing speed of 11 and 7 pages per minute is extreme.
- Facilitator: What is the difference in attractiveness between a printing speed of 10 and 7 pages per minute?
  - Decision-maker: The difference in attractiveness between a printing speed of 10 and 7 pages per minute is very strong.

- Facilitator: What is the difference in attractiveness between a printing speed of 9 and 7 pages per minute?
  - Decision-maker: The difference in attractiveness between a printing speed of 9 and 7 pages per minute is strong.
- Facilitator: What is the difference in attractiveness between a printing speed of 8 and 7 pages per minute?
  - Decision-maker: The difference in attractiveness between a printing speed of 8 and 7 pages per minute is moderate.

After completing the judgements in the last column, the facilitator can proceed to the diagonal cells. The questioning protocol for the diagonal cells would be as follows:

Questioning Protocol Example for Diagonal Cells:

- Facilitator: What is the difference in attractiveness between a printing speed of 11 and 10 pages per minute?
  - Decision-maker: The difference in attractiveness between a printing speed of 11 and 10 pages per minute is very weak.
- Facilitator: What is the difference in attractiveness between a printing speed of 10 and 9 pages per minute?
  - Decision-maker: The difference in attractiveness between a printing speed of 10 and 9 pages per minute is weak.
- Facilitator: What is the difference in attractiveness between a printing speed of 9 and 8 pages per minute?
  - Decision-maker: The difference in attractiveness between a printing speed of 9 and 8 pages per minute is weak.

Finally, the facilitator can proceed to the remaining cells in the matrix. The questioning protocol for the remaining cells would be as follows:

Questioning Protocol Example for Remaining Cells:

- Facilitator: What is the difference in attractiveness between a printing speed of 11 and 9 pages per minute?
  - Decision-maker: The difference in attractiveness between a printing speed of 11 and 9 pages per minute is weak.
- Facilitator: What is the difference in attractiveness between a printing speed of 10 and 8 pages per minute?

  - Decision-maker: The difference in attractiveness between a printing speed of 10 and 8 pages per minute is moderate.

- Facilitator: What is the difference in attractiveness between a printing speed of 11 and 8 pages per minute?

  - Decision-maker: The difference in attractiveness between a printing speed of 11 and 8 pages per minute is strong.

By following this questioning protocol, the facilitator can systematically elicit the decision-maker's judgements and build a matrix of MACBETH judgements with those judgements.

3.1.3 LP-MACBETH

LP-MACBETH (Linear Programming — MACBETH) is the numerical implementation of the MACBETH method as a linear programming problem. In the field of Multiple Criteria Decision Analysis (MCDA), the MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) approach has gained recognition for its ability to handle imprecise assessments and preference information. MACBETH provides a structured framework for eliciting and incorporating the values and preferences of the decision-makers into the decision-making process (Bana e Costa and Oliveira, 2012).

MACBETH focuses on the classification of differences in value increments between performance levels into predefined importance classes. These classes represent the strength of preference for the corresponding value increments. The process involves pairwise comparisons of categories, where decision-makers assess the strength of the increase in attractiveness realized by moving from one level to another on a specific criterion.

A linear programming model is employed to test the consistency of the preference information provided and to find consistent values for the value differences. The model maximizes a minimum slack on the desired inequalities, seeking a solution that satisfies the constraints.

The MACBETH approach offers a practical and systematic method for capturing imprecise assessments of the decision-makers and incorporating them into comprehensive decision analysis. Its numerical implementation provides a robust and structured framework for handling imprecision and ambiguity in decision-making processes, enhancing the reliability and transparency of MCDA applications.

Further details on the implementation of MACBETH can be found in the seminal work by Bana e Costa and Vansnick, 1994 and the comprehensive textbook on Multiple Criteria Decision Analysis by Belton and Stewart, 2002.

The LP-MACBETH has the objective of minimizing the difference between the value of the best alternative and the value of the worst or neutral alternative, subject to constraints that ensure the consistency of the judgements. The constraints are as follows:

- The value of the worst or neutral alternative is set to zero. This is an arbitrary assignment.

- The value difference between two alternatives is zero if they are considered indifferent. Namely, the difference between the value of the same alternative is zero.

- The value difference between two alternatives is greater than or equal to the difference in importance classes if they are considered different.

- The value difference between two alternatives is greater than or equal to the difference between two other alternatives plus the difference in importance classes if they are considered different.

LP-MACBETH

$$
\begin{aligned}
&Min[v(x^+) - v(x^-)] \\
&\text{Subject to} \\
&v(x^-) = 0 \text{ (arbitrary assignment)} \\
&v(x) - v(y) = 0, \forall (x, y) \in C_0 \text{ (indifference)} \\
&v(x) - v(y) \geq i, \forall (x, y) \in C_i \cup \ldots \cup C_s \text{ with } i, s \in \{1, 2, 3, 4, 5, 6\} \text{ and } i \leq s \\
&v(x) - v(y) \geq v(w) - v(z) + i - s', \forall (x, y) \in C_i \cup \ldots \cup C_s \text{ and } \forall w, z \in C_{i'} \cup \ldots \cup C_{s'}, \\
&\text{ with } i, s, i', s' \in \{1, 2, 3, 4, 5, 6\}, i \leq s, i' \leq s' \text{ and } i > s'
\end{aligned}
\tag{3.1}
$$

3.1.4 M-MACBETH

M-MACBETH is a software that deploys the MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) method (Bana e Costa et al., 2017). It is an interactive approach that requires only qualitative judgements about differences in values to help decision-makers quantify the relative attractiveness of options. The user's qualitative preference judgement is captured through an interactive questioning procedure that compares two alternatives at a time. Using mathematical programming, the consistency of judgement is automatically verified, and a numerical scale is generated based on seven semantic categories: no, very weak, weak, moderate, strong, very strong, and extreme difference of attractiveness. Weighting scales for decision criteria are generated similarly, and an overall score for each option is calculated (Greco et al., 2016; Bana e Costa et al., 2013). An example of the M-MACBETH software interface is shown in Figure 3.2.

A feature of M-MACBETH is the ability to detect inconsistencies. If inconsistencies are detected in the judgements, the software itself provides suggestions to resolve these inconsistencies. These suggestions are then validated by the decision-makers (DMs).

The DMs can also perform sensitivity analysis to evaluate the impact of the missing values on the overall evaluation results. By systematically varying the missing values and observing the changes in the outcomes, decision-makers might gain insights into the robustness of their evaluations and identify potential uncertainties associated with the missing information (Greco et al., 2016; Oliveira et al., 2019).

3.1.5 MACBETH model building with multiple decision-makers

The MACBETH method can be used to build value functions with multiple decision-makers.

There can be different degrees of interaction between the decision-makers, ranging from individual judgements to group discussions and consensus-building exercises. The goal is to capture the diverse

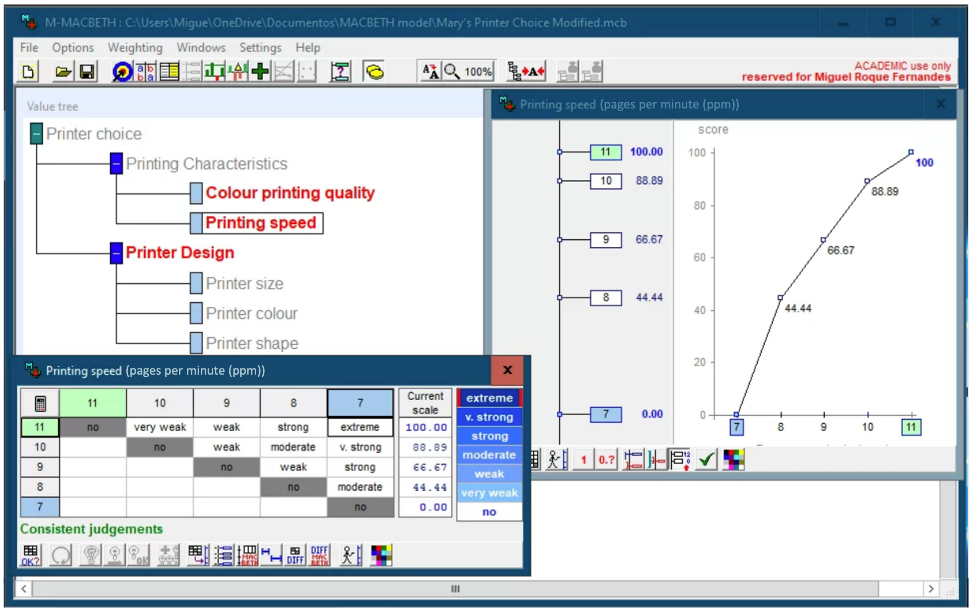


Figure 3.2: M-MACBETH software interface. The background window shows the value tree which includes the criteria. The windows on top show the matrix of judgements for a criterion and the value function for the criterion.

perspectives and preferences of the decision-makers and integrate them into a coherent and robust model.

In decision conferences, multiple decision-makers are brought together to build the model. Most often, there are four different stages in a meeting: an exploratory part where the issues are discussed, a structuring and building part where the model is built, a part where the model is explored, and a part where an agreement is reached on the way forward (Phillips and Bana e Costa, 2007).

MACBETH voting can be used to reach a consensus on each judgement. In this methodology, judgements are elicited from each decision-maker, and the group discusses the rationale behind them with the aim of creating a collective value function (Mateus et al., 2017).

In Delphi processes, decision-makers provide their judgements independently. The Delphi process can be iterative, with decision-makers revising their choices based on feedback (Trevelyan and Robinson, 2015).

3.2 Value Functions

Value Functions are used to convert the performance of an option on a criterion into a partial value score. They can be constructed using expert judgements or empirical data. See Figure 3.3 for an example of the visual representation of a value function shown by M-MACBETH. The corresponding mathematical equation that defines the piecewise linear value function is shown below.

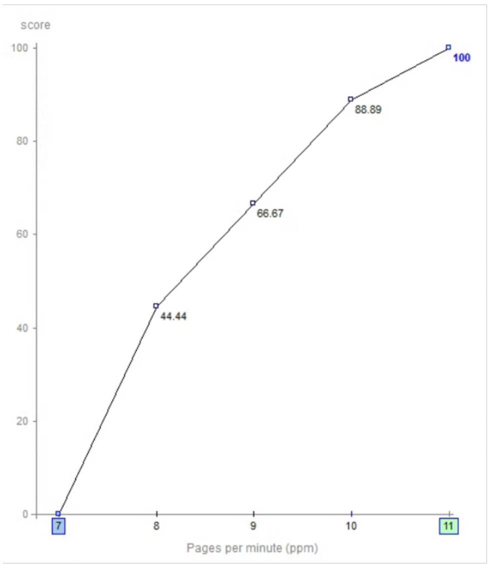


Figure 3.3: Example of a value function defined in a piecewise linear manner represented on a plane defined by a performance axis (Pages per minute (ppm)) and a score axis (Partial value score).

$$v_{value}(x) = \begin{cases} 44.44(x-7), & 7 \leq x < 8 \\ 22.23(x-8)+44.44, & 8 \leq x < 9 \\ 22.22(x-9)+66.67, & 9 \leq x < 10 \\ 11.11(x-10)+88.89, & 10 \leq x \leq 11 \end{cases} \tag{3.2}$$

3.2.1 Types of Value Functions

The descriptors of performance of the options on a criterion can be of different types and can be classified according to three aspects: their relationship with the criterion, type of data, and type of scale:

- Relationship with the criterion: direct (natural), indirect (proxy) or constructed;
- Type of data: quantitative or qualitative (which can also be pictorial);
- Type of scale: continuous (interval or ratio), discrete (nominal or ordinal).

Relationship with the criterion

The relationship with the criterion can be classified into three categories: direct, indirect, and constructed descriptors. Direct descriptors directly reflect the effects associated with a criterion. Indirect descriptors indicate causes rather than immediate effects related to a criterion. Constructed Descriptors are descriptors developed to fit the specific decision context and are typically needed when the criterion is subjective, intangible, or involves multiple interconnected aspects. See Figure 3.4 for examples of direct, indirect, and constructed descriptors.

Direct (Natural) Descriptor

Example:
"number of people affected by respiratory diseases"

Indirect (Proxy) Descriptor

Example:
"degree of concentration of pollutants that cause respiratory diseases"

Constructed Descriptor

- One-dimension qualitative scales
finite sets of verbal descriptions of one dimension plausible impacts
- Multidimensional descriptors
sets of multidimensional impact levels
- Pictorial descriptors
finite set of reference visual representations of impacts
Example: sketches, drawings, pictures, videos, computer simulations...

Figure 3.4: Examples of direct, indirect, and constructed descriptors (Bana e Costa et al., 2022).

Type of data

The type of data refers to whether the descriptor is quantitative or qualitative. Quantitative descriptors are numerical and can be continuous or discrete. Qualitative descriptors are non-numerical and can be pictorial or textual. See Figure 3.5 for examples of quantitative and qualitative descriptors.

Quantitative

Example:
"The patient's blood pressure is 120/80 mmHg."

Qualitative

Example:
"The patient's blood pressure is normal."

Figure 3.5: Examples of quantitative and qualitative descriptors

Type of scale

Value scales can be classified into two main categories: continuous and discrete. Continuous scales are used to represent value functions that are defined over a continuous domain. Discrete scales are used to represent value functions that are defined over a discrete domain.

After gathering the criteria and the alternatives, the decision-maker is asked to provide judgements on the differences in attractiveness between successive alternatives along a specific descriptor. These semantic judgements, ranging from very weak to extreme, can be translated into a value function to represent the preferences of the decision-makers.

Construction of value functions from MACBETH judgements

To explain how MACBETH semantic judgements are translated into a value function, let's consider an example where the descriptor is *Life Expectancy at birth (years)*. Bana e Costa et al., 2023a

The decision maker is asked to provide judgements on the attractiveness or preference between these alternatives using the MACBETH scale (no, very weak, weak, moderate, strong, very strong, extreme). The semantic judgements are captured by comparing the alternatives pairwise. The questions are asked to the decision-maker in a similar way as explained in 3.1.2: "What is the difference in attractiveness between a life expectancy of 90 and 80?" The decision-maker then provides a semantic MACBETH judgement. The same question is asked for the other pairs of alternatives. Here's an example of how the decision-maker may provide judgements, that are also shown in Figure 3.6:

- 90 years of life expectancy at birth is chosen as the best alternative
- 70 years of life expectancy at birth is chosen as the neutral alternative
- The difference in attractiveness in life expectancy at birth between 90 years and 70 years is classified by the decision-maker as moderate
- The difference in attractiveness in life expectancy at birth between 80 years and 70 years is classified by the decision-maker as very weak
- The difference in attractiveness in life expectancy at birth between 90 years and 80 years is classified by the decision-maker as weak

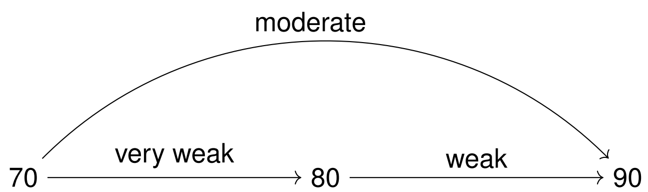


Figure 3.6: Example of semantic judgements for the Life Expectancy at birth (years) descriptor.

We can represent this information with a matrix of judgements as shown in Table 3.7.

Note:

- The cells represent the semantic judgements between the alternatives.
- The diagonal cells are empty because the alternatives are not being compared to themselves.
- The matrix is symmetric, so we only need to fill cells above the diagonal.
- The matrix is not necessarily complete. The decision-maker may not be able to provide a judgement for every pair of alternatives.

Based on these judgements, we can construct a value function that represents the preferences of the decision-makers. The value function describes how the partial values assigned to each alternative vary across the descriptor range. The function for this example is shown in Figure 3.8. The next step is to validate the value function with the decision-maker.

	70	*80*	*90*
70	—	very weak	moderate
80		—	weak
90			—

Figure 3.7: Example of a MACBETH matrix of judgements for the Life Expectancy at birth (years) descriptor.

Figure 3.8: Example of a value function for the Quality of Service descriptor.

3.2.2 Shapes of value functions

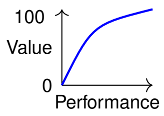
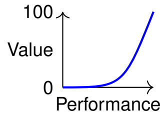
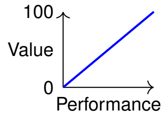
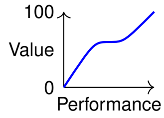
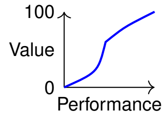
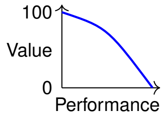
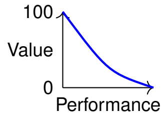
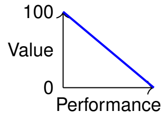
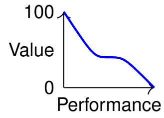
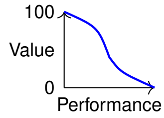
In MACBETH judgements, we can say that (Parnell, 2012):

- If all the judgements between successive similar differences are the same, then the value function is linear.
- If the judgements between successive similar differences are decreasing, then the value function is concave.
- If the judgements between successive similar differences are increasing, then the value function is convex.
- If the judgements between successive similar differences are decreasing until the inflexion point and then increasing, then the value function is S-seat.
- If the judgements between successive similar differences are increasing until the inflexion point and then decreasing, then the value function is S-sigmoid.

To visually represent the value function of a quantitative descriptor, one simply plots the descriptor on the x-axis and the partial value on the y-axis. For qualitative descriptors, value functions can be characterized by their shape, and can be classified into different categories. We start by classifying them into increasing and decreasing functions. Then, we can classify them into linear, concave, convex, S-shaped, and S-seat functions (Bana e Costa et al., 2023a; Parnell, 2012).

Refer to Table 3.1 for a visual representation of the different types of value functions.

Table 3.1: Shape-based taxonomy for Value Functions. (Bana e Costa et al. (2023a))

	Concave	Convex	Linear	S-seat	S-sigmoid
Increasing					
Decreasing					

Increasing functions

Increasing functions are functions that increase as the performance of the option on the criterion increases. They can be classified into five categories: increasing concave, increasing convex, increasing linear, increasing S-seat, and increasing S-sigmoid functions:

- Increasing Concave: Successive increases in performance increase value less and less
- Increasing Convex: Successive increases in performance increase value more and more
- Increasing Linear: Successive increases in performance increase value by the same amount
- Increasing S-seat: Successive increases in performance increase value less and less, until the inflexion point, then successive increases in performance increase value more and more
- Increasing S-sigmoid: Successive increases in performance increase value more and more, until the inflexion point, then successive increases in performance increase value less and less

Decreasing functions

Decreasing functions are functions that decrease as the performance of the option on the criterion increases. They can be classified into five categories: decreasing concave, decreasing convex, decreasing linear, decreasing S-seat, and decreasing S-sigmoid functions:

- Decreasing Concave: Successive increases in performance decrease value less and less
- Decreasing Convex: Successive increases in performance decrease value more and more
- Decreasing Linear: Successive increases in performance decrease value by the same amount
- Decreasing S-seat: Successive increases in performance decrease value less and less, until the inflexion point, then successive increases in performance decrease value more and more
- Decreasing S-sigmoid: Successive increases in performance decrease value more and more, until the inflexion point, then successive increases in performance decrease value less and less

3.2.3 Delta functions respect the constant trade-off attitude property

When working with value functions, there is a property that can be present called the constant trade-off attitude property. In the following section, this property will presented.

The constant trade-off attitude property

Kirkwood, 1996 defines the delta property or the constant risk aversion property for preferences in a decision problem with a single evaluation attribute as follows: The property of constant risk aversion exists when altering all possible outcomes of an uncertain option by a fixed amount results in the certainty equivalent of that option also changing by the same fixed amount.

