Assessment of Sustainability-Driven Operational Innovations in the Airline Industry

A Multi-Criteria Decision-Making Perspective

Master thesis submitted to Delft University of Technology
in partial fulfillment of the requirements for the degree of

Master of Science
in Management of Technology

Faculty of Technology, Policy and Management

by

Glenn De Decker
Student number: 5184800

To be defended in public on September 23rd, 2021

First Supervisor:	Dr.ir. Z. Roosenboom-Kwee, Section Economics of Technology and Innovation
Second Supervisor & Chair:	Dr. J. Rezaei, Section Transport and Logistics
External Supervisor:	M. van den Ham, MSc, Transavia Airlines CV

Keywords: Airline strategy, sustainability, innovation management, multi-criteria decision-making, Best-Worst Method

Executive Summary

The European aviation sector agrees that a mix of several smaller operational innovations collectively have great potential to help shortening the pathway towards net zero aviation. Nevertheless, from the operator's point of view, hardly any insights have been provided on how the sustainability perspective aligns with industry specific corporate objectives. In order to improve the effectiveness of sustainability-driven decision-making in airlines, this study aimed to address this gap. Specifically, the goal was to view the evaluation of operational innovations as a multi-criteria decision-making (MCDM) problem. To that end, a case study was conducted at Transavia, the low-cost arm of Europe's third largest airline group.

From discussions with seven company experts and relevant literature, the definitions of the main criteria to support the evaluation of operational improvements in flight operations were synthesized. Using the Best-Worst Method (BWM) developed by Rezaei (2015), inter-criteria preference data from seventeen senior and executive airline managers were collected. The pairwise comparison vectors were analyzed through the aggregation of individual preferences as well as probabilistic modeling.

Results from both approaches showed that "Safety" plays by far the most important role in the evaluation of operational innovations. It is followed by the four criteria, "Employee Experience", "Passenger Experience", "Finance", and "Sustainability", between which there seems to be no evident relative preference based on the aggregation of individual data. From the probabilistic approach could be concluded with relative certainty that Finance ≻ Sustainability ≻ Employee Experience ≻ Passenger Experience. This order showed conformity to an overall intuitive ranking provided by a separate group of decision-makers.

Still, neither the aggregation of individual data, nor the probabilistic model revealed practically meaningful differences between the magnitudes of the weights assigned to the four intermediate criteria. Both approaches established however, that the criterion related to on-time performance (OTP) is clearly the least important.

The fact that the Sustainability dimension takes second place to safety in the appraisal of eco-focused innovation projects, may seem strange to some outsiders. However, given the legal responsibilities of some of the managers involved, and ongoing campaigns that promote the firm's safety policy, the results are in fact not unexpected. What is more interesting, is the question of how the firm will balance these two dimensions, which are sometimes diametrically opposed. It is expected that through increased global focus on sustainability, the dichotomy will only become more relevant in the coming years. Accordingly, a supportive framework provided in this report gives the opportunity to connect the safety perspective with sustainability-driven initiatives.

Surprisingly, the findings from the aggregated preferences of airline leadership also reveal that the importance placed on the passenger and employee dimensions is almost equal to the relative preference for finance and sustainability. Several of the consulted

decision-makers argue that through perceived correlations, all other dimensions will eventually benefit from increased satisfaction among passengers and crew. The long-term relationship between satisfaction and profitability is partly supported by the literature. Although both customers and employees might eventually demand sustainable business practices, it is doubtful whether this also assures that enough attention is paid to sustainability in the short to medium-term. This could be an interesting direction for future research.

In the meantime, the operator is recommended to reflect on the importance placed on sustainability – which seems not to be particularly prominent among the other corporate objectives that vie for attention – and decide whether priorities need to be reevaluated. For airlines like Transavia, which have been expanding their capacity and resulting carbon-intensity in the past decades, these considerations are especially important. In an industry where economies of scale are the endgame, the current relative preference for sustainability might – in light of regulatory mechanisms that impose a CO2-ceiling with a reference year that lies before the period of accelerated capacity growth – pose a threat to long-term competitiveness. In that regard, ongoing campaigns may be required to promote the firm's sustainability policy in a similar way as is currently done for safety.