Corner, 1994 explains how the constant trade-off attitude property or the delta property (Pratt, 1964) for preferences in a decision problem with a single evaluation attribute guarantees the exponential form:

$$v_i(x_i) = \frac{1 - exp[c_i(x_i^0 - x_i)]}{1 - exp[c_i(x_i^0 - x_i^*)]}, c_i \neq 0 \tag{3.3}$$

Note: This is the form for an attribute with increasing preferences. The form for an attribute with decreasing preferences is similar and is presented in the next page.

Under certain conditions, the exponential form degenerates to the linear form:

$$v_i(x_i) = \frac{x_i - x_i^0}{x_i^* - x_i^0}, c_i = 0 \tag{3.4}$$

where c_i are the trade-off attitude constants, and x_i^0 and x_i^* are the lowest and highest possible levels of x_i.

The delta property holds for attribute x_i if, given the midvalue, x_i^2, of an interval along that attribute $[x_i^1, x_i^3]$, then $x_i^2 + \delta$ is the midvalue of $[x_i^1 + \delta, x_i^3 + \delta]$ for any δ. The midvalue x_i^2 of an interval $[x_i^1, x_i^3]$ on an attribute x_i is that value within the interval such that a decision-maker would give up the same amount of some other attribute to move from x_i^1 to x_i^2 as from x_i^2 to x_i^3.' (Corner, 1994)

Figure 3.9 shows an example of a delta function that satisfies the constant trade-off attitude property. The figure shows the midvalue of an interval along the attribute x_i and the midvalue of the interval shifted by δ. The delta property can be summed in the following definition:

$$\begin{aligned} &\text{if} \\ &x_i^m = \text{midvalue of } [x_i', x_i''] \\ &\text{then} \\ &x_i^m + \delta = \text{midvalue of } [x_i' + \delta, x_i'' + \delta] \end{aligned} \tag{3.5}$$

Family of continuous functions that satisfy the constant trade-off attitude property

In this dissertation, the name delta function will be used to refer to a family of functions that satisfy the constant trade-off attitude property and are defined as:

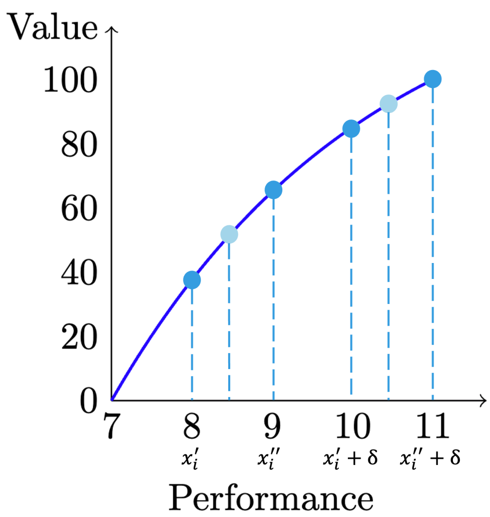


Figure 3.9: Example of a delta function that satisfies the constant trade-off attitude property.

Increasing value function:

$$v_i(x_i) = \frac{1 - exp[c_i(x_i^0 - x_i)]}{1 - exp[c_i(x_i^0 - x_i^*)]}, c_i \neq 0$$

$$v_i(x_i) = \frac{x_i - x_i^0}{x_i^* - x_i^0}, c_i = 0$$

Decreasing value function:

$$v_i(x_i) = \frac{1 - exp[-c_i(x_i - x_i^*)]}{1 - exp[-c_i(x_i^0 - x_i^*)]}, c_i \neq 0$$

$$v_i(x_i) = \frac{x_i^* - x_i}{x_i^* - x_i^0}, c_i = 0$$

where c_i are the trade-off attitude constants, x_i^0 are the lowest possible levels of x, and x_i^* are the highest possible levels of x (Corner, 1994).

Note that the constant trade-off attitude is analagous to the constant risk aversion property, but the former is more used in the context of value functions and the latter in the context of utility functions. The family of functions that satisfy the constant risk aversion property are usually defined by a ρ parameter, and are equivalent to the family of functions that satisfy the constant trade-off attitude property, which are defined by a c parameter, in the following manner:

$$\rho = \frac{1}{c} \quad \Leftrightarrow \quad c = \frac{1}{\rho} \tag{3.6}$$

3.2.4 Building value functions with MACBETH

LP-MACBETH is the formulation of the conditions from the MACBETH method as a linear programming problem. The categorical judgements of pairwise comparisons lead to a system of inequalities that can be solved using linear integer programming. The resulting matrix is called the Basic Differences Ma-

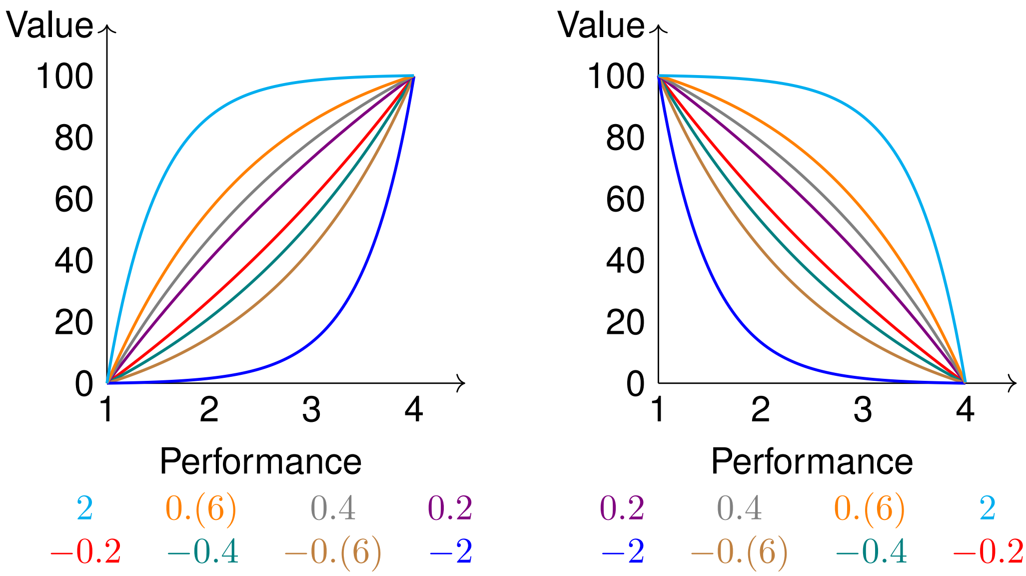


Figure 3.10: Example of increasing (left side) and decreasing (right side) delta functions with different trade-off attitude constants c. From left to right, the trade-off attitude constants of the increasing delta functions are 2, 0.(6), 0.4, 0.2, -0.2, -0.4, -0.6, -2. From left to right, the trade-off attitude constants of the decreasing delta functions are -2, -0.(6), -0.4, -0.2, 0.2, 0.4, 0.6, 2.

trix. From the Basic Differences Matrix, the MACBETH Basic scale is used to build the value function. The Basic scale is scaled (normalized) to the range of the value function [minimum value, maximum value] to obtain the value function.

An example of how MACBETH Basic Differences are calculated from MACBETH Semantic Judgements and then used to build a value function is shown in Figure 3.11. All the numbers in the matrix of Basic Differences are positive integers.

The lowest values in the matrix of Basic Differences are to the leftmost and bottommost cells, and the highest values are to the rightmost and topmost cells. The lowest value in this matrix is 0, and the highest value is 9.

All the pairwise comparisons where the decision-maker answered ‘no’ are assigned a value of 0.

The MACBETH Basic scale is equal to the last column of the matrix, which in this case is [0, 4, 6, 8, 9].

3.3 Framework for Developing Decision Support Systems

In this section, a framework for understanding and developing Decision Support Systems (DSS) is presented. The present framework is based on the work of (Sprague, 1980), which provides a structured approach to developing DSS, with an emphasis on various technological, methodological, and organizational aspects that are essential in the creation of effective decision support tools.

Definitions for DSS range from more restrictive to broader interpretations. A more restrictive definition of DSS defines DSS as interactive computer-based systems that aid decision-makers in solving unstructured complex problems by providing analytical tools, models, and data visualization capabilities.

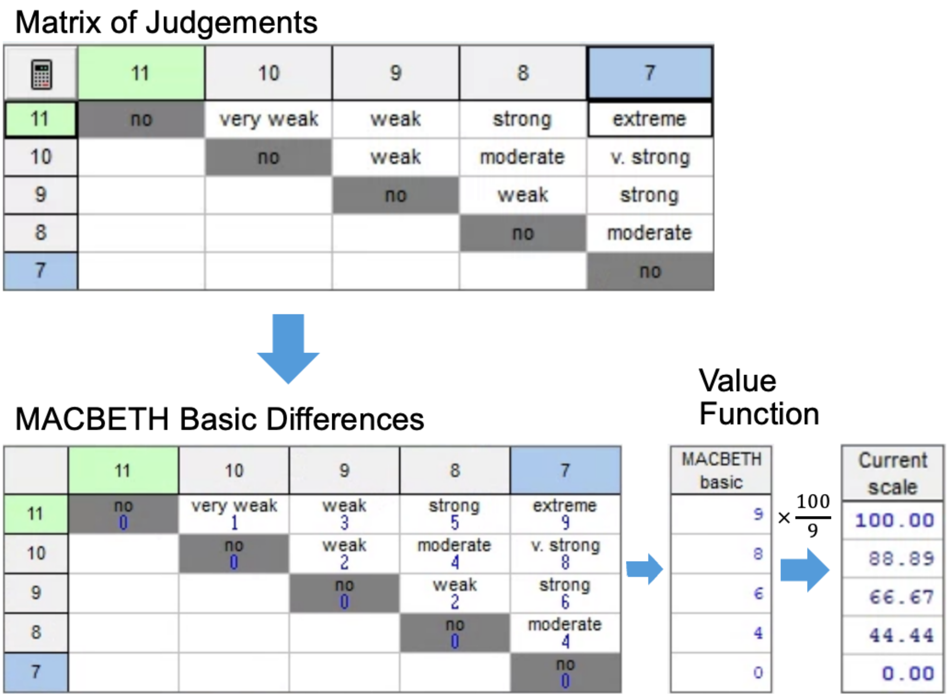


Figure 3.11: Diagram showing how MACBETH Basic Differences are calculated from MACBETH Semantic Judgements and then used to build a value function.

A more comprehensive definition of DSS includes any system that supports decision-making processes, regardless of the level of structure or complexity of the problem. This broader definition encompasses a wide range of tools and technologies that assist decision-makers in making informed and effective decisions.

The framework is divided into two main parts. The first describes the three different levels of technology, the developmental process, and the roles of the participants in the development process. The second part focuses on a descriptive model to understand and access the objectives for the performance and capabilities of the DSS.

The three levels of technology are:

- Data: Access and manipulate data to provide information for decision-making.
- Model: Use analytical models to support decision processes.
- Knowledge: Incorporate expert knowledge into the system to better guide decision-making.

The developmental process should be iterative, with continuous feedback and evaluation to ensure that the system meets its requirements and objectives. Thus combining the analysis, design, construction and implementation into a single step that is repeated iteratively. The process should actively involve users and stakeholders to ensure that the system is user-friendly and meets the needs of the users.

The roles of the participants are relevant to implementing DSS in an organization. The participants are the Manager (User), the Intermediary, the DSS Builder, the Technical Supporter and the Toolsmith, and include Executives and Professionals, Managers, Information Specialists, System Designers and Researchers.

3.4 Challenges

This chapter reviews existing research on adjusting value functions to fit data, the aggregation of judgements in group decisions, how to handle missing values in incomplete matrices and the number of judgements needed for effective decision-making.

There have been reports of concerns about some MCDA models in Healthcare not being able to represent all relevant views and preferences because of the small number of participants in the decision-making process (Oliveira et al., 2019). This issue can be addressed with more robust methods for aggregating or reconciling judgements in group decisions.

There have also been reports from participants of difficulties in the interpretation and understanding of MCDA models in Healthcare (Oliveira et al., 2019). This can be addressed by ensuring that Decision Support Systems (DSS) are user-friendly, intuitive and understandable.

Another challenge is the need to work with limited resources, such as the time of decision-makers and costs (Oliveira et al., 2019). This can be addressed by developing efficient and effective methods for decision-making that maximize the use of available resources, including the time and expertise of decision-makers.

A challenge relates to the different levels of numeracy and fluency of the decision-makers. The model-building methods need to be tailored to the decision-makers' numeracy and fluency (Fasolo and Bana e Costa, 2014).

This work will focus on three main challenges:

- Designing curve fitting tools to adjust value functions to fit data.
- Designing tools to reconcile value functions in group decisions.
- Developing approaches to analyse the variability of incomplete matrices.

In the following sections, these challenges will be discussed in more detail.

3.4.1 Designing curve fitting tools

Some applications require a continuous value function to represent the preferences of the decision-makers. However, the value functions obtained from MACBETH judgements may not always be continuous. In such cases, it is necessary to adjust the value functions to ensure continuity and smoothness. Tools to adjust value functions to fit data and analyse their shapes are needed or can be useful (Vieira et al., 2020). This process involves curve-fitting techniques to approximate the value functions based on the available data points. The goal is to find a function that best fits the data and can be used to represent the preferences of the decision-makers accurately.

When exploring what kinds of functions can be used to represent the value functions, there are some possibilities:

- Polynomial functions: Polynomial functions can capture complex relationships between the performance of the options and the partial values. However, polynomial functions may not have the constant trade-off attitude property and thus may not be suitable for representing the preferences of the decision-makers accurately. Moreover, higher-order polynomials can be prone to overfitting and may not generalize well to new data.

- Exponential functions: Exponential functions can model the exponential growth or decay of the value functions. Namely, the delta functions that respect the constant trade-off attitude property.
- Other functions: Other functions, such as logarithmic, power, or sigmoid functions, can also be used to represent the value functions. These functions can capture different types of relationships between the performance of the options and the partial values.

For fitting the value functions, we can use curve fitting techniques such as least squares regression, nonlinear optimization, or machine learning algorithms. These techniques aim to find the best-fitting function that minimizes the error between the observed data points and the predicted values. By adjusting the parameters of the function, a continuous and smooth value function that accurately represents the preferences of the decision-makers can be obtained.

Least Squares Regression

Least squares regression is a common curve-fitting technique that minimizes the sum of the squared differences between the observed data points and the predicted values. The goal is to find the best-fitting function that describes the relationship between the independent and dependent variables. In the context of value functions, least squares regression can be used to fit a function to the observed data points and obtain a continuous value function that represents the preferences of the decision-makers.

The least squares regression problem can be formulated as follows:

$$\min_f \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{3.7}$$

3.4.2 Designing tools to reconcile value functions

In the case of group decisions involving multiple decision-makers, a challenge is how to aggregate the judgements of different individuals. Each decision maker may have their preferences and priorities, and finding a consensus can be complex. Aggregating the judgements requires a systematic and fair approach to ensure that the collective preferences are appropriately represented.

In the MACBETH method, having multiple decision-makers with differing judgements can pose challenges in reaching a group compromise judgement. The process typically involves presenting and discussing arguments, voting, discussing disparities in votes if they occur, and then voting again. This iterative process aims to reach a consensus among the decision-makers, but it can be time-consuming and may require multiple rounds of discussion and voting to achieve a satisfactory compromise. Additionally, managing and incorporating the diverse perspectives of decision-makers can be complex, as it requires careful consideration of each individual's input and preferences (Bana e Costa et al., 2014; Bana e Costa and Vansnick, 1994).

3.4.3 Developing approaches to analyse the variability of incomplete matrices

When a matrix of judgements is incomplete, meaning that not all pairwise comparisons have been made, we need to address the issue of missing values. In such cases, various techniques can be employed to estimate or infer the missing values based on the available information. These techniques aim to fill in the gaps and create a complete matrix of judgements, which is necessary for the subsequent analysis and calculation of the total values for each option.

One challenge is determining how many judgements should be asked of the decision-maker. The number of judgements required tends to grow with the number of options, which can be time-consuming and burdensome for the decision-maker. Therefore, it is crucial to minimize the number of judgements while still obtaining sufficient information for the evaluation process. That is, it's important to find a balance between the number of judgements and the robustness of the results.

In the MACBETH method, the minimal number of judgements required is n-1, where n represents the number of states being assessed. However, it is recommended to ask for additional judgements to perform consistency checks and obtain higher precision in numerical assessments (Oliveira et al., 2018).

The number of judgements required tends to grow with the number of options. However, not all possible matrices of judgements are consistent.

If one or more of the decision-makers do not provide a judgement for a particular pair of options, the matrix of judgements is incomplete. See Figure 3.12 for an example of an incomplete matrix of judgements.

		Options			
		Option 1	Option 2	Option 3	Option 4
Options	Option 1	—	v. weak	?	v. strong
	Option 2		—	v. weak	moderate
	Option 3			—	v. weak
	Option 4				—

Figure 3.12: Example of an incomplete matrix of judgements. In this example, a decision-maker has not provided a judgement for Option 1 compared to Option 3. The missing value is represented by a question mark.

M-MACBETH fills incomplete matrices by transitivity

The M-MACBETH DSS handles an incomplete matrix of judgements by completing the remaining entries by transitivity (Bana e Costa et al., 2016). This process is done by solving the linear programming problem LP-MACBETH. After this process, in M-MACBETH the resulting value functions are presented to the DMs for validation and necessary adjustments (Bana e Costa et al., 2022).

Possible completions of incomplete matrices

Each missing value in an incomplete matrix of judgements can be filled with any of the possible semantical categorical judgements. The missing value from the example from Figure 3.12 can be filled with any of the following semantical categorical judgements: very weak, weak, moderate, strong, or very strong. When the matrix is filled with each one of these possible values, we get a different scale. See Figure 3.13 for the possible completions of the incomplete matrix of judgements from Figure 3.12.

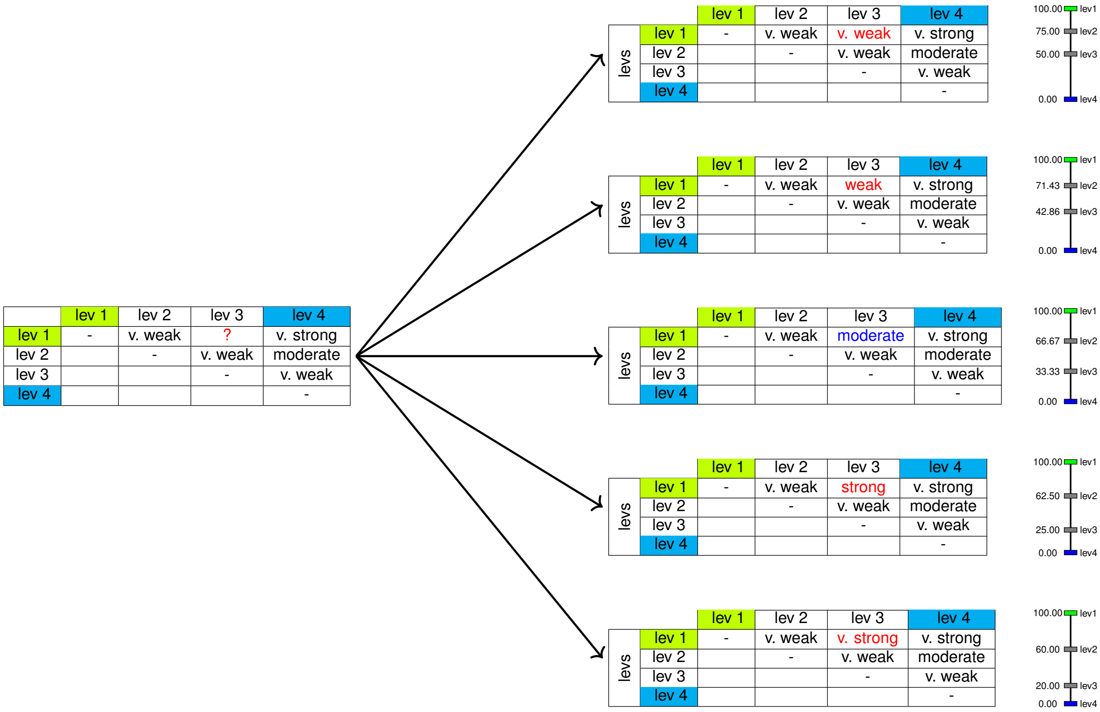


Figure 3.13: Example of an incomplete matrix of judgements. In this example, a decision-maker has not provided a judgement for Option 1 compared to Option 3. The missing value is represented by a question mark. We see that the possible values for this missing value are very weak, weak, moderate, strong, and very strong. Macbeth automatically fills in the missing judgement by transitivity with moderate. When the matrix is filled with each one of these possible values, we get a different scale.