The results presented in this study can be interpreted three ways:

1. As a high-level framework to assess the priority of sustainability relative to the five other dimensions vital to the firm's operations. In this interpretation, the results can be used by airline managers to develop an effective corporate strategy for the appraisal of operational innovations in flight operations;

2. As the conceptual foundation for the quantitative comparison of a set of projects with respect to the definitions of the six perspectives presented in this report;

3. Separately, knowing the relative preference from key decision-makers can help other practitioners to improve the performance of their innovative efforts based on the six dimensions. Concretely, if practitioners want to introduce a new eco-oriented operational improvement, they should also explicitly highlight the performance of the alternative's attributes in the safety, employee experience, and passenger experience dimensions.

ACRONYMS

- **A/C** Aircraft.
- **AHP** Analytic Hierarchy Process.
- **AIP** Aggregation of Individual Priorities.
- **ALARP** As Low As Reasonably Practicable.
- **ANP** Analytic Network Process.
- **AVF** Additive Value Function.
- **BWM** Best-Worst Method.
- **CASK** Cost per Available Seat Kilometer.
- **CBA** Cost-Benefit-Analysis.
- **COG** Center of Gravity.
- **COI** Current Operating Income.
- **DCF** Discounted Cash Flow.
- **DEMATEL** Decision-Making Trial and Evaluation Laboratory.
- **DM** Decision-Maker.
- **ELECTRE** ELimination and Choice Expressing REality.
- **ETS** Emission Trading System.
- **EXI** Employee Experience (Index).
- **GHG** Greenhouse Gas.
- **HV** IATA code of Transavia Airlines CV.
- **IATA** International Air Transport Association.
- **IPA** Importance - Performance Analysis.
- **KPI** Key Performance Indicator.
- **LCC** Low Cost Carrier.
- **MADM** Multi-Attribute Decision-Making.
- **MCDM** Multi-Criteria Decision-Making.
- **MCMC** Markov Chain Monte Carlo.
- **MODM** Multi-Objective Decision-Making.
- **NOX** Nitrogen Oxides.
- **NPV** Net Present Value.
- **OTP** On-Time Performance.
- **PROMETHEE** Preference Ranking Organization METHod for Enrichment of Evaluations.
- **PSI** Project Safety Index.

- **PXI** Passenger Experience (Index).
- **RI** Risk Index.
- **ROA** Real Options Analysis.
- **ROE** Return On Investment.
- **SAF** Sustainable Aviation Fuel.
- **TAT** Turnaround Time.
- **TOPSIS** Technique for Order of Preference by Similarity to Ideal Solution.
- **UGT** Unscheduled Ground Time.
- **VC** Venture Capital.
- **VIKOR** Multicriteria Optimization and Compromise Solution (from Serbian; "VIseKriterijumska Optimizacija I Kompromisno Resenje").
- **WSM** Weighted Sum Model.
- **capex** Capital Expenditures.
- **opex** Operational Expenditures.

List of Figures

List of Tables

Contents

1

INTRODUCTION

FROM an outsider's perspective the airline industry may look like it is on the leading edge of technological innovation. However, a highly *optimized* industry like aviation should not be mistaken for a high-technology industry. The basic design of the aircraft flying today date back to the 1960s, the airlines' business models to the 1990s, and large parts of their IT infrastructure to the 2000s. In stark contrast to the tech giants, margins on airfares are shockingly thin. In addition, the airline industry is extremely conservative. The undivided attention of the industry goes towards compliance to standard operating procedures. This mentality yielded the industry the most impressive safety record in all of transportation. But it is innately conflicting with innovative efforts.

Fortunately, the future of air transport as the preferred mode for medium to long-haul transit is at least partly safeguarded by its conceptual superiority over alternatives. For the ground-based competition, the majority of the cost of operating a network comes from the links between nodes, making it an increasingly less efficient substitute when distance increases (van de Weijer, 2019). For aviation, those links are made of air and therefore practically free. Partly because of this, air traffic is expected to continue its growth in the coming years, despite the temporary downturn caused by the pandemic (Uitbeijerse, 2020).

However, the downsides of air transport pose important preconditions on the future survival of airlines. These downsides include CO2, particulate matter, nitrogen, and noise emissions. As a result, the public support for the aviation sector is diminishing (Ministry of Infrastructure and Water Management, 2019). While there is uncertainty in terms of the regulatory mechanisms, it is clear that externalities will not remain unpriced. This means that an industry with an innate belief in slow, incremental change and limited resources will have to learn to innovate quickly in order to respond to an increasingly louder call for sustainable business practices.