Chapter 4

Methodological Approach

This chapter details the methodological approach used to address the research questions related to the application of the MACBETH method in decision support systems. The focus is on developing a framework to fit a continuous value function, reconciling value functions from multiple decision-makers, and assessing the stability of incomplete judgement matrices.

4.1 Overview of the Methodological Approach

The methodological approach can be divided into design and implementation phases, with continuous feedback loops between the two.

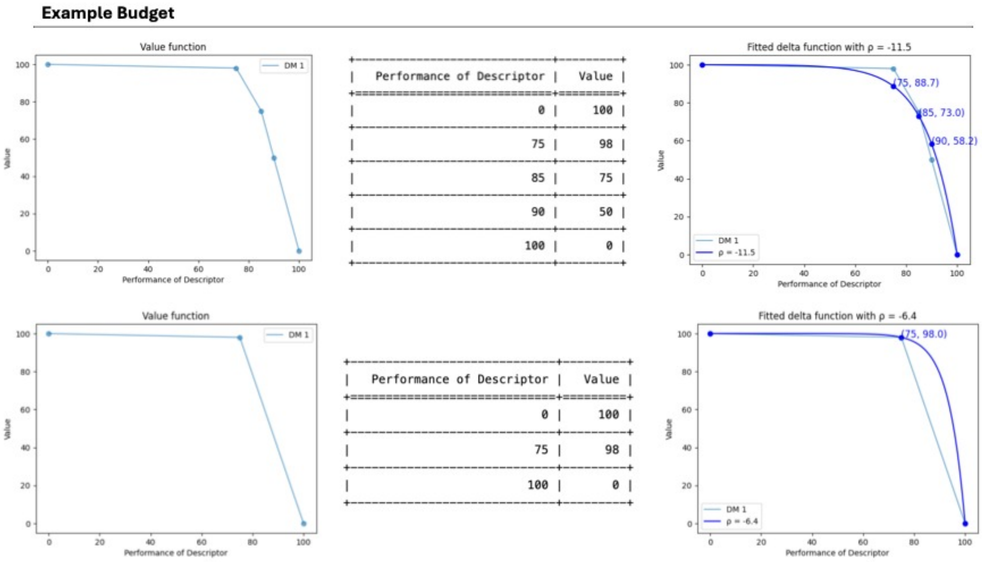
The design phase involves gathering requirements, defining the use cases, and creating quick prototypes to validate the methodologies. The implementation phase involves developing the tools and algorithms to apply the methodologies to data. The feedback loops ensure that the tools and methodologies are refined and improved based on the results obtained during the implementation phase.

During the design phase, regular meetings with experts in the field were held to gather requirements and define the use cases. Quick prototypes were developed to validate the methodologies and ensure that they meet the requirements. The quick prototypes were created using Python and Jupyter notebooks, which allowed for rapid development and testing of the methodologies. Figma was used to create wireframes and mockups of the user interface to ensure that the tools were user-friendly and intuitive. See Figure 4.1 for an example of results from the prototype for a case explored during the design phase using Jupyter Notebooks.

For the implementation phase of the project, the front end was developed using React.js, a popular JavaScript library for building user interfaces and the back end was developed using Flask, a lightweight web framework for Python.

The work on curve fitting tools was presented at a DECISING meeting and the APDIO 2024 Conference.

The implementation phase involves developing the tools and algorithms to apply the methodologies to real-world data. The feedback loops ensure that the tools and methodologies are refined and improved based on the results obtained during the implementation phase.



Performance of Descriptor	Value
0	100
75	98
85	75
90	50
100	0

Performance of Descriptor	Value
0	100
75	98
100	0

Figure 4.1: Example of results from the prototype for a case explored during the design phase using Jupyter Notebooks.

4.2 Curve Fitting

The objective was to develop methods and tools to fit an exponential value function to a set of MACBETH value scores, ensuring the curve adheres to the delta property, thus providing a continuous representation of value functions as opposed to piecewise representations.

The method designed uses the delta function to fit the value functions to the data points. The delta function was chosen because it satisfies the constant trade-off attitude property. The goal is to find the optimal c that best fits the data points.

The method for curve fitting uses a nonlinear Least Squares Regression technique to find the parameter c. The tool used for this was a Python package named scipy.optimize, each enable the program to fine-tune the parameter of the delta function, ensuring the best possible fit.

```
import numpy as np
from scipy.optimize import curve_fit

def delta(x: np.ndarray, x0: float, xplus: float, y0: float, yplus: float, rho: float,
    increasing: bool = False) -> np.ndarray:
    '''
    The delta function is a function that respects the constant trade-off property.
    The function is defined for increasing functions as:
    delta(x) = (1 - exp(- (x - x0) / rho)) / (1 - exp(- (xplus - x0) / rho)) * (yplus -
    y0) + y0
    and for decreasing functions as:
    delta(x) = (1 - exp(- (xplus - x) / rho)) / (1 - exp(- (xplus - x0) / rho)) * (yplus
    - y0) + y0
```

```
    where x0 is the minimum performance, xplus is the maximum performance,
    y0 is the value at the minimum performance, yplus is the value at the maximum performance
    and rho is the trade-off constant.

    Arguments:
    x: np.ndarray - the performance of the
    x0: float - the minimum performance of the descriptor
    xplus: float - the maximum performance of the descriptor
    y0: float - the value at the minimum performance
    yplus: float - the value at the maximum performance
    rho: float - the trade-off constant
    increasing: bool - whether the descriptor is increasing or decreasing

    Returns:
    float - the value of the delta function at x
    '''
    if increasing:
        return (1 - np.exp(-(xplus - x) / rho)) / (1 - np.exp(-(xplus - x0) / rho)) * (yplus - y0) + y0
    return (1 - np.exp(-(x - x0) / rho)) / (1 - np.exp(-(xplus - x0) / rho)) * (yplus - y0) + y0

def fit_delta(data: np.ndarray):
    '''
    Fits the delta function to the data points and returns the optimal rho value.

    Arguments:
    data: np.ndarray - the data points

    Returns:
    float - the optimal rho value
    '''
    x0 = data[0, 0, 0]
    xplus = data[0, -1, 0]

    y0 = data[0, 0, 1]
    yplus = data[0, -1, 1]

    x = data[:, :, 0]
    y = data[:, :, 1]

    def loss_fn(x: np.ndarray, y: np.ndarray, rho: float) -> float:
        '''
        The loss function is the sum of the squared differences between the data points and the delta function.

        Arguments:
        x: np.ndarray - the performance of the descriptor
```

```
        y: np.ndarray - the value of the descriptor
        rho: float - the trade-off constant

        Returns:
        float - the loss value
        '''
        return np.sum(np.square(y - delta(x, x0, xplus, y0, yplus, rho)))

    # Use curve_fit from scipy.optimize to find the optimal rho value
    # There are two minima, so we need to find both and choose the one with the lowest loss
    # The function can be increasing or decreasing, so we need to find the optimal rho for both cases

    x_ = x.flatten()
    y_ = y.flatten()
    optim_rho1 = curve_fit(lambda x, rho: delta(x, x0, xplus, y0, yplus, rho), x_, y_, p0=[10])[0][0]
    optim_rho2 = curve_fit(lambda x, rho: delta(x, x0, xplus, y0, yplus, rho), x_, y_, p0=[-10])[0][0]
    optim_rho1inc = curve_fit(lambda x, rho: delta(x, x0, xplus, y0, yplus, rho, increasing=True), x_, y_, p0=[10])[0][0]
    optim_rho2inc = curve_fit(lambda x, rho: delta(x, x0, xplus, y0, yplus, rho, increasing=True), x_, y_, p0=[-10])[0][0]

    options = [optim_rho1, optim_rho2, optim_rho1inc, optim_rho2inc]

    optim_rho = options[np.argmin([loss_fn(x_, y_, rho) for rho in options])]

    # Calculate R^2
    # R^2 is the coefficient of determination,
    # which is the proportion of the variance in the dependent variable that is predictable from the independent variable
    # popt, pcov = curve_fit(lambda x, rho: delta(x, x0, xplus, y0, yplus, rho), x_, y_, p0=[optim_rho])
    # residuals = y_ - delta(x_, x0, xplus, y0, yplus, optim_rho)
    # ss_res = np.sum(residuals**2)
    # ss_tot = np.sum((y_ - np.mean(y_))**2)
    # r2 = 1 - (ss_res / ss_tot)
    r2 = 1 - loss_fn(x_, y_, optim_rho) / np.sum(np.square(y - np.mean(y)))

    return [optim_rho, r2]
```

Listing 4.1: Python code for curve fitting

4.3 Reconciling Value Functions

When building a matrix of judgements with multiple decision-makers, it is common to have different judgements for the same pair of options. In Figure 4.2, we see an example of a matrix of judgements with three decision-makers with one judgement that is different for the three decision-makers.

		Options			
		Option 1	Option 2	Option 3	Option 4
Options	Option 1	-	v. weak	weak	v. strong
	Option 2		-	v. weak	moderate
	Option 3			-	v. weak
	Option 4				-

		Options			
		Option 1	Option 2	Option 3	Option 4
Options	Option 1	-	v. weak	moderate	v. strong
	Option 2		-	v. weak	moderate
	Option 3			null	v. weak
	Option 4				null

		Options			
		Option 1	Option 2	Option 3	Option 4
Options	Option 1	-	v. weak	strong	v. strong
	Option 2		-	v. weak	moderate
	Option 3			-	v. weak
	Option 4				-

		Options			
		Option 1	Option 2	Option 3	Option 4
Options	Option 1	-	v. weak	?	v. strong
	Option 2		-	v. weak	strong
	Option 3			-	v. weak
	Option 4				-

Figure 4.2: Example of aggregation of judgements in group decisions. In this example, the judgements of three decision-makers need to be aggregated to obtain a single matrix of judgements. In red are the judgements that are not consistent with the majority of the decision-makers.

Reconciliation of value functions can be addressed by designing a method to integrate different individual value functions from multiple decision-makers into a single, unified collective value function that reflects a collective view.

There are several decisions to be made regarding the process of reconciliation as can be seen in Figure 4.3. They all have arguments for and against them. For example, an aggregation without correcting the shape of the value function of the minorities might be more suited to applications where every decision maker is equally important. On the other hand, an aggregation with shape correction might be more suited to applications where the majority of the decision-makers are more important than the minority. For example, when setting policies, not doing this correction might result in policies that are not suited to any of the decision-makers and thus are worse than either of the policies that the decision-makers would choose individually. For these nuanced situations, it is advantageous to have a set of approaches that can be used in different situations and metrics to evaluate the quality of the aggregation. With these metrics, we can notice when nuanced situations like these arise and choose the approach that is more suited to the situation.

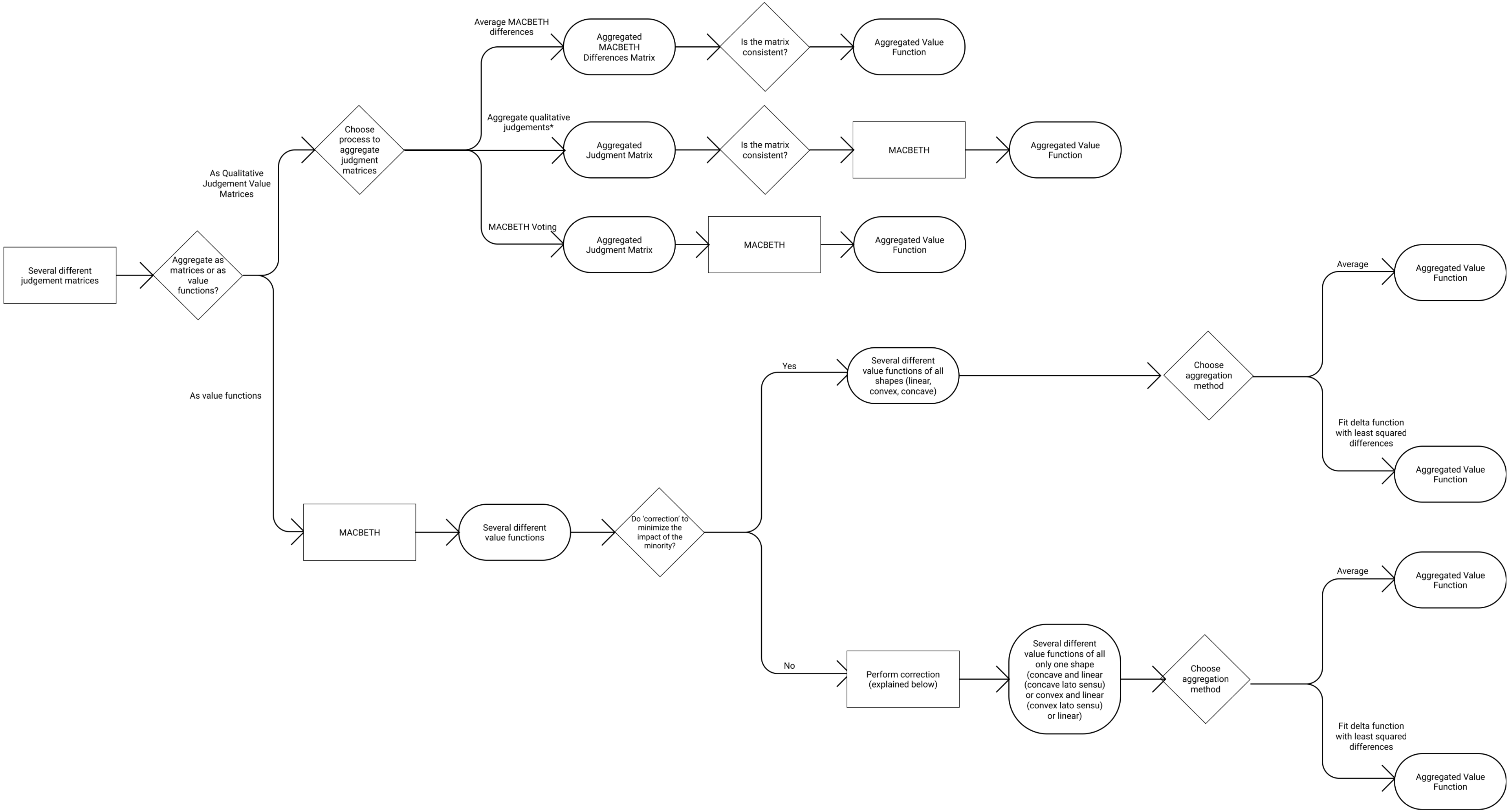


Figure 4.3: Flowchart of the aggregation process. There are several decisions to be made on how to aggregate the value functions of the decision-makers. Note that the flowchart is not exhaustive and there are other possible approaches with pros and cons by taking different decisions not represented in the flowchart. For example, when correcting the shape of the value functions, an option can be made to fully disregard the value functions of the minorities instead of correcting them to linear, as was done in the flowchart below. This limits the influence of the minority to a linear shape while not fully disregarding their value functions.

Possible approaches to aggregate judgements in group decisions

One of the objectives of this thesis is to develop a method that can aggregate judgements in group decisions.

List, 2012 formally studies judgement aggregation, where a group seeks collective judgements on interconnected propositions based on individuals' judgements. Majority voting risks inconsistent judgements, generalizing Condorcet's voting paradox. List, 2012 models judgements aggregation using logic: the agenda contains propositions and negations, individuals submit judgement sets, and an aggregation rule produces collective judgements. Impossibility theorems show proposition-wise rules like majority voting fail absent degeneracy conditions. However, relaxing independence allows non-dictatorial possibilities like premise-based voting (List, 2012).

When there are differences between the judgements of the decision-makers, it is important to note that the option to ask the decision-makers to revise their judgements might not exist.

Majority Voting

Majority voting is a method for aggregating judgements in group decisions. In majority voting, the majority opinion is chosen as the collective decision.

Using majority voting on consistent matrices of judgements may lead to inconsistent matrices of judgements. An intuition for this hypothesis is the doctrinal paradox (List, 2012), which may occur for cardinality constraints. The doctrinal paradox states that premise-based and conclusion-based rules can produce opposite outcomes for the same combination of individual judgments and can be explained through the following example:

Consider the logical propositions p, q, and r, where p and q are jointly necessary and sufficient for r.

$$p \land q \leftrightarrow r \tag{4.1}$$

The logical value of the two first propositions p and q is not known, and three decision-makers are asked to provide their input. The three decision-makers provide the following input:

Table 4.1: Input of the decision-makers.

Decision-maker	p	q
1	True	True
2	True	False
3	False	True

The majority of the decision-makers agree that the logical value of p and q is true. However, the majority of the decision-makers may disagree on the logical value of r.

$$p \land q \Leftrightarrow r \tag{4.2}$$

If p and q are true, then r is true. If p or q is false, then r is false.

When starting by aggregating by majority voting the judgements of p and q, p and q are true and thus r is true. However, when starting by aggregating by majority voting the judgements of r, r is false. See Table 4.2 for the example of majority voting.

Table 4.2: Example of paradox in majority voting.

Decision-maker	p	q	r
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True	True	?

Additive Group Utility Functions

The aggregated value function is the weighted average of the individual value functions (Keeney and Raiffa, 1976; Harsanyi, 1955).

If each decision-maker has a value function $u_i(x)$, then the group value function is the weighted average of the individual value functions as seen in Equation 4.3 (Keeney and Raiffa, 1976; Harsanyi, 1955).

$$u(x) = \sum_{i=1}^{N} \lambda_i u_i(x), \lambda_i \geq 0 \tag{4.3}$$

The weights are the importance of each decision-maker. The weights can be obtained by asking the decision-makers to rate the importance of each decision-maker (Cooke and Cooke, 1992).

In the following subsection the possible approaches to aggregate value functions in the context of MACBETH are discussed.

Aggregation of the semantical judgements through the value function building process

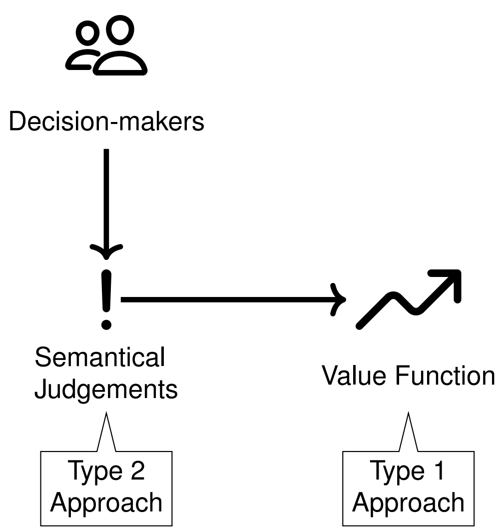


Figure 4.4: Stages of value function building

There are several approaches that aggregate the data of the decision-makers at different stages of the value function building process. The stages of the value function building process are shown in

Figure 4.4 and are the following:

1. **Aggregation of the value functions**
   The value functions of the decision-makers are aggregated after the value function for each decision-maker group is built.

2. **Aggregation of the semantical judgements**
   The semantical judgements of the decision-makers are aggregated before the value function is built.

To reconcile the semantical judgements, an option would be to do a decision conference or another survey round, e.g. by presenting the value functions shapes or value scores to the DMs and let them choose. This is a good approach and there is a drawback regarding questions that arise about how well the DMs understand the meaning of the value functions shapes. In a delphi method, the decision-makers can be shown a visual representation of the resulting value function according to the matrix of judgements they filled out and are also the value functions of the other decision-makers, along with their distribution. Then, they are asked to change their judgements if they want to. This process is repeated until consensus is reached. In some cases, it's not possible to do this approach because of resource constraints such as time.

It's unclear how to do an average of semantical judgements without direct contact between the decision-makers. The approaches for reconciliating judgements without communication between decision-makers are not ideal as there are questions regarding the incomparability of the semantical judgements from different decision-makers when there isn't a discussion between them. Without a discussion, the semantical judgements are not comparable because e.g. a weak judgement from one decision-maker can be a moderate judgement from another decision-maker. Also, these approaches have difficulty guaranteeing that the resulting matrix of semantical judgements is consistent.

In the following section, two approaches using the value scores of the decision-makers are presented using an example.

Example for the aggregation of value functions

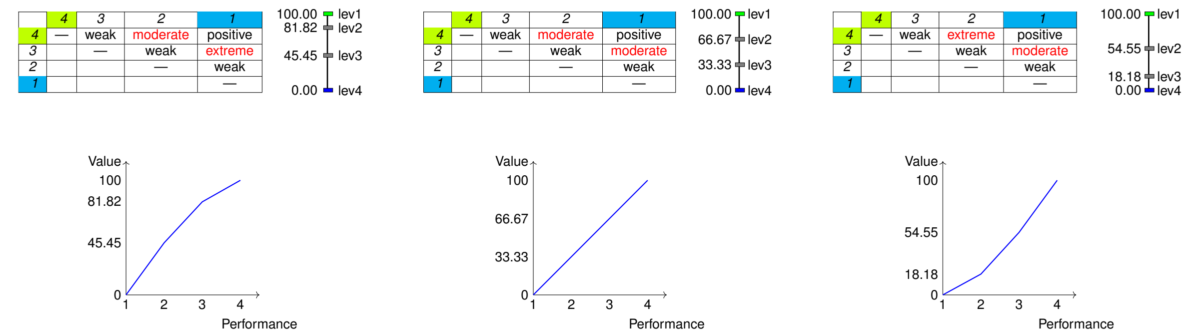
Consider the following example:

7 (N) decision-makers are asked to fill a matrix of MACBETH semantical judgements. The answers of all the decision-makers are similar. In this case, there are six judgements to fill in the matrix of judgements, only two judgments vary between the decision-makers, the judgement of difference in attractiveness between a performance of 2 and 4 and the judgement of difference in attractiveness between a performance of 1 and 3. The decision-makers are grouped according to their answers, forming three groups of decision-makers. The resulting value functions are concave, linear, and convex, respectively. The resulting value functions are shown in Figure 4.5.

In the figure, for each group of decision-makers, there is the matrix of MACBETH judgements, the calculated value scale by the software M-MACBETH and the visual representation of the value function.

The first group (c for concave) has 3 decision-makers, the second group (l for linear) has 2 decision-makers and the third group (v for convex) has 2 decision-makers. The shapes of the value functions resulting from their judgements are concave, linear and convex, respectively. The first group of decision-makers represents the majority of the decision-makers, and the second and third groups each represent a minority of the decision-makers.

In Table 4.3, the value scores and the shape of the value functions are shown for each group of decision-makers.



Decision-makers Group 1 (c = 3, c / N = 3 / 7 = 43% (majority)) Decision-maker Group 2 (l = 2, l / N = 2 / 7 = 28.5%) Decision-maker Group 3 (v = 2, v / N = 2 / 7 = 28.5%)

Figure 4.5: Macbeth Semantical Judgements of 7 Decision-makers

Table 4.3: Initial data of the groups of decision-makers and respective value function shape

Decision-makers Group	N	Value function shape	Scale
1	3	Concave	100, 81.82, 45.45, 0
2	2	Linear	100, 66.67, 33.33, 0
3	2	Convex	100, 54.55, 18.18, 0

Correction of outliers

In this subsection, a methodology for the correction of the outliers is proposed. See Figure 4.6 for a flowchart of the shape correction process.

This methodology will be applied to the previous example before the aggregation of the value scores.

In this case, the majority of the decision-makers say that the shape of the value function is concave (Decision-makers Group 1). Then, the shape of the value functions of the other decision-makers is changed the limit of a concave function which is a linear function. The scale of the Decision-Makers Group 3 becomes equal to the scale of Decision-Makers Group 2. The corrected data of the groups of decision-makers and respective value function shape are shown in Table 4.4.

Table 4.4: Groups of decision-makers and respective value function shape where the scale of decision-makers group 3 is equal to the scale of decision-makers group 2

Decision-makers Group	N	Value function shape	Scale
1	3	Concave	100, 81.82, 45.45, 0
2	2	Linear	100, 66.67, 33.33, 0
3	2	Linear	100, 66.67, 33.33, 0

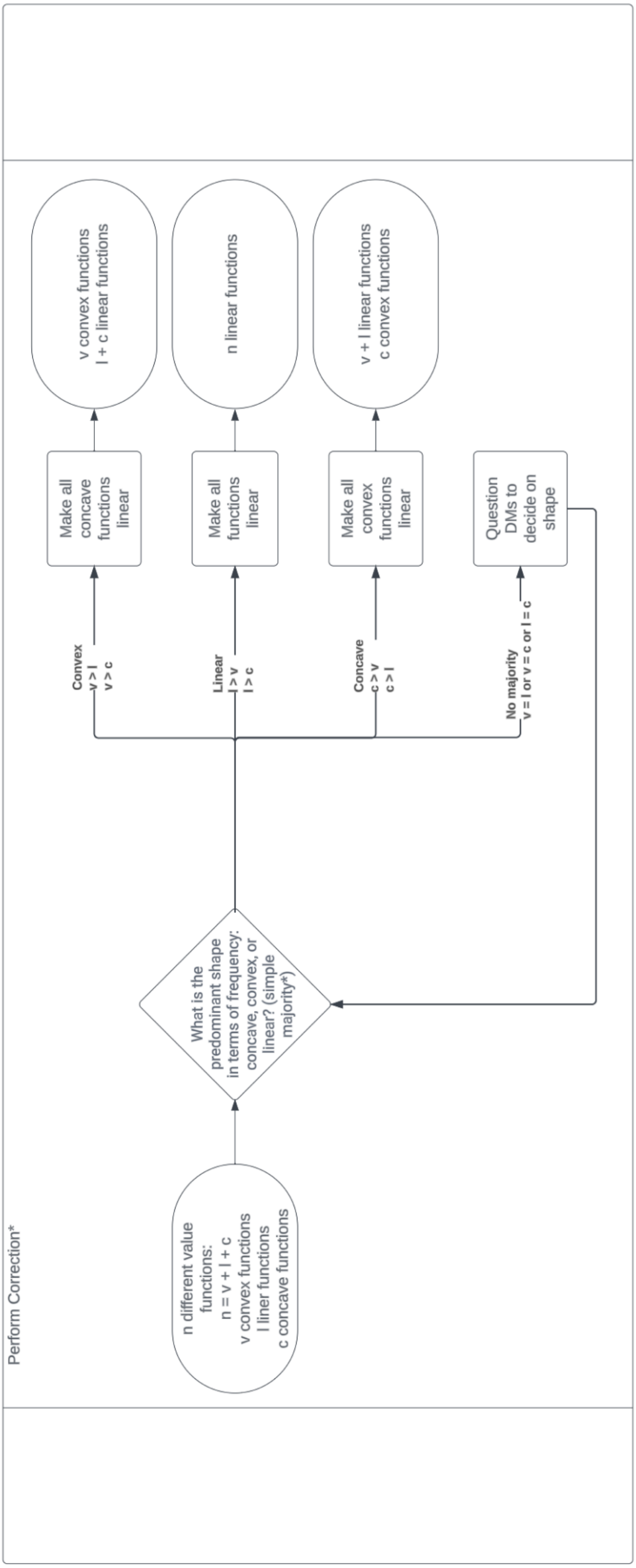


Figure 4.6: Flowchart of the shape correction process

Approaches for the aggregation of value scores

Two approaches to aggregate the value scores will be shown:

- **Approach a:** Direct average the points of the value functions
- **Approach b:** Minimization of the square error between the value functions and a function with delta property

Approach a: Direct average the points of the value functions

To do a direct average of the points of the value functions, the mean of the points of the value scale is calculated:

$$mean = \frac{n_{DMG_1} \times v_{DMG_1} + n_{DMG_2} \times v_{DMG_2} + n_{DMG_3} \times v_{DMG_3}}{n_{DMG_1} + n_{DMG_2} + n_{DMG_3}}$$
$$\frac{3 \times 45.45 + 2 \times 33.33 + 2 \times 33.33}{7} = 38.52$$
$$\frac{3 \times 81.82 + 2 \times 66.67 + 2 \times 66.67}{7} = 73.16$$

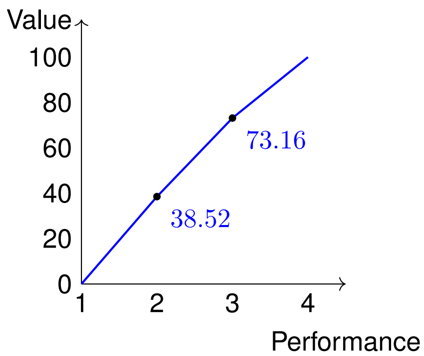


Figure 4.7: Resulting value function

The resulting function, shown in Figure 4.7, is somewhat between the two functions. The function is concave (and not s-seated, linear or convex).

Approach b: Minimize the square error between the value functions and a function with delta property

In this approach, the squared error between the value functions and a function with delta property is minimized: The partial delta value function is defined as:

$$\frac{1 - exp[-(x_i - x_i^0)/\rho]}{1 - exp[-(x_i^+ - x_i^0)/\rho]} \times 100, \rho \neq infinity$$
$$\frac{x_i - x_i^0}{x_i^+ - x_i^0} \times 100, otherwise$$

In this case, $x_i^0 = 1$, and $x_i^+ = 4$,, so the partial delta value function is:

$$v(x_i) = \frac{1 - exp[-(x_i - 1)/\rho]}{1 - exp[-(4 - 1)/\rho]} \times 100, \rho \neq infinity$$

The mean squared error, which is the loss function to be minimized, is:

$$\mathcal{L} = MSE = \frac{1}{n}\sum_{i=1}^{n}(v(x_i) - v_i)^2$$

where v_i is the value of the value function at x_i.

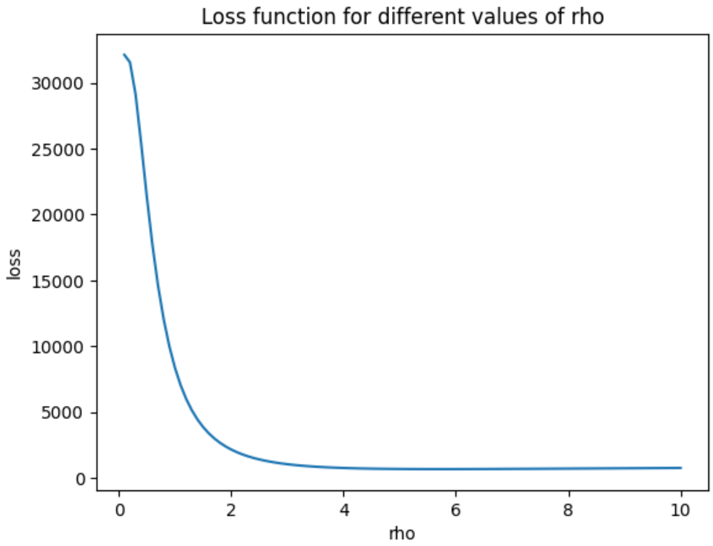


Figure 4.8: Loss function for different values of ρ. The minimum is at $\rho = 5.7$ and the root of the loss function at the minimum is $\sqrt{\mathcal{L}} = \sqrt{654.7} = 25.6$

The loss function for different values of ρ is shown in Figure 4.8.

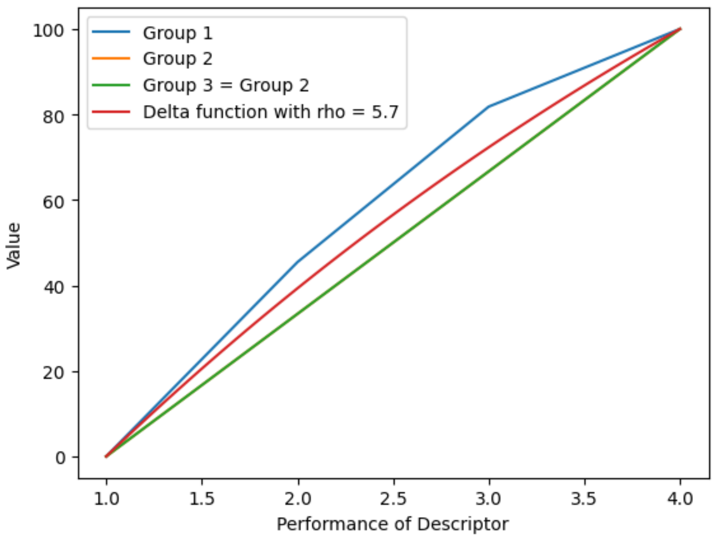


Figure 4.9: Aggregated value function and the three original value functions of the decision-makers. The aggregated value function is the delta function that minimizes the loss function defined in Equation 4.3.

If we minimize the loss function, we get the value $\rho = 5.7$. The root of loss value for this value of $\sqrt{\mathcal{L}} = \sqrt{654.7} = 25.6$. The values at performance levels 1, 2, 3 and 4 are 0, 39.32, 72.31 and 100, respectively. The aggregated value function with the value functions of the decision-makers is shown in

Figure 4.9.

Comparison of the two approaches a and b

In this subsection,the two approaches a and b outlined above are compared.

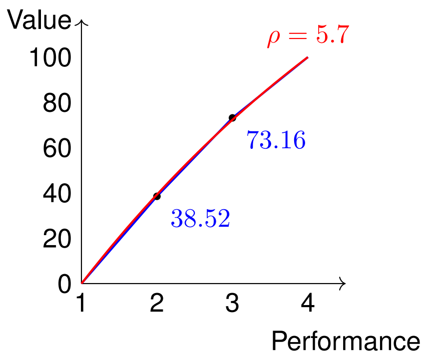


Figure 4.10: Comparison of the value functions resulting from the two approaches. The piecewise linear function defined by the average of the points of the value functions (Approach 1a) is shown in blue and the delta function that minimizes the loss function defined in Equation 4.3 (Approach 1b) is shown in red.

The two functions resulting from different methods seem similar, however, one of them is defined piece-wise and the other is a continuous curve.

Table 4.5: Results of the two approaches along with original data

Decision-makers Group	N	Value function shape	Scale
1	3	Concave	100, 81.82, 45.45, 0
2	2	Linear	100, 66.67, 33.33, 0
3	2	Linear	100, 66.67, 33.33, 0
Approach 1a	7	Concave	100, 73.16, 38.52, 0
Approach 1b	7	Concave	100, 72.31, 39.32, 0

Analysis of correction of outliers through an example

In this subsection, the method of fitting the data to a function with the delta property after correcting for outliers is applied to a different dataset with more decision-makers.

Consider 11 decision-makers who filled out a matrix of MACBETH semantical judgements. A value function was generated for each decision-maker. The resulting value functions are shown in Table 4.6.

In this case, the majority of decision-makers gave judgements that originated convex value functions. Then, there is the option to correct the concave value functions to linear as explained in Fig. 4.6. The corrected data is shown in Table 4.6 (b).

The points of the value functions after the correction of the outliers were fitted to a function with the delta property. The resulting value function is shown in Figure 4.12.

Table 4.6: Value functions for 11 different decision-makers. For each value function, the shape and the scale for the different 4 levels of performance are shown. Table (a) shows the original data and Table (b) shows the data after the outliers are corrected. In this case, the majority of decision-makers gave judgements that originated convex value functions. All the concave value functions were corrected to linear (shown in blue in Table (b)).

(a)

ID	Shape of value function	lev4	lev3	lev2	lev1
1	Concave	100	81.82	63.64	0
2	Concave	100	81.82	54.55	0
3	Concave	100	81.82	54.55	0
4	Concave	100	81.82	45.45	0
5	Linear	100	66.67	33.33	0
6	Linear	100	66.67	33.33	0
7	Convex	100	54.55	27.27	0
8	Convex	100	54.55	27.27	0
9	Convex	100	54.55	18.18	0
10	Convex	100	54.55	18.18	0
11	Convex	100	54.55	9.09	0

(b)

ID	Shape of value function	lev4	lev3	lev2	lev1
1	Linear	100	66.67	33.33	0
2	Linear	100	66.67	33.33	0
3	Linear	100	66.67	33.33	0
4	Linear	100	66.67	33.33	0
5	Linear	100	66.67	33.33	0
6	Linear	100	66.67	33.33	0
7	Convex	100	54.55	27.27	0
8	Convex	100	54.55	27.27	0
9	Convex	100	54.55	18.18	0
10	Convex	100	54.55	18.18	0
11	Convex	100	54.55	9.09	0

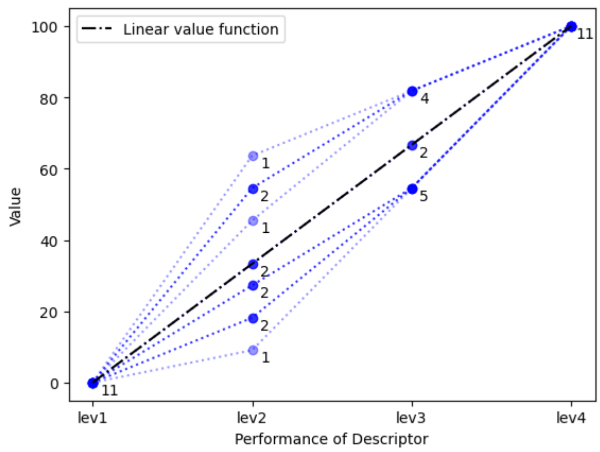


Figure 4.11: Value functions for different decision-makers. The label on each point shows how many data points are in those coordinates.

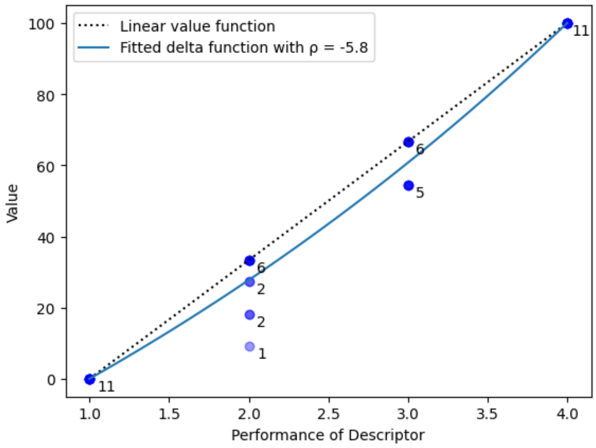


Figure 4.12: Fitted delta value function for the data in Figure 4.11. The label on each point shows how many data points are in those coordinates. The resulting value function is concave.

4.4 Stability of Incomplete Matrices

The objective was to investigate methods to quantify and mitigate the uncertainty and instability associated with incomplete judgement matrices in MACBETH analyses.