1

1.1. PROBLEM DEFINITION

This research project was issued by the Flight Operations Support & Innovation department of Transavia Airlines CV (herein Transavia and sometimes referred to by their IATA code HV), a Dutch low-cost carrier (LCC). This department is tasked with selecting sustainable innovation projects in the operational domain. Projects grouped in this category are limited to the way in which aircraft are used by the airline and are generally characterized by low asset-intensity. This means that drastic changes to the physical aircraft are out of scope. Nevertheless there are numerous smaller operational improvements that together have great potential to reduce CO2 emissions (SEO Economics and Netherlands Aerospace Centre, 2021). For Transavia examples of such operational improvements include small aerodynamic enhancements for the fleet, a switch to lighter materials (e.g. in passenger seats or brake discs) or an AI tool to accurately predict the weight of the passengers on board, which allows for a better estimation of the most fuel-efficient altitude and airspeed combination. With these types of projects, the airline aims at an 8% CO2 reduction by 2030.

Currently, the selection of projects is done through a combination of practical experience (the term "gut feeling" was used by one of the practitioners) and traditional discounted cash flow (DCF) methods. However, it was clear from preliminary meetings with senior managers and an initial literature scan that DCF methods do not sufficiently capture the complexity of the problem. They convey a false sense of unambiguity, while in reality they are only as accurate as the original assumptions. Furthermore, these methods would underestimate the potential of projects of the high-risk, high reward type, and often fail to capture the long-term strategic importance of the investment decision (Schilling, 2017).

To respond to these gaps, some scholars have promoted the idea of treating innovation projects as real call options (Bowman and Hurry, 1993; Chan et al., 2007; Perlitz et al., 2002; Schilling, 1998). However, the further one deviates from derivatives pricing – the original application of real options analysis (ROA) – the more inaccurate the analogy with financial options becomes. This gets especially problematic when the total risk profile is dominated by project-specific risks rather than market risk (Steffens and Douglas, 2007).

Even if this could somehow be accounted for, a purely financial proxy for project success will not be suitable in the context of sustainable aviation. The airline does not rule out a scenario where accepting potential carbon taxes, emission rights purchases, and carbon compensation obligations is on paper more cost-effective than investing in actual mitigation. However, it fears that this strategy will impact its long-term competitiveness in a market where eco-minded travelers will increasingly weigh sustainability efforts into their ticket purchasing decisions. Furthermore, the airline currently depends on state-backed loans to its holding company Air France - KLM and is afraid that political goodwill once lost could be very hard to regain.

Aside from economics, sustainability, brand equity, and public relations, there are several other business objectives that should be included in the decision-making process. In its high-level strategic goals, the firm uses the "SPOEFS" dimensions. SPOEFS is an acronym for Safety, Passenger experience (PXI), On-time performance (OTP), Employee experience (EXI), Finance, and the most recently added KPI; Sustainability. However, this framework has not yet been operationalized at the project level. Additionally, the framework contains conflicting objectives both within and between domains that need to be made explicit to the decision-maker. For instance, specifying a higher minimum acceptable altitude at which the landing gear should be lowered will score well across safety KPIs but also increases the drag of the aircraft which leads to additional greenhouse gas (GHG) and noise emissions. An example of a paradox within a domain is the well-known fact that more fuel efficient, less carbon emitting engines tend to produce higher amounts of nitrogen oxides (NOX) due to increased pressures and temperatures in the combustor.

These are just two of the many examples, highlighting that the six variables cannot be optimized for peak performance at the same time. Meanwhile, solving for these variables independently may yield sub-optimal results overall. In order to leverage new opportunities that consider all aspects of the corporate strategy, sustainability efforts have to be evaluated as an integrated part of the airline planning.

Although many projects targeting GHG reductions have been proposed, they do not always ensure the required shift in business practices. A likely explanation is that due to parallel requirements in other domains, the sustainability dimension alone is not sufficient to ensure management buy-in. For that reason it can be valuable to determine the relative importance of the criteria from the perspective of airline leadership. Knowing which aspects of a project are viewed as more important can be beneficial to the integration of sustainability into the innovation management decision-making process. Considering the above, the problem statement can be defined as follows:

The process of selecting sustainability-related innovation projects in airlines requires complex strategic decisions that rely on multiple and often conflicting goals and involve many different departments within the firm, making it inherently a multiple-criteria decision-making (MCDM) problem, whereas it is not yet treated as such.