The number of judgements required tends to grow with the number of options. However, not all possible matrices of judgements are consistent. For example, for a 3x3 matrix, there are 216 possible matrices of judgements, but only 91 are ordinally consistent. The number of cardinally consistent matrices is even lower, as not all ordinally consistent matrices can be transformed into cardinally consistent matrices (see Table 4.7 for some examples of the number of possibilities for different matrix sizes).

Table 4.7: Number of cells and possibilities for different matrix sizes

Size	No. of Cells/-judgements	No. of Possible Matrices	No. of Ordinally Consistent Matrices	No. of Cardinally Consistent Matrices
2×2	1	6	6	6
3×3	3	216	91	91
4×4	6	46656	2548	2548
5×5	10	60466176	111384	...
6×6	15	4,702E+11	...	...
7×7	21	2,194E+16	...	...
8×8	28	6,141E+21	...	...
9×9	36	1,031E+28	...	...
10×10	45	1,039E+35	...	...
...	...	...	...	
n×n	$\frac{n(n-1)}{2}$	$6^{\frac{n(n-1)}{2}}$	...	...

Exploration of Incomplete Matrices filled by Transitivity through an example

Table 4.8: Matrix of judgements with cell labels

	lev 1	lev 2	lev 3	lev 4
lev 1	-	j12	j13	j14
lev 2		-	j23	j24
lev 3			-	j34
lev 4				-

In this subsection, an incomplete matrix with two missing judgements was used to explore how LP-MACBETH behaves when filling in the missing judgements by transitivity in incomplete matrices.

To do so, the following question was explored: what is the value proposed by transitivity for the cell j13 using M-MACBETH when j12 = v. weak, j23 = v. weak, j24 = v. weak, j34 = v. weak, and j14 is changed between v. weak, weak, moderate, strong, v. strong, and extreme? This is the example of an incomplete matrix of judgements shown in Figure 4.8.

This example was analysed. j14 was varied between v. weak, weak, moderate, strong, v. strong, and extreme.

For each j14, the semantical categorical judgements of j13 that produced an inconsistent matrix and the semantical categorical judgement proposed by transitivity by LP-MACBETH were registered. The

Figure 4.13: Example of an incomplete matrix of judgements with two missing judgements

	lev 1	lev 2	lev 3	lev 4
lev 1	-	v. weak	j13	j14
lev 2		-	v. weak	v. weak
lev 3			-	v. weak
lev 4				-

results of the analysis are shown in Table 4.9.

For example, for j14 = very strong, the possible semantical categorical judgements of j13 are v. weak, weak, moderate, strong and very strong. The semantical categorical judgement proposed by transitivity by LP-MACBETH is moderate. The semantical categorical judgement extreme produces an inconsistent matrix.

Table 4.9: Results of the analysis. For every possible value of j14 (first column), the possible values of j13 are shown in green and the value proposed by transitivity is marked with a star *.
For example, for j14 = moderate (the third column of the present table), the possible semantical categorical judgements of j13 are v. weak, weak and moderate (marked in green). The semantical categorical judgement proposed by transitivity by LP-MACBETH is weak (marked by the star *). The semantical categorical judgements strong, v. strong and extreme produce inconsistent matrices (marked in red)

j14	v. weak		weak		moderate		strong		v. strong		extreme
v. weak	*										
weak	*										
moderate			*								
strong				*							
v. strong					*						
extreme						*					

Metrics to measure variability

Several different metrics were explored to measure variability in the value functions resulting from incomplete matrices:

- Maximum Chebyshev distance between the value functions, which is the maximum absolute difference between the value functions.
- Maximum Manhattan distance between the value functions.
- Maximum Euclidean distance between the value functions.
- Mean Absolute Deviation between each value function resulting from a matrix completion and the value function resulting from the matrix filled by transitivity with LP-MACBETH.

The metrics were analysed using 4×4 incomplete matrices.

Analysis of different metrics to measure the variability of 4×4 incomplete matrices

To do this analysis, the value functions for all the possible 4×4 matrices with all the different possible judgements were calculated. The resulting table has 2548 rows. The first 10 rows are shown in Table 4.10.

This table was then used to analyse different metrics using Microsoft Excel.

Table 4.10: Value functions for all the possible 4×4 matrix with all the different possible judgements

j12	j13	j14	j23	j24	j34	lev1	lev2	lev3	lev4
very weak	very weak	very weak	very weak	very weak	very weak	100	66,67	33,33	0
very weak	very weak	weak	very weak	very weak	very weak	100	66,67	33,33	0
very weak	very weak	weak	very weak	weak	very weak	100	75	50	0
very weak	very weak	weak	very weak	weak	weak	100	80	60	0
...	...	...	...	...	...	...	...	...	...
extreme	extreme	extreme	extreme	extreme	strong	100	62,5	25	0
extreme	extreme	extreme	extreme	extreme	very strong	100	64,71	29,41	0
extreme	extreme	extreme	extreme	extreme	extreme	100	66,67	33,33	0

Different metrics were calculated to measure the variability in the value functions resulting from incomplete matrices. This analysis was done by fixing 5 of the matrix entries and allowing the other entries to vary. All possible combinations were calculated and the first rows of the results are shown in Table 4.11.

Table 4.11: Analysis of matrix stability. 5 of the matrix indices were fixed and 1 was left to vary. All possible combinations were calculated and the resulting data is shown in the table.

j12	j13	j14	j23	j24	j34	mean1	mean2	var1	var2	std1	std2	max1	max2	min1	min2	max deviation1	max deviation2	maxChebychev	maxManhattan	maxEuclidean
1	1	1	1	1	?	66,67	33,33	0,00	0,00	0,00	0,00	66,67	33,33	66,67	33,33	0,00	0,00	0,00	0,00	0,00
1	1	1	1	?	1	66,67	33,33	0,00	0,00	0,00	0,00	66,67	33,33	66,67	33,33	0,00	0,00	0,00	0,00	0,00
1	1	1	?	1	1	66,67	33,33	0,00	0,00	0,00	0,00	66,67	33,33	66,67	33,33	0,00	0,00	0,00	0,00	0,00
1	?	1	1	1	1	66,67	33,33	0,00	0,00	4,94	4,94	66,67	45,45	54,55	33,33	12,12	12,12	12,12	29,24	17,85
?	1	1	1	1	1	66,67	33,33	0,00	0,00	0,00	0,00	66,67	33,33	66,67	33,33	0,00	0,00	0,00	0,00	0,00
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1	2	4	1	?	1	68,93	39,05	210,55	12,26	0,00	0,00	66,67	33,33	66,67	33,33	0,00	0,00	0,00	0,00	0,00
1	2	5	1	?	1	70,63	42,10	233,81	5,67	0,00	0,00	66,67	33,33	66,67	33,33	0,00	0,00	0,00	0,00	0,00
1	2	6	1	?	1	71,64	43,87	243,03	3,13	8,75	4,38	92,31	46,15	66,67	33,33	25,64	12,82	25,64	43,46	29,10

Legend: j12, ..., j34: entry in the judgement matrix; mean1, mean2: mean of the values for lev2 and lev3, respectively; var1, var2: variance of the values for lev2 and lev3, respectively; std1, std2: standard deviation of the values for lev2 and lev3, respectively; max1, max2: maximum value of the values for lev2 and lev3, respectively; min1, min2: minimum value of the values for lev2 and lev3, respectively; max deviation1, max deviation2: maximum deviation between value vectors (for all possible pairs), respectively; maxChebychev: maximum Chebychev distance between value vectors (for all possible pairs); maxManhattan: maximum Manhattan distance between value vectors (for all possible pairs); maxEuclidean: maximum Euclidean distance between value vectors (for all possible pairs).

Mean Absolute Deviation (MAD)

There are many measures of forecast accuracy, with different degrees of applicability and interpretability (Hyndman and Koehler, 2006).

Mean Absolute Deviation (MAD) is a measure of forecast error that calculates the average of the absolute errors between the forecasted values and the actual values. It weights each error equally and is not affected by the direction of the errors. MAD is calculated as follows:

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}| \tag{4.4}$$

where x_i is the actual value, $\hat{x}_i$ is the forecasted value, and n is the number of observations.

In the context of incomplete matrices, MAD can be used to assess the stability of the value functions obtained from different completions of the matrix. In this case, in the equation above, x_i would be the value function obtained from the completions of the matrix, and $\hat{x}_i$ would be the value function obtained from the original matrix. The MAD would then provide a measure of the variability of the value functions obtained from different completions of the matrix.

From the analysis and exploration of incomplete matrices, MAD was chosen to measure the variability in the value functions resulting from incomplete matrices.

Exploration of possible completions of incomplete matrices through an example

In this example, MAD was used to measure the variability of the incomplete matrix of judgements from example from Figure 3.12, shown again in Figure 4.14.

The missing value in the matrix was filled with each one of the possible values for the missing judgement. The value scores were then calculated for each of the possible completions of the matrix. The MAD of the value score was calculated for each performance level and for all the non-reference levels combined. See Figure 4.15 for the results of this analysis.

The MAD was calculated to be 4.77 for lev2, 9.57 for lev3, and 7.17 overall. The results are shown in Table 4.12.

		Options			
		Option 1	Option 2	Option 3	Option 4
Options	Option 1	—	v. weak	?	v. strong
	Option 2		—	v. weak	moderate
	Option 3			—	v. weak
	Option 4				—

Figure 4.14: Example of an incomplete matrix of judgements

Table 4.12: Mean absolute deviation between the scales proposed by MACBETH by transitivity and the possible scales for lev2, lev3 and overall, respectively

	lev2	lev3	Overall
MAD	4.77	9.57	7.17

4.5 Conception of a novel DSS

This chapter has outlined the methodological frameworks designed to address each research question, facilitating the development of robust decision-support tools that leverage the MACBETH method's capabilities. These approaches aim to enhance decision-making processes, particularly in complex, multi-criteria environments like healthcare.

To operationalize the methods outlined in the previous chapter, a web-based decision support system (DSS) was developed. The proposed web DSS provides users with a platform to input their data, perform analyses, and visualize results. Each module in the proposed web DSS corresponds to a specific methodological approach.

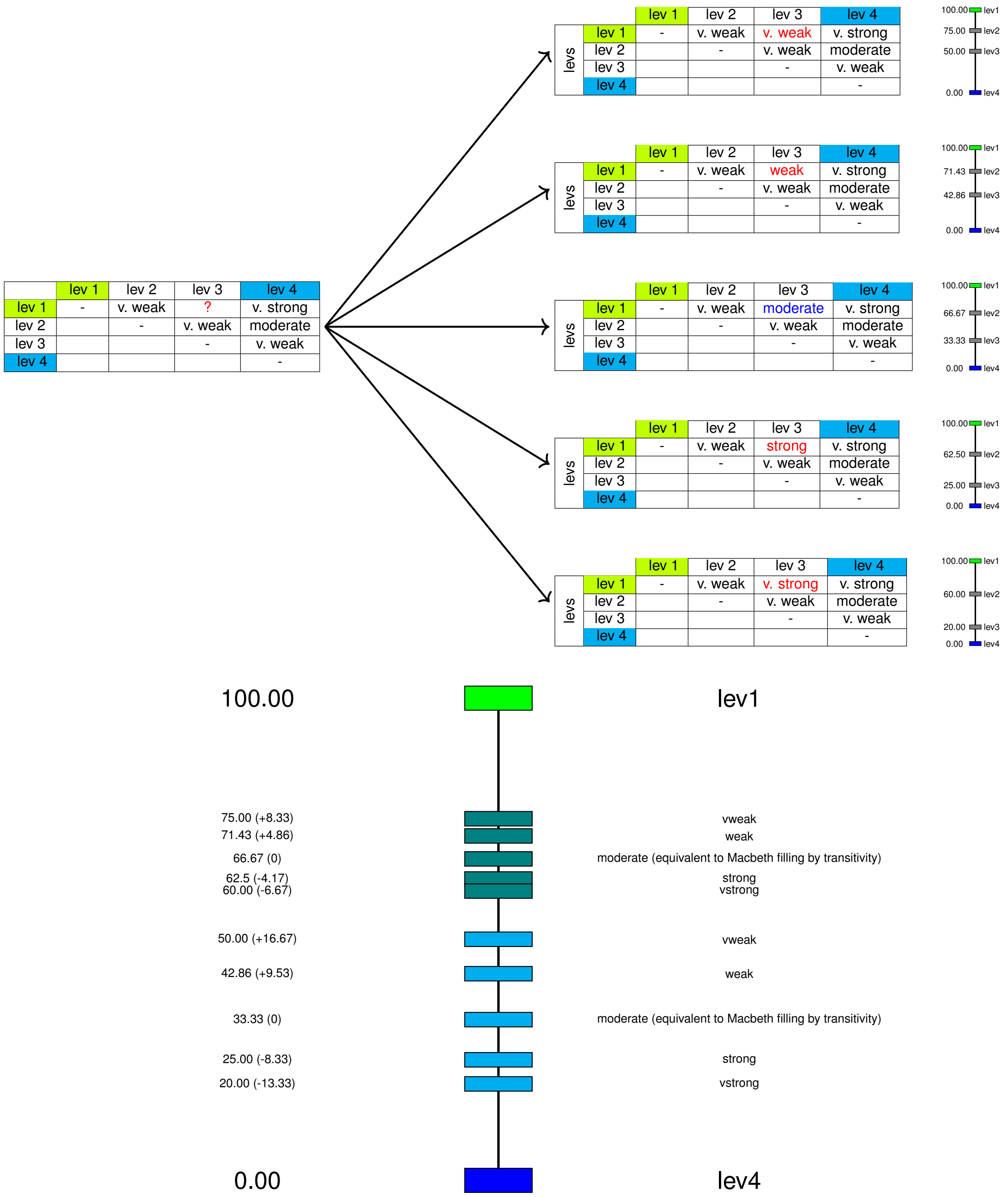


Figure 4.15: Example of an incomplete matrix of judgements. In this example, a decision-maker has not provided a judgement for Option 1 compared to Option 3. The missing value is represented by a question mark. We see that the possible values for this missing value are very weak, weak, moderate, strong, and very strong. Macbeth automatically fills in the missing judgement by transitivity with moderate. When the matrix is filled with each one of these possible values, we get a different scale. For each value of the scales, the difference between this value and the value of the scale proposed by Macbeth by transitivity was calculated. The mean absolute deviation was calculated to be 4.77 for lev2, 9.57 for lev3 and 7.17 overall.

Chapter 5

Building a decision support system

The methods outlined in the previous chapter were implemented in a Web DSS available on the website https://research.miguelroquefernandes.com/. A DSS assists users in making decisions by providing interactive tools and access to relevant data and models.

In the following sections, the conception and development of the proposed web DSS are discussed, including the design of the user interface, the architecture of the system, and the implementation of the different modules.

5.1 Conception of the Decision Support System

During the ideation and design phase of the DSS, several key considerations were taken into account to ensure the system's effectiveness and usability. These considerations included the target audience, the functionality of the system, and the user interface design. The DSS was designed to be user-friendly, intuitive, and accessible, allowing users to easily navigate the system and utilize its features. Making the system web-based enhances accessibility and usability, allowing users to access it from various devices and locations. To improve the ease of use, the DSS was designed with a clean and simple interface, with clear instructions and guidance provided to users. The system was also designed to be interactive, allowing users to input data and visualize results.

5.2 Development of the DSS

An initial mockup of the DSS was created to outline the system's structure and functionality using Figma, a design tool. The mockup was used as a reference during the development process to ensure that the final product met the design specifications. The front end of the proposed web DSS was written in TypeScript and React, using the Next.js framework. The deployment was done on Vercel, a cloud platform that allows for easy deployment and scaling of web applications with integrated Continuous Integration and Continuous Development (CI/CD) capabilities.

The backend of the proposed web DSS was written in Python using the Flask web framework. The

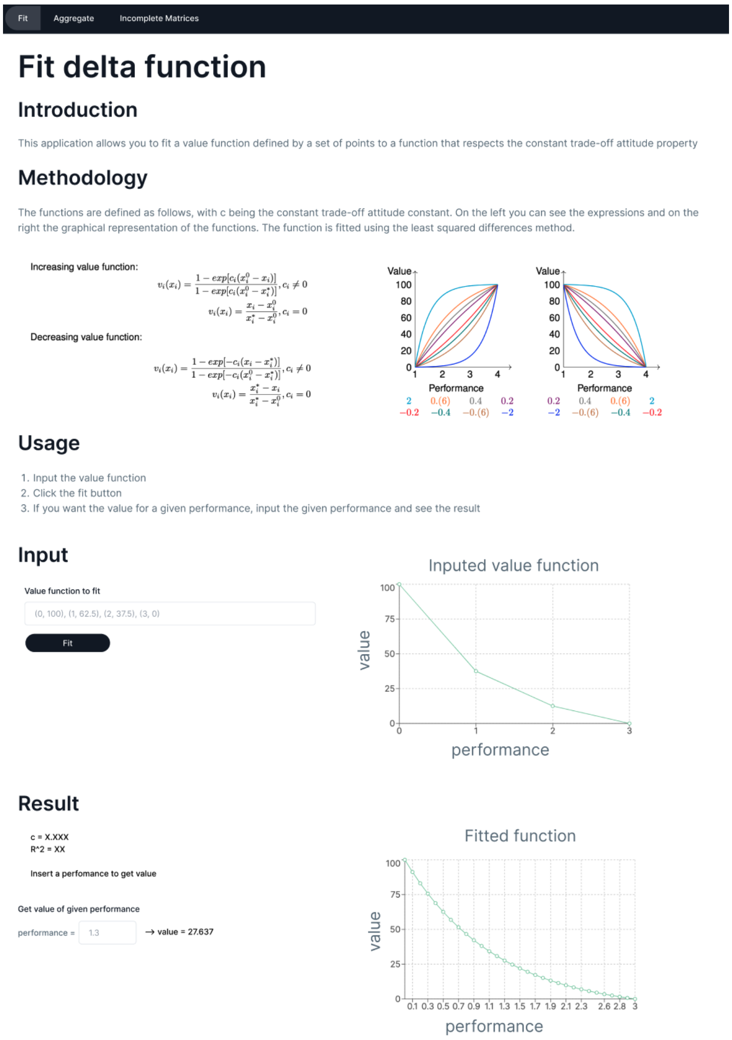


Figure 5.1: Mockup of the first module of the Decision Support System. Mockups like this one were used as a reference during the development process to ensure that the final product met the design specifications.

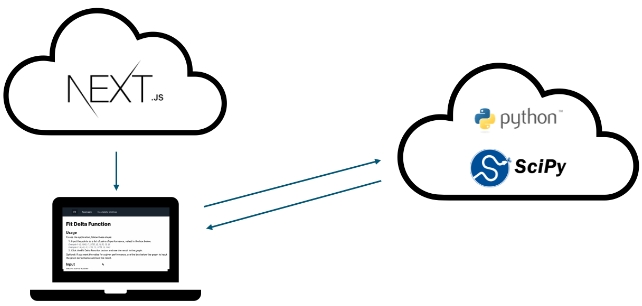


Figure 5.2: Architecture of the Decision Support System. The system consists of a front end written in TypeScript and React, and a back end written in Python using the Flask web framework.

backend handles the data processing and analysis, as well as the communication with the frontend interface. The backend was initially deployed on a virtual private server (VPS) using Docker, Jenkins and NGINX. Later, the backend was migrated to Render.com to improve scalability and reliability. The

CI/CD pipeline was set up using Jenkins to automate the deployment process and ensure the stability of the system. The libraries used in the development of the proposed web DSS include *NumPy* and *SciPy*.

NumPy is a library for numerical computing in Python, providing support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays (Harris et al., 2020).

SciPy is a library that provides additional functionality for scientific computing, including optimization (Virtanen et al., 2020).

The first and second modules of the proposed web DSS were implemented using *scipy.optimize.curve_fit* to fit the data to a function with the delta property and aggregate the data of the decision-makers. The third module was implemented using *scipy.optimize.linprog* to implement LP-MACBETH and *NumPy* was used to calculate the mean absolute deviation.

5.3 User Interface

The user interface of the proposed web DSS was designed to be user-friendly and intuitive. The interface is divided into three pages: the page to fit the data to a function with the delta property, the page to aggregate the data of the decision-makers and the page to analyse the stability of incomplete matrices.

Each page contains instructions on how to use the module and different sections for inputting data and viewing results.

When opening the proposed web DSS, the user is presented with the first module, and they can navigate to the other modules using the navigation bar at the top of the page. There are three different modules in the proposed web DSS, each with its user interface and functionality:

- Module 1: Fitting the data to a function with the delta property
- Module 2: Aggregating the data of the decision-makers
- Module 3: Analysing the stability of incomplete matrices

5.3.1 Module 1: Fitting the data to a function with the delta property

The user interface for fitting the data to a function with the delta property is shown in Figure 5.3.

The interface is divided into three sections: the usage instructions, the input form, and the result section.

In the usage instructions section, the user is provided with a brief overview of the module and instructions on how to use it.

The input form allows the user to input the data in the format of a list of (performance level, value) separated by commas. After inputting the data, the user can click the ”Fit Delta Function” button to fit the data to a function with the delta property.

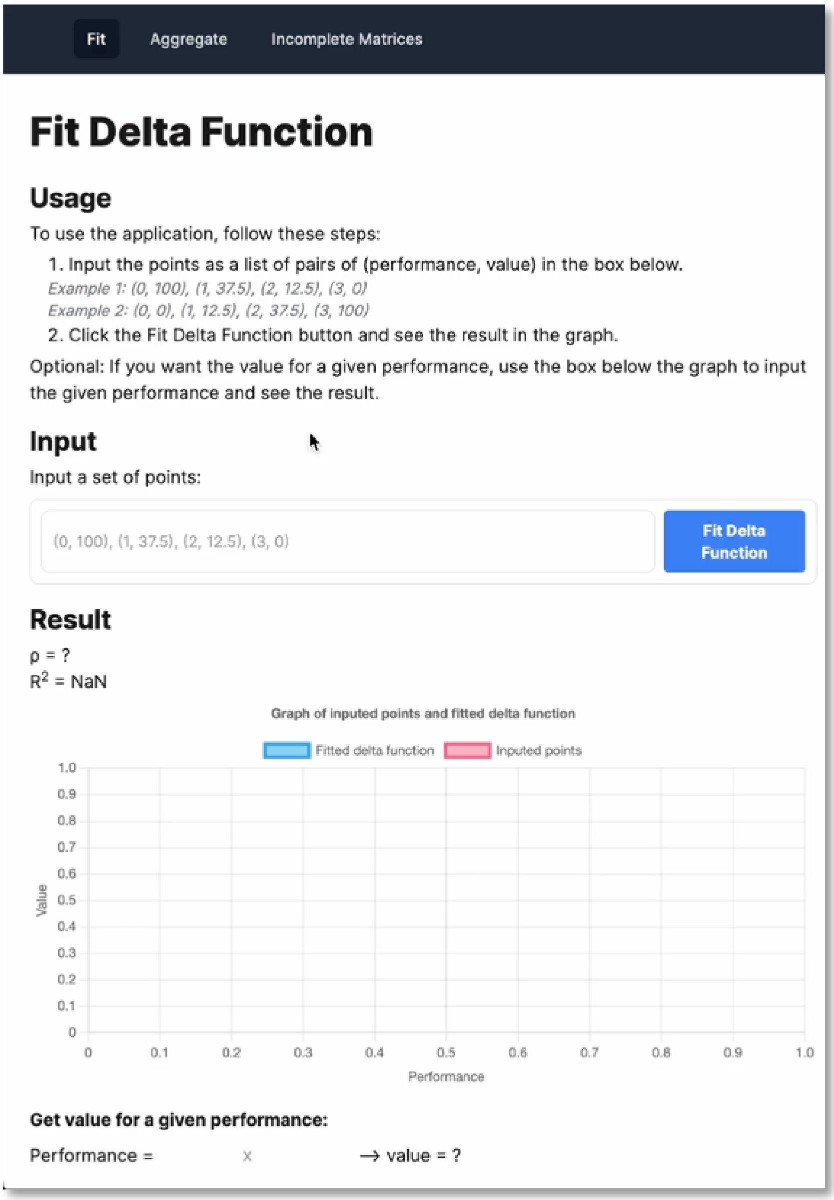


Figure 5.3: User interface of Module 1 of the proposed web DSS. This module is for fitting the data to a function with the delta property.

In the result section, the data is fitted to a function with the delta property, and the resulting function is shown in a graph. The trade-off constant c and coefficient of determination R-squared are also displayed.

The user can then hover over the graph to see the values at each performance level and choose to hide or show the data points and the fitted function by clicking on the corresponding legend items.

At the end of the result section, the user can get the value of the function at a specific performance level by inputting the performance in the text box and reading the value shown next to it.

5.3.2 Module 2: Aggregating the data of the decision-makers

The user interface for aggregating the data of the decision-makers is shown in Figure 5.4.

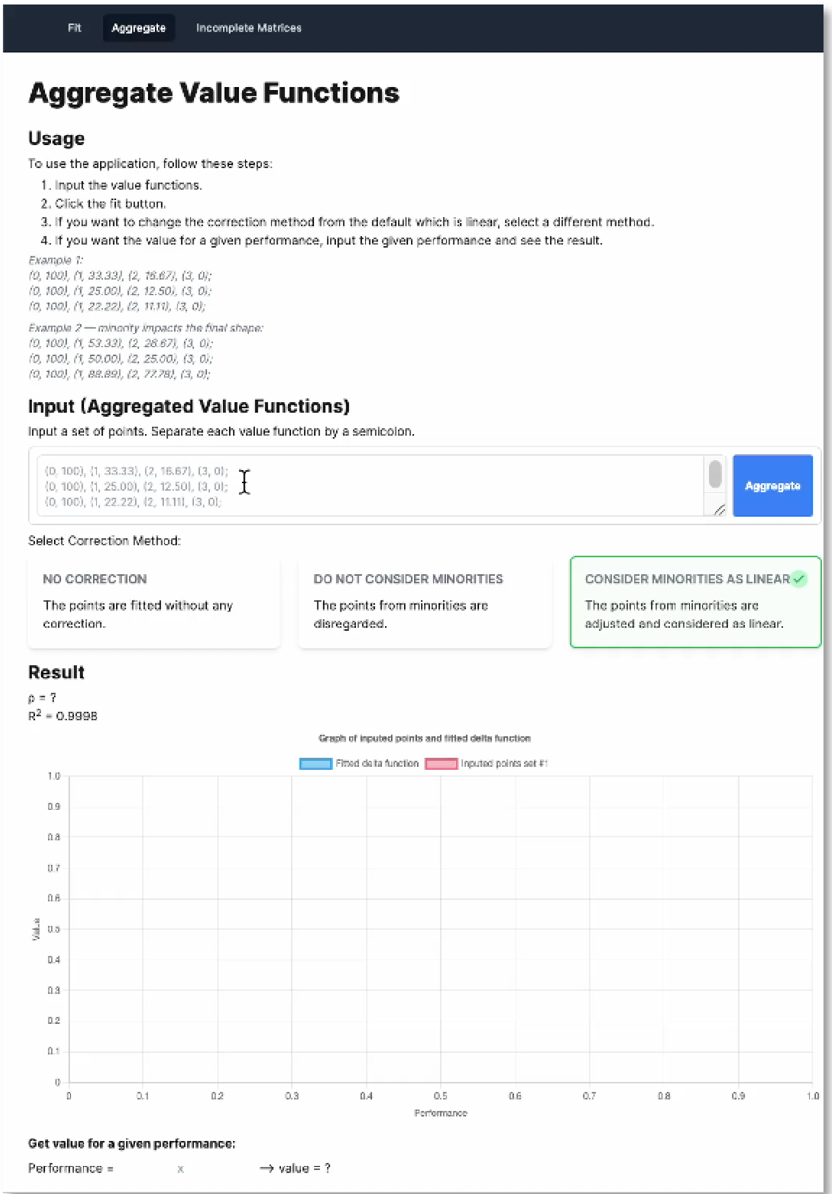


Figure 5.4: User interface for aggregating the data of the decision-makers.

Similar to the first module, the interface of Module 2 is divided into three sections, the usage instructions section provides information on how to use the module, the input form allows the user to input the data, and the result section shows the aggregated function.

The user can input the data in the format of a list of value scales. The value scale for each decision-maker needs to be separated from the others by a semicolon. Each value scale is written as a list of (performance level, value) separated by commas.

Before aggregating the data, the user can choose to correct the outliers in the data by selecting one of the correction methods.

There are three options for the correction of outliers:

- **No correction**

  No correction is done to the outliers and the function is fitted to the data as it is

- **Do not consider minorities – Correction by removing the outliers**
  The outliers are removed from the data and the function is fitted to the remaining data
- **Consider minorities as linear – Correction by replacing the outliers with a linear value function**
  The outliers are replaced by a linear value function and the function is fitted to the adjusted data

After selecting the correction method, the option box will be highlighted in green to indicate the selected option.

After clicking the ”Aggregate” button, the data is aggregated and the resulting function is shown in a graph. The trade-off constant c and coefficient of determination R-squared are also shown.

Similar to the first module, the user can hover over the graph to see the values at each performance level and choose to hide or show the data points and the aggregated function by clicking on the corresponding legend items. The user can also input a performance level to get the value of the function at that level.

5.3.3 Module 3: Analysing the stability of incomplete matrices

The user interface for analysing the stability of incomplete matrices is shown in Figure 5.5.

As with the other modules, the page starts with a usage instructions section.

The user can choose the size of the matrix and input judgements in the matrix of judgements. If one of the judgements is missing, the user can input a question mark in the corresponding matrix cell.

While changing the judgements, the thermometer and the table with values and Mean Absolute Deviation will automatically update.

To change the size of the matrix, the user can choose the size from the dropdown and the size of the matrix will be updated accordingly. For instance, if the user wants to choose a 4×4 matrix, the user can select option 4 in the dropdown and the matrix will be updated to a 4×4 matrix.

To input the matrix of judgements, the user should click each cell and select the judgement from the dropdown. When the user selects a judgement, the thermometer and the table with values and Mean Absolute Deviation will automatically update.

The thermometer, shown in Figure 5.5 to the right of the matrix, shows the value scale for the inputted judgements. The bigger green rectangles correspond to the value scale of the matrix filled by transitivity with LP-MACBETH. The smaller green rectangles correspond to the value scale of the matrices that correspond to other possible completions of the matrix.

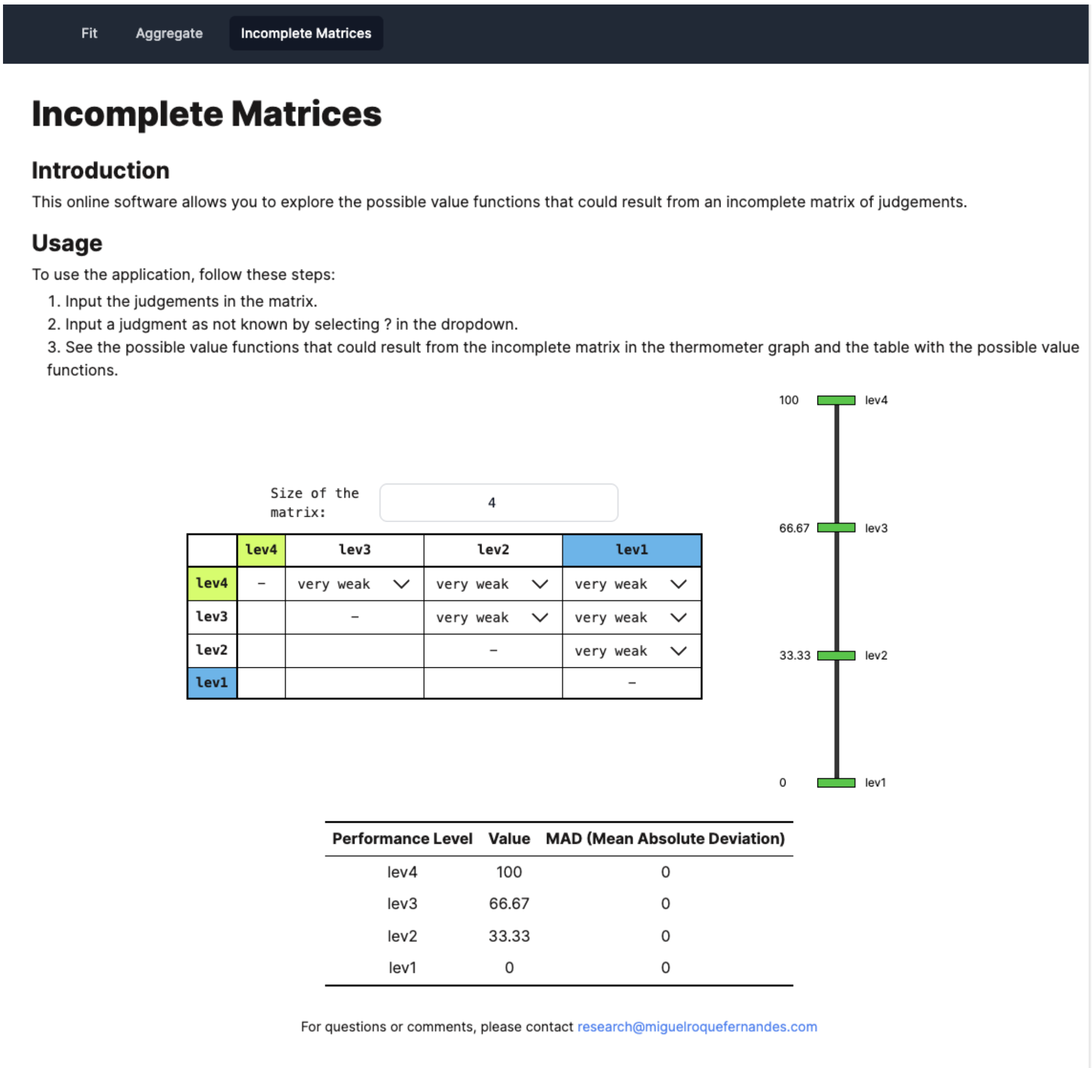


Figure 5.5: User interface for analysing the stability of incomplete matrices. In this example, a 4×4 matrix is shown filled with very weak judgements. To the right of the matrix is the thermometer showing the value scale for the inputted judgements and below the matrix is the table with values and Mean Absolute Deviation.

Chapter 6

Applications of the developed DSS

The present chapter presents the applications of the proposed web DSS in real-world decision-making problems. These problems already exist and there is a need for a solution. The proposed web DSS was designed to address these problems and provide decision-makers with a tool to make informed decisions based on data and analysis.

The developed DSS was applied in three different cases to demonstrate its functionality and utility. The cases were chosen to showcase the versatility of the DSS and its ability to address different types of decision-making problems. The cases were as follows:

- Case 1: Fitting a function following the delta property with a case previously used in the literature
- Case 2: Aggregating value functions of a criterion in the context of building a Population Health Index
- Case 3: Analysing the stability of an incomplete matrix of judgements in the context of public health decision-making

6.1 Case 1: Fitting a function for a budget in the public sector

Case 1 focuses on the first module of the proposed web DSS, which fits the data to a function with the delta property. The data used in this case was related to the EURO-HEALTHY project.

EURO-HEALTHY (Shaping EUROpean policies to promote HEALTH equitY) was an initiative aimed at creating a comprehensive Population Health Index (PHI) to analyse and address health disparities across regions in the European Union. The project employed a sociotechnical decision-aid process that combined both qualitative and quantitative data through a collaborative value-modelling framework. This approach was structured in two phases: the first involved the selection and structuring of performance indicators into a value tree using Web-Delphi methods, and the second phase involved building a hierarchical additive value-function model supported by the MACBETH decision-support system. This framework ensures the incorporation of diverse expert opinions and stakeholder inputs to create a robust, evidence-based tool for policy-making (Bana e Costa et al., 2023b).

The EURO-HEALTHY project was particularly focused on providing policymakers with actionable insights to design and implement policies aimed at improving population health and reducing health inequalities across EU regions. The final output of EURO-HEALTHY included a Web-GIS platform for interactive visualization, analysis, and comparison of health data across the EU, fostering informed discussions and decision-making on public health issues (Bana e Costa et al., 2023b).

For the present case, the method of fitting the data to a function with the delta property was applied to a set of points from a criterion of the model developed during the EURO-HEALTHY project.

The set of points was for the criterion “Expenditure on care for elderly % of GDP”. The set of points used was the same as used previously in the literature (Vieira et al., 2020).

```
(0,0), (0.9, 50), (2.3, 100)
```

The data was input in the first module of the proposed web DSS in the format of a list of (performance levels, values) separated by commas. After clicking the button “Fit Delta Function”, the data was fitted to a function with the delta property and the resulting function was shown in a graph. The trade-off constant c and coefficient of determination R-squared were also shown as seen in Figure 6.1.

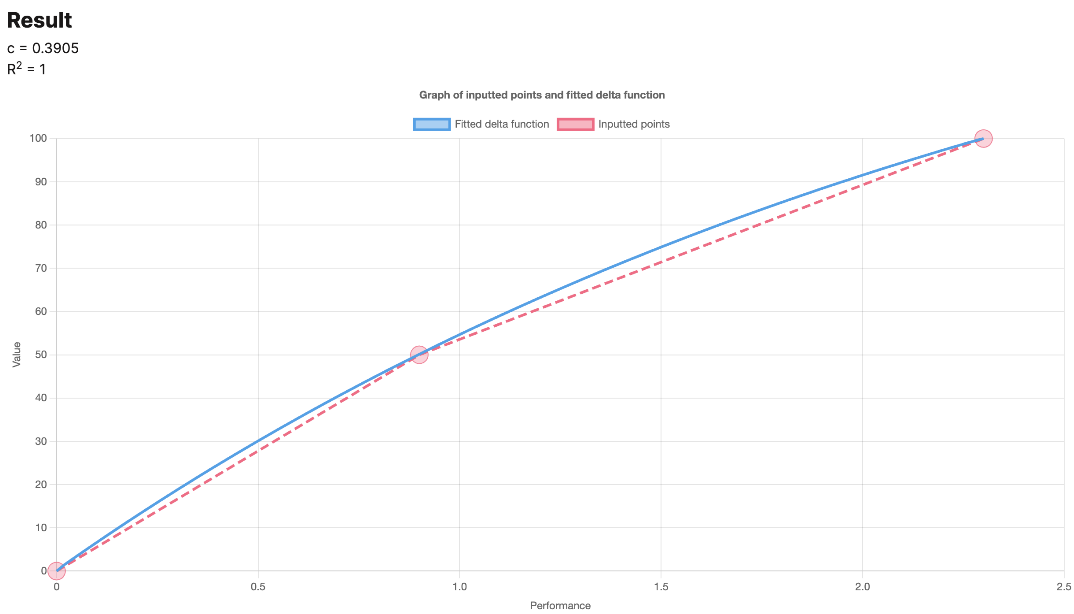


Figure 6.1: Fitting a function with the delta property for the criterion “Expenditure on care for elderly % of GDP”.

6.2 Case 2: Aggregating value functions in the context of building a Population Health Index

The method of aggregating value functions was applied to a set of value functions from the EURO-HEALTHY project (Bana e Costa et al., 2023b).

The three different options for the correction of outliers were tested: no correction, correction by removing the outliers, and correction by replacing the outliers with a linear value function.

The judgements on the difference of attractiveness between the three alternatives for the criterion 'Expenditure on care for elderly % of GDP' had been elicited in a Web-Delphi platform (Bana e Costa et al., 2023b) from 14 decision-makers. The data corresponds only to the diagonal of the matrix of judgements. Refer to Table 6.1 for the judgements on this criterion for the 14 decision-makers.

Most of the decision-makers judged the difference in attractiveness between the three alternatives as all being in the same category. For example, decision-maker 1 judged the difference between 0.8% and 0.0% as moderate, the difference between 1.5% and 0.8% as moderate, and the difference between 2.3% and 1.5% as moderate. Decision-makers 2 and 4 gave the same judgements. Decision-maker 5 gave similar judgements, categorizing the differences as strong instead of moderate. All of these judgements result in a linear value function.

For the other decision-makers, we see that the majority classified the difference between 0.8% and 0.0% in a category that expresses a greater difference than the difference between 1.5% and 0.8%. For example, decision-maker 3 classified the difference between 0.8% and 0.0% as extreme, the difference between 1.5% and 0.8% as strong, and the difference between 2.3% and 1.5% as moderate.

On the other hand, the minority expressed an opposite view, classifying the difference between 0.8% and 0.0% as smaller than the difference between 1.5% and 0.8% and the difference between 2.3% and 1.5%.

For example, Decision-maker 7 classified the difference between 0.8% and 0.0% as weak, the difference between 1.5% and 0.8% as moderate, and the difference between 2.3% and 1.5% as very strong.

Table 6.1: Extract of Web-Delphi results for the criterion "Expenditure on care for elderly % of GDP" with judgements of difference in attractiveness between the three options for the criterion for the 14 decision-makers.

Decision-maker ID	0,0 - 0,8	0,8 - 1,5	1,5 - 2,3
1	moderate	moderate	moderate
2	moderate	moderate	moderate
3	extreme	strong	moderate
4	moderate	moderate	moderate
5	strong	strong	strong
6	strong	moderate	weak
7	weak	moderate	very strong
8	extreme	strong	moderate
9	strong	moderate	moderate
10	very strong	strong	strong
11	very strong	strong	moderate
12	weak	weak	weak
13	extreme	very strong	strong
14	very strong	very strong	very strong

The columns represent the difference in attractiveness between the options. The first column is the difference between the first and second options, that is, between 0,0 and 0,8, the second column is the difference between the second and third options, that is, between 0,8 and 1,5, and the third column is the difference between the first and third options, that is, between 1,5 and 2,3.

M-MACBETH was used with the judgements of the 14 decision-makers to obtain the individual value function of each decision-maker. The matrix of judgements of each decision-maker was filled by transitivity and the value function was calculated. The value functions are shown in Table 6.2.

Table 6.2: Value scores for the criterion "Expenditure on care for elderly % of GDP" for different Decision-makers

Decision-maker ID	**0,0**	**0,8**	**1,5**	**2,3**
1	0	33,33	66,67	100
2	0	33,33	66,67	100
3	0	46,15	76,92	100
4	0	33,33	66,67	100
5	0	33,33	66,67	100
6	0	44,44	77,78	100
7	0	20,00	50,00	100
8	0	46,15	76,92	100
9	0	40,00	70,00	100
10	0	38,46	69,23	100
11	0	41,67	75,00	100
12	0	33,33	66,67	100
13	0	40,00	73,33	100
14	0	33,33	66,67	100

To allow input of the value functions in the proposed web DSS, the data was input in the format of pairs of (performance level, value) separated by commas, and with each set of value scores from each decision-maker separated by semicolons. Note that the decimal places were separated by a dot.

For case 2, the data was input as follows to the second module of the proposed web DSS:

```
(0,0), (0.8, 33.33), (1.5, 66.67), (2.3, 100);
(0,0), (0.8, 33.33), (1.5, 66.67), (2.3, 100);
(0,0), (0.8, 46.15), (1.5, 76.92), (2.3, 100);
(0,0), (0.8, 33.33), (1.5, 66.67), (2.3, 100);
(0,0), (0.8, 33.33), (1.5, 66.67), (2.3, 100);
(0,0), (0.8, 44.44), (1.5, 77.78), (2.3, 100);
(0,0), (0.8, 20.00), (1.5, 50.00), (2.3, 100);
(0,0), (0.8, 46.15), (1.5, 76.92), (2.3, 100);
(0,0), (0.8, 40.00), (1.5, 70.00), (2.3, 100);
(0,0), (0.8, 38.46), (1.5, 69.23), (2.3, 100);
(0,0), (0.8, 41.67), (1.5, 75.00), (2.3, 100);
(0,0), (0.8, 33.33), (1.5, 66.67), (2.3, 100);
(0,0), (0.8, 40.00), (1.5, 73.33), (2.3, 100);
(0,0), (0.8, 33.33), (1.5, 66.67), (2.3, 100);
```

After clicking the button 'Aggregate', the 14 value functions of the decision makers and the aggregated value function were plotted in a graph. The aggregated value function was hidden in the DSS by clicking on the legend item, as this will be explained in the next paragraphs. See Figure 6.2 for a screenshot with the individual value functions of the decision-makers as shown in the proposed web DSS.

The three options for the correction of outliers were tested: no correction, correction by removing the outliers, and correction by replacing the outliers with a linear value function.

The results of the aggregation of the value functions for the three different options for the correction of

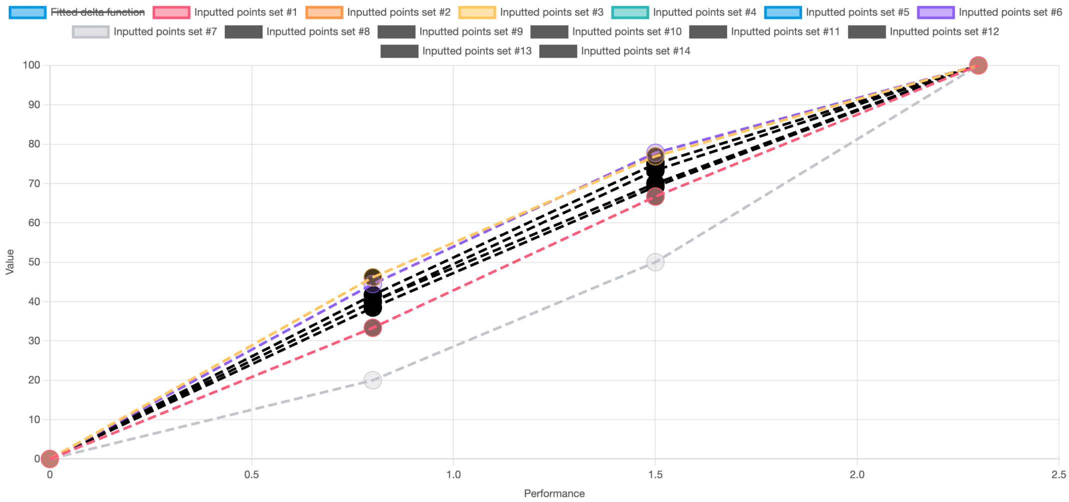


Figure 6.2: Visual representation of the individual value functions of the decision-makers. The fitted value function was hidden in the DSS by clicking on the legend item.

outliers are shown in Figures 6.3, 6.4, and 6.5. The trade-off constant c and coefficient of determination R-squared were also shown in the proposed web DSS.

The first figure, Figure 6.3, shows the aggregation of the value functions without correction of outliers. For this option, the trade-off constant c was 0.1169 and the coefficient of determination R-squared was 0.9833.

The second figure, Figure 6.4, shows the aggregation of the value functions with the correction of outliers by removing the outliers. For this option, the trade-off constant c was 0.1699 and the coefficient of determination R-squared was 0.9916.

The third figure, Figure 6.5, shows the aggregation of the value functions with the correction of outliers by replacing the outliers with a linear value function. For this option, the trade-off constant c was 0.1578 and the coefficient of determination R-squared was 0.9917.

In an increasing function, positive trade-off constants result in concave functions and negative trade-off constants result in convex functions. Examining Table 6.2 or Figure 6.2, it is possible to see that the value functions are increasing and that most of the value functions are concave.

As expected, the trade-off constant c was higher for the options that perform a correction on the outliers. This is because the correction of outliers changes the shape of the value function of the outliers, making them more similar to the other value functions.

Also as was expected, the coefficient of determination R-squared was higher for the options that perform a correction on the outliers. This is because the correction of outliers makes the value functions more similar to each other, which results in a fit of the aggregated value function to the data with a higher coefficient of determination R-squared.

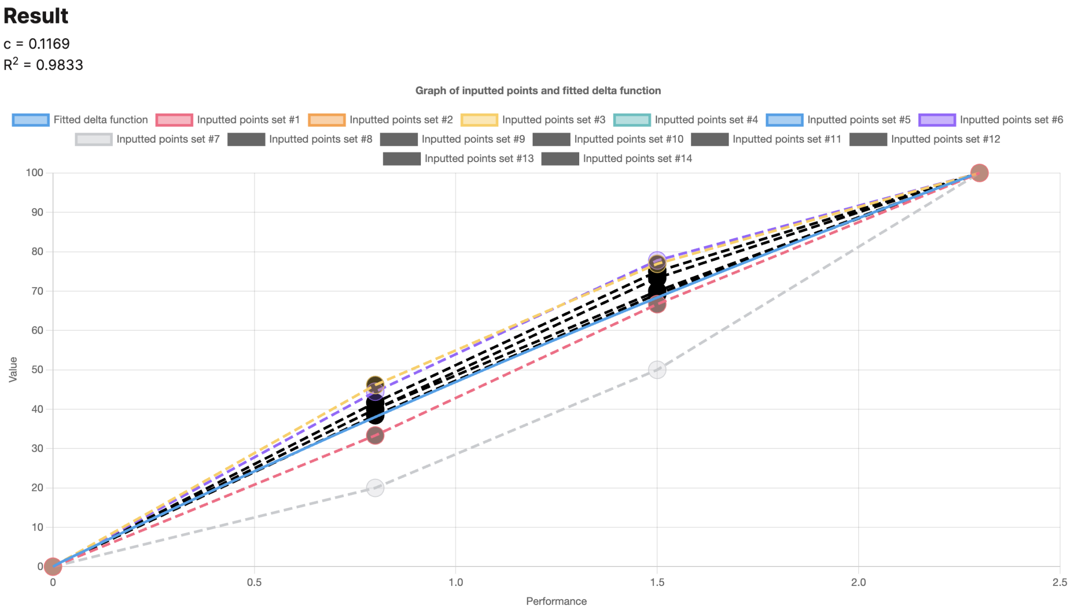


Figure 6.3: Aggregating value functions without correction of outliers. The trade-off constant c was 0.1169 and the coefficient of determination R-squared was 0.9833.

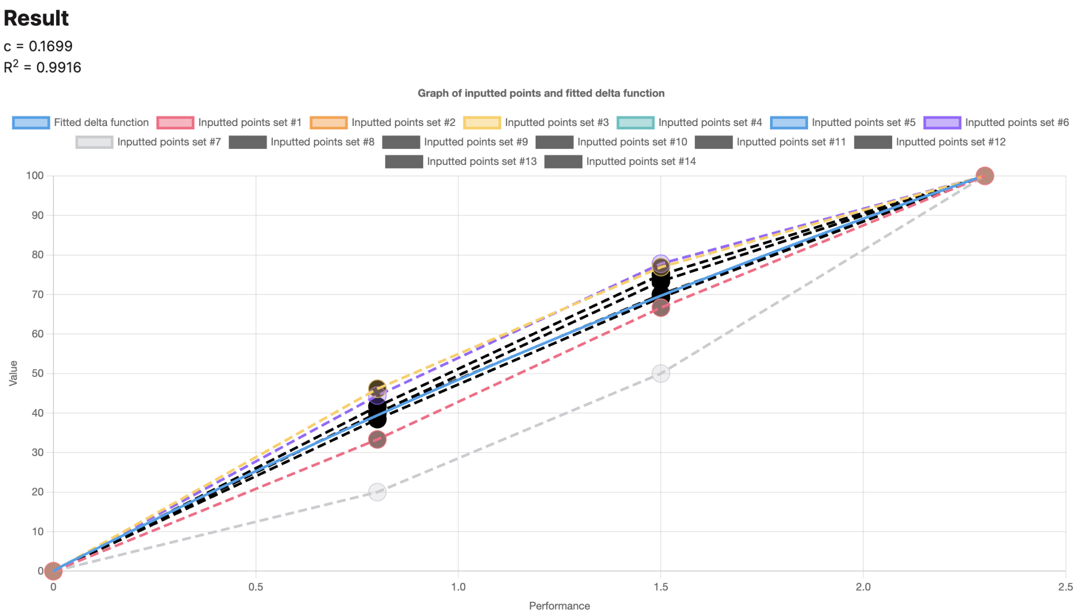


Figure 6.4: Aggregating value functions with correction of outliers by removing the outliers. The trade-off constant c was 0.1699 and the coefficient of determination R-squared was 0.9916.

6.3 Case 3: Incomplete matrix of judgements in the context of a healthcare decision

This case demonstrates the application of the third module of the proposed web DSS, which is used to analyse the stability of incomplete matrices.

A matrix of judgements from one of the decision-makers from the EURO-HEALTHY project for the criterion ”Expenditure on care for elderly % of GDP” was selected (Bana e Costa et al., 2023b). Note

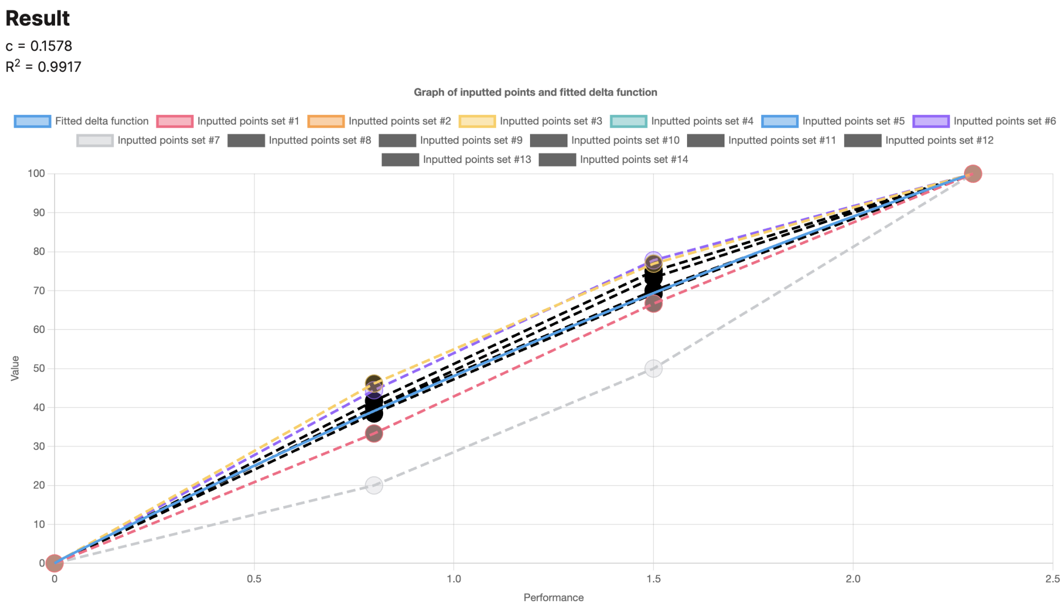


Figure 6.5: Aggregating value functions with correction of outliers by replacing the outliers with a linear value function. The trade-off constant c was 0.1578 and the coefficient of determination R-squared was 0.9917.

that this is also one of the matrices of judgements from the previous case. The matrix of judgements selected was the one from the decision-maker 6. The judgements of the decision-maker are shown in Table 6.3 and the matrix of judgements is shown in Table 6.4.

Table 6.3: Matrix of judgements of decision-maker 6 for the criterion "Expenditure on care for elderly % of GDP"

Decision-maker ID	**0,0 - 0,8**	**0,8 - 1,5**	**1,5 - 2,3**
6	strong	moderate	weak

Table 6.4: Incomplete matrix of judgements of decision-maker 6 for the criterion "Expenditure on care for elderly % of GDP"

	2.3	1.5	0.8	0
2.3	-	weak	?	?
1.5		-	moderate	?
0.8			-	strong
0				-

The possible completions of the matrix are:

This matrix was input into the third module of the proposed web DSS as can be seen in Figure 6.6. The matrix size chosen in the dropdown box was 4, as the size of the matrix of judgements of decision-maker 6 is 4x4. Next, the matrix entries were input into the matrix of judgements by clicking each cell and selecting the corresponding judgement from the dropdown menu. For the missing judgements, the option with the question mark was selected.

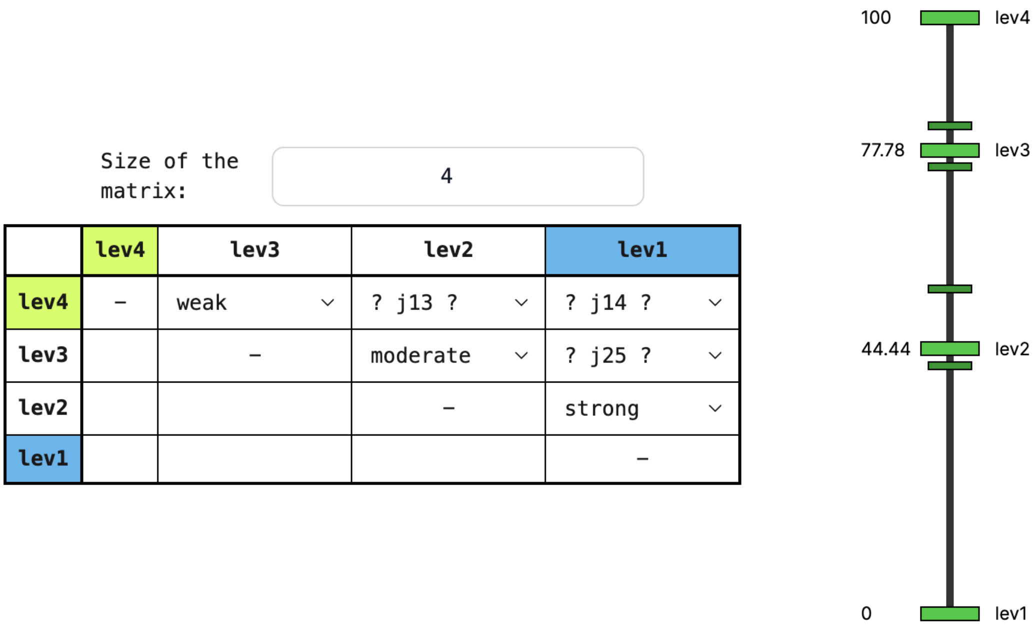
After inputting the data, the thermometer on the right and the table on the bottom with the performance levels, values and Mean Absolute Deviation (MAD) automatically updated and the results were shown in the proposed web DSS, as shown in Figure 6.6. The bigger rectangles in the thermometer

Table 6.5: Completions of the incomplete matrix and respective value scores

j13	j14	j25	lev1	lev2	lev3	lev4
moderate	strong	strong	0	54.55	81.82	100
moderate	very strong	strong	0	54.55	81.82	100
moderate	very strong	very strong	0	54.55	81.82	100
moderate	extreme	strong	0	54.55	81.82	100
moderate	extreme	very strong	0	54.55	81.82	100
moderate	extreme	extreme	0	54.55	81.82	100
strong	strong	strong	0	44.44	77.78	100
strong	very strong	strong	0	44.44	77.78	100
strong	very strong	very strong	0	44.44	77.78	100
strong	extreme	strong	0	44.44	77.78	100
strong	extreme	very strong	0	44.44	77.78	100
strong	extreme	extreme	0	44.44	77.78	100
very strong	very strong	very strong	0	44.44	77.78	100
very strong	extreme	very strong	0	44.44	77.78	100
very strong	extreme	extreme	0	44.44	77.78	100
extreme	extreme	extreme	0	41.67	75	100

represent the points of the value function as given by MACBETH (filling the matrix by transitivity). The smaller rectangles represent the points of value functions created with other possible completions of the matrix.

The MAD was calculated by the DSS for each level of performance. For levels lev1 and lev4, the MAD was 0, for level lev2 the MAD was 3.96 and for level lev3 the MAD was 1.69. The MAD for level lev2 was higher than the MAD for level lev3.



Performance Level	Value	MAD (Mean Absolute Deviation)
lev4	100	0
lev3	77.78	1.69
lev2	44.44	3.96
lev1	0	0

Figure 6.6: Input of the incomplete matrix of judgements of decision-maker 6 for the criterion “Expenditure on care for elderly % of GDP”. The bigger rectangles in the thermometer represent the points of the value function as given by MACBETH (filling the matrix by transitivity). The smaller rectangles represent the points of value functions created with other possible completions of the matrix.

Chapter 7

Discussion and Conclusions

This final chapter provides a comprehensive review of the findings and the insights gained from this research. The chapter recapitulates the central achievements of the thesis, articulating how these contribute to the broadening of the knowledge in this field. Furthermore, it provides an examination of potential developments and further research that could extend and enhance this work.

The research conducted has made contributions to curve fitting of value functions, value function reconciliation and the analysis of incomplete judgement matrices in the context of MACBETH. The findings have implications for decision-making processes in various domains including healthcare.

Using curve fitting techniques to fit value functions to a function with the delta property has provided insights into the advantages, limitations, and challenges of this approach. This approach can be used to expand the application of MACBETH to a wider range of decision-making scenarios and enhance the quality of decisions made by individuals and organizations.

Reconciliation of value functions can be a necessity when dealing with multiple decision-makers and limited resources to gather more data. This work explored different methodologies to aggregate value functions and provided a comparison of the results obtained from these methodologies. The approach that provided the best results was the use of a loss function with the squares of the differences in the points to fit a delta function and the correction of outliers was left as an option for the user.

Incomplete matrices are common in real-world decision-making scenarios, and the tool developed can be utilized to analyse the stability of incomplete matrices and provide insights into the potential outcomes associated with incomplete data. The research associated with incomplete data and the developed methodologies to address those issues have provided insights into the challenges (incomplete matrices can have very different degrees of stability) and opportunities (the tool developed can be used to research new methodologies to address incomplete matrices and improve the quality of decisions).

The proposed methodologies for fitting value functions, reconciling value functions, and analysing incomplete matrices can be used in cases similar to the ones presented in this work, that is, processes that involve multiple decision-makers and incomplete matrices of judgements.

7.1 Discussion of the results

There are several strengths and limitations to the research conducted in this thesis. The strengths include the development of a web Decision Support System that is user-friendly and intuitive, freely available online, and versatile, allowing for the application of different decision-making scenarios. For example, all three modules of the DSS can be applied to matrices of judgements of different sizes and the second module of the DSS is flexible, allowing the user to choose between three options for the correction of outliers: no correction, correction by removing the outliers, and correction by replacing the outliers with a linear value function. The third module of the DSS allows the exploration of matrices of MACBETH judgements that can generate insights into the stability of incomplete matrices and the potential outcomes associated with incomplete data.

The limitations of the research include the fact that the DSS is limited to the methodologies developed in this work, which means that in applications, it needs to be used in conjunction with other DSS. For example, the curve fitting module only allows the user to fit the data to a function with the delta property, and the aggregation module only allows the user to aggregate value functions with three options for the correction of outliers.

Another limitation is the lack of the ability to import or export data in a file format like CSV or Excel. This would allow the user to save the data and results and use them in other tools or share them with other users.

Future work could expand the functionality of the DSS to include additional methodologies and options for the user to explore.

7.2 Summary and Key Achievements

The present work aimed to address three main challenges within the MACBETH framework: curve fitting, aggregate value functions, and measuring the variability of incomplete matrices of judgements.

The key achievements of the present work are:

- Identification and selection of challenges in MCDA in HTA that can be addressed with new methodologies and tools.
- Development of a methodology to fit data to a function with the delta property.
- Development of a methodology to aggregate value functions.
- Development of a methodology to analyse the stability of incomplete matrices.
- Implementation of LP-MACBETH, the formulation of the linear programming problem of the conditions to create value functions from a matrix of MACBETH judgements in the Python programming language.
- Development of a web Decision Support System to apply all three methodologies developed, with a user-friendly interface that allows the input of data and visualization of results.

- Application of the methodologies developed to real-world cases in the context of building a Population Health Index, a case in the literature where the proposed tools and methodologies are useful.

7.3 Future Work

The present work has made several advancements in the area of MCDA in HTA.

However, it also opens up several avenues for future research. The most important ones are listed below.

- Investigation of the optimal number of judgements to ask the decision-maker — A critical area of future research would involve investigating the optimal number of judgements that should be ideally asked of the decision-maker. This line of investigation could pave the way for developing more efficient and effective decision-making models.

- Investigation of the impact of the number of decision-makers on the results of the aggregation of value functions — Another area of future research could involve investigating the impact of the number of decision-makers on the results of the aggregation of value functions. This could help decision-makers understand the implications of involving more or fewer decision-makers in the decision-making process.

- Investigation of the impact of incomplete matrices on the decision-making process in various domains — Investigate how incomplete matrices of judgements need to be addressed in different decision-making scenarios and how the methodologies developed in this work can be adapted to address these challenges. Research how to define recommendations for thresholds of uncertainty in the decision-making process for different domains.

- Indicate which missing judgements reduce variability the most — Another area of future research could involve identifying which missing judgements would reduce the uncertainty or variability in the decision-making process the most. This could help decision-makers prioritize which judgements to focus on when completing an incomplete matrix. This could be done by modifying the third module of the DSS to show the impact of each missing judgement on the variability of the matrix for instance with a color gradient. The methodology could be similar to the one used in the third module of the DSS and it would be more computationally intensive.

- Implement tools from the DSS into more complete DSS — The tools developed in the proposed web DSS could be integrated into more comprehensive decision support systems (such as M-MACBETH) to enhance their functionality and usability.

- Iterate the DSS with feedback from users — The DSS could be iterated based on feedback from users to improve its intuitiveness, user-friendliness, usability and effectiveness.

In conclusion, while this thesis has filled some gaps in the understanding of judgements in multi-criteria decision analysis, there are still uncharted territories to explore. Future investigations, coupled with the constant evolution of technology, are likely to open up novel perspectives and bring about more accurate and efficient decision-making processes. This continuous progress is crucial in a world where quality decisions often serve as the bedrock for successful outcomes in various fields ranging from business and economics to social and environmental planning.

Bibliography

S. Afonso, A. C. Vieira, C. Pereira, and M. D. Oliveira. Advancing hospital-based health technology assessment: evaluating genomic panel contracting strategies for blood tumors through a multimethodology. *International Journal of Technology Assessment in Health Care*, 39(1):e76, 2023. ISSN 0266-4623, 1471-6348. doi: 10.1017/S0266462323002751. URL https://www.cambridge.org/core/product/identifier/S0266462323002751/type/journal_article.

A. Angelis and P. Kanavos. Value-Based Assessment of New Medical Technologies: Towards a Robust Methodological Framework for the Application of Multiple Criteria Decision Analysis in the Context of Health Technology Assessment. *PharmacoEconomics*, 34(5):435–446, May 2016. ISSN 1179-2027. doi: 10.1007/s40273-015-0370-z. URL https://doi.org/10.1007/s40273-015-0370-z.

A. Angelis and P. Kanavos. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework. *Social Science & Medicine*, 188:137–156, Sept. 2017. ISSN 0277-9536. doi: 10.1016/j.socscimed.2017.06.024. URL https://www.sciencedirect.com/science/article/pii/S0277953617303933.

C. Bana e Costa, L. Angulo-Meza, and M. Oliveira. O MÉTODO MACBETH E APLICAÇÃO NO BRASIL. Carlos A. Bana e Costa1 Lidia Angulo-Meza2 Mónica D. Oliveira3 ENGEVISTA - UFF - BRASIL. *ENGEVISTA*, 15, Jan. 2013. doi: 10.22409/engevista.v15i1.484.

C. A. Bana e Costa and M. D. Oliveira. A multicriteria decision analysis model for faculty evaluation. *Omega*, 40(4):424–436, Aug. 2012. ISSN 03050483. doi: 10.1016/j.omega.2011.08.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0305048311001265.

C. A. Bana e Costa and J.-C. Vansnick. MACBETH — An interactive path towards the construction of cardinal value functions. *International Transactions in Operational Research*, 1(4):489–500, Oct. 1994. ISSN 0969-6016. doi: 10.1016/0969-6016(94)90010-8. URL https://www.sciencedirect.com/science/article/pii/0969601694900108.

C. A. Bana e Costa, J. C. Lourenço, M. D. Oliveira, and J. C. Bana e Costa. A Socio-technical Approach for Group Decision Support in Public Strategic Planning: The Pernambuco PPA Case. *Group Decision and Negotiation*, 23(1):5–29, Jan. 2014. ISSN 0926-2644, 1572-9907. doi: 10.1007/s10726-012-9326-2. URL http://link.springer.com/10.1007/s10726-012-9326-2.

C. A. Bana e Costa, J.-M. De Corte, and J.-C. Vansnick. On the Mathematical Foundations of MACBETH. In S. Greco, M. Ehrgott, and J. R. Figueira, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, International Series in Operations Research & Management Science, pages 421–463. Springer, New York, NY, 2016. ISBN 978-1-4939-3094-4. doi: 10.1007/978-1-4939-3094-4_11. URL https://doi.org/10.1007/978-1-4939-3094-4_11.

C. A. Bana e Costa, J.-M. De Corte, and J.-C. Vansnick. M-MACBETH User's Guide. Technical report, 2017. URL https://m-macbeth.com/wp-content/uploads/2017/10/M-MACBETH-Users-Guide.pdf.

C. A. Bana e Costa, J.-M. De Corte, and J.-C. Vansnick. THE MACBETH APPROACH. page 41, 2022. Type: article.

C. A. Bana e Costa, M. D. Oliveira, A. C. L. Vieira, L. Freitas, T. C. Rodrigues, J. Bana e Costa, A. Freitas, and P. Santana. Collaborative development of composite indices from qualitative value judgements: The EURO-HEALTHY Population Health Index model. *European Journal of Operational Research*, 305(1):475–492, Feb. 2023a. ISSN 0377-2217. doi: 10.1016/j.ejor.2022.05.037. URL https://www.sciencedirect.com/science/article/pii/S037722172200426X.

C. A. Bana e Costa, M. D. Oliveira, A. C. L. Vieira, L. Freitas, T. C. Rodrigues, J. C. B. e Costa, Â. Freitas, and P. Santana. Collaborative development of composite indices from qualitative value judgements: The EURO-HEALTHY population health index model. *Eur. J. Oper. Res.*, 305(1):475–492, 2023b. doi: 10.1016/J.EJOR.2022.05.037. URL https://doi.org/10.1016/j.ejor.2022.05.037.

D. Beiderbeck, N. Frevel, H. A. von der Gracht, S. L. Schmidt, and V. M. Schweitzer. Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX*, 8:101401, May 2021. ISSN 2215-0161. doi: 10.1016/j.mex.2021.101401. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8374446/.

V. Belton and T. J. Stewart. *Multiple Criteria Decision Analysis*. Springer US, Boston, MA, 2002. ISBN 978-1-4613-5582-3 978-1-4615-1495-4. doi: 10.1007/978-1-4615-1495-4. URL http://link.springer.com/10.1007/978-1-4615-1495-4.

A. Beyer and B. Fasolo. Application of MACBETH to the elicitation of preferences for treatment outcomes (Revised 2014).

A. C. w. W. Charnes. *Management Models and Industrial Applications of Linear Programming*. Wiley, Jan. 1961. ISBN 978-0-471-14850-0.

R. M. Cooke and R. M. Cooke. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Environmental Ethics and Science Policy Series. Oxford University Press, Oxford, New York, Jan. 1992. ISBN 978-0-19-506465-0.

J. L. Corner. Operationalizing approximate multiattribute utility functions for use in practice. *European Journal of Operational Research*, 79(1):73–84, Nov. 1994. ISSN 03772217.

doi: 10.1016/0377-2217(94)90396-4. URL https://linkinghub.elsevier.com/retrieve/pii/0377221794903964.

C. A. B. E. Costa, A. C. Vieira, M. Nóbrega, A. Quintino, M. D. Oliveira, and J. B. E. Costa. Collaborative Value Modelling in corporate contexts with MACBETH. *Procedia Computer Science*, 162:786–794, 2019. ISSN 18770509. doi: 10.1016/j.procs.2019.12.051. URL https://linkinghub.elsevier.com/retrieve/pii/S1877050919320629.

B. Fasolo and C. A. Bana e Costa. Tailoring value elicitation to decision makers' numeracy and fluency: Expressing value judgments in numbers or words. *Omega*, 44:83–90, Apr. 2014. ISSN 03050483. doi: 10.1016/j.omega.2013.09.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0305048313000960.

F. A. F. Ferreira and S. P. Santos. Two decades on the MACBETH approach: a bibliometric analysis. *Annals of Operations Research*, 296(1):901–925, Jan. 2021. ISSN 1572-9338. doi: 10.1007/s10479-018-3083-9. URL https://doi.org/10.1007/s10479-018-3083-9.

D. for Communities and Local Government: London. Multi-criteria analysis: a manual, Jan. 2019. URL https://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf.

P. Gongora-Salazar, S. Rocks, P. Fahr, O. Rivero-Arias, and A. Tsiachristas. The Use of Multicriteria Decision Analysis to Support Decision Making in Healthcare: An Updated Systematic Literature Review. *Value in Health*, Nov. 2022. ISSN 1098-3015. doi: 10.1016/j.jval.2022.11.007. URL https://www.sciencedirect.com/science/article/pii/S1098301522047386.

S. Greco, M. Ehrgott, and J. R. Figueira, editors. *Multiple Criteria Decision Analysis: State of the Art Surveys*, volume 233 of *International Series in Operations Research & Management Science*. Springer New York, New York, NY, 2016. ISBN 978-1-4939-3093-7 978-1-4939-3094-4. doi: 10.1007/978-1-4939-3094-4. URL https://link.springer.com/10.1007/978-1-4939-3094-4.

C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

J. C. Harsanyi. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy*, 63(4):309–321, 1955. URL http://www.jstor.org/stable/1827128.

J. M. Hummel, M. D. Oliveira, C. A. Bana e Costa, and M. J. IJzerman. Supporting the Project Portfolio Selection Decision of Research and Development Investments by Means of Multi-Criteria Resource Allocation Modelling. In K. Marsh, M. Goetghebeur, P. Thokala, and R. Baltussen, editors, *Multi-Criteria Decision Analysis to Support Healthcare Decisions*, pages 89–103. Springer International Publishing, Cham, 2017. ISBN 978-3-319-47538-7 978-3-319-47540-0. doi: 10.1007/978-3-319-47540-0_6. URL `http://link.springer.com/10.1007/978-3-319-47540-0_6`.

R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, Oct. 2006. ISSN 01692070. doi: 10.1016/j.ijforecast.2006.03.001. URL `https://linkinghub.elsevier.com/retrieve/pii/S0169207006000239`.

R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: Preferences and value tradeoffs*. Decisions with multiple objectives: Preferences and value tradeoffs. Cambridge University Press, New York, NY, US, 1976. ISBN 978-0-521-44185-8 978-0-521-43883-4. Pages: xix, 569.

C. W. Kirkwood. *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets*. Cengage Learning, Belmont, 1ª edição edition, July 1996. ISBN 978-0-534-51692-5.

C. List. The theory of judgment aggregation: an introductory review. *Synthese*, 187(1):179–207, July 2012. ISSN 0039-7857, 1573-0964. doi: 10.1007/s11229-011-0025-3. URL `http://link.springer.com/10.1007/s11229-011-0025-3`.

R. J. G. Mateus, J. C. Bana E Costa, and P. V. Matos. Supporting Multicriteria Group Decisions with MACBETH Tools: Selection of Sustainable Brownfield Redevelopment Actions. *Group Decision and Negotiation*, 26(3):495–521, May 2017. ISSN 0926-2644, 1572-9907. doi: 10.1007/s10726-016-9501-y. URL `http://link.springer.com/10.1007/s10726-016-9501-y`.

M. D. Oliveira, A. Agostinho, L. Ferreira, P. Nicola, and C. Bana e Costa. Valuing health states: is the MACBETH approach useful for valuing EQ-5D-3L health states? *Health and Quality of Life Outcomes*, 16(1):235, Dec. 2018. ISSN 1477-7525. doi: 10.1186/s12955-018-1056-y. URL `https://doi.org/10.1186/s12955-018-1056-y`.

M. D. Oliveira, I. Mataloto, and P. Kanavos. Multi-criteria decision analysis for health technology assessment: addressing methodological challenges to improve the state of the art. *The European journal of health economics: HEPAC: health economics in prevention and care*, 20(6):891–918, Aug. 2019. ISSN 1618-7601. doi: 10.1007/s10198-019-01052-3.

G. S. Parnell, editor. *Handbook of decision analysis*. Wiley, Hoboken, New Jersey, 2012. ISBN 978-1-118-51584-6 978-1-118-51583-9 978-1-118-51582-2. Medium: electronic resource.

L. D. Phillips. Decision analysis in 2005. *OR47: Keynote and Tutorials. The OR Society: Birmingham*, 2005.

L. D. Phillips and C. A. Bana e Costa. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations Research*, 154 (1):51–68, July 2007. ISSN 0254-5330, 1572-9338. doi: 10.1007/s10479-007-0183-3. URL `http://link.springer.com/10.1007/s10479-007-0183-3`.

L. D. Phillips, P. Humphreys, D. Embrey, and D. L. Selby. A socio-technical approach to assessing human reliability. pages 253–276. John Wiley & Sons, West Sussex, UK, 1990. ISBN 978-0-471-92381-7. URL http://eu.wiley.com/. Num Pages: 500.

T. L. Saaty. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281, June 1977. ISSN 00222496. doi: 10.1016/0022-2496(77)90033-5. URL https://linkinghub.elsevier.com/retrieve/pii/0022249677900335.

T. L. Saaty. Analytic Hierarchy Process. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd, 2005. ISBN 978-0-470-01181-2. doi: 10.1002/0470011815.b2a4a002. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011815.b2a4a002. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470011815.b2a4a002.

C. Sorenson, M. Drummond, and P. Kanavos. ENSURING VALUE FOR MONEY IN HEALTH CARE : The role of health technology assessment in the European Union. 2008.

R. H. Sprague. A framework for the development of decision support systems. *MIS Quarterly*, 4(4): 1–26, 1980. ISSN 02767783. URL http://www.jstor.org/stable/248957.

E. G. Trevelyan and P. N. Robinson. Delphi methodology in health research: how to do it? *European Journal of Integrative Medicine*, 7(4):423–428, 2015. ISSN 1876-3820. doi: https://doi.org/10.1016/j.eujim.2015.07.002. URL https://www.sciencedirect.com/science/article/pii/S1876382015300160. Diagnostic Techniques and Outcome Measures for Integrated Health.

A. C. Vieira, M. D. Oliveira, and C. A. Bana e Costa. Enhancing knowledge construction processes within multicriteria decision analysis: The Collaborative Value Modelling framework. *Omega*, 94:102047, July 2020. ISSN 03050483. doi: 10.1016/j.omega.2019.03.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0305048318312532.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

D. v. Winterfeldt and W. Edwards. *Decision Analysis and Behavioral Research*. Cambridge University Press, Aug. 1986. ISBN 978-0-521-25308-6. Google-Books-ID: eUo7AAAAIAAJ.