1.2. RESEARCH APPROACH

The choice for using an MCDM approach to address this problem was based on the decision environment this study is meant to support. Figure 1.1 (p.4) highlights the position of the Transavia case study within the hierarchy of decision-making problems as proposed by Raiffa (1982).

It shows the two main perspectives on decision-making: i) "Individual", where the decision to invest in a project for instance, is made by a single person or an entity acting as an individual; and ii) "Plural", where there are two or more parties involved in the decision. While in this case it may appear to outsiders that the decision is made by the airline as a single entity, in reality it is the result of an explicit or implicit internal negotiation between managers with contradicting views. In this context, it is therefore more relevant to view the decision problem through the latter perspective (highlighted with the solid red line in Figure 1.1).

Subsequently, there are two sub-categories within the perspective of plural decision-making: a) "Separate and Interactive", where the payoff for one party will depend on the decision of the other party; and b) "Joint", where a joint agreement is sought by two or more parties and the consequences and payoffs are shared. In this context, the actual situation at the company is more accurately described by the joint decision-making perspective (as highlighted with the dotted red line in Figure 1.1).

Although the process does not directly involve *politicians*, it still shares some characteristics with the category of "wicked problems"[1] (defined by Rittel and Webber (1973)) that require a "political" decision-making approach.

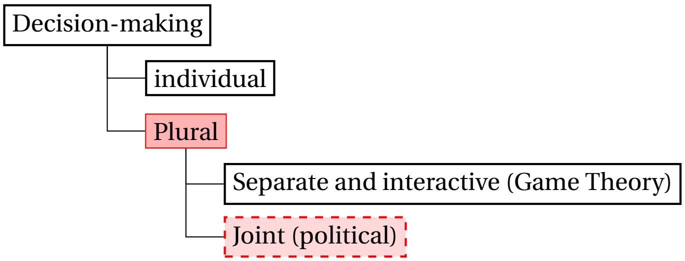

Figure 1.1: Relevant decision-making perspectives, highlighting the main (solid line) and sub-categories (dotted line) of perspectives through which the case study is viewed - by author, based on Raiffa (1982).

First of all, problem solving with regards to sustainable aviation is never definitively finished. The extent of both the positive and negative externalities is not fully known and the things that are known may change over time. This means that the decision-makers can always optimize their evaluation process when new information becomes available (for example on the "NOX-paradox" explained in section 1.1).

Second, there are no formal decision rules to determine the correctness of the decision. In other words, there are no true good or false answers to this problem. Singular focus on sustainability aspects will reduce the negative consequences for the environment in terms of noise and pollution, but it will also severely impact the company's bottom line and possibly put thousands of jobs on the line. However, in future market conditions, doing the opposite might have the same effect. What the exact balance between financial and sustainability indicators should be will always depend on who you ask.

Finally, there is no ultimate test for the solution of the problem. Any decision to allow or block the investments in operational improvements will generate waves of consequences over an extended period of time. There is no way to trace all these consequences, especially not before the decision is made. As an example, consider the fact that the vast majority of total HV capacity is dedicated to leisure traffic, a market that is to a large degree commoditized. These passengers might be indifferent as to whether they choose Transavia to get to their holiday destination or opt for a rival LCC. Possible reasons for switching behavior could be related to overprioritization of sustainable business practices (at the expense of other elements of the value proposition), but might just as likely be caused by negligence of passengers' ecological concerns.

These three characteristics of the Transavia case study imply that the problem can be meaningfully viewed from a "political" decision-making angle. Aside from MCDM, another popular approach to support project appraisal in this domain is the Cost-Benefit-Analysis (CBA). However, the limited amount of empirical data on the attitude of transport decision-makers towards appraisal tools, does not favor the CBA. Reasons for this

[1] Parts of the discussion regarding wicked problems in aviation was adapted from De Decker et al. (2020)

might be related to the DCF assumptions that are inherent to CBA. The disadvantages of DCF for complex appraisal problems have been previously discussed in section 1.1. According to Annema et al. (2015), transportation DMs find the aggregate outcome (the composite result) of CBAs "pretentious". They seem especially interested in appraisal tools which show clearly to them the political important trade-offs in transportation issues. One of the main advantages that MCDM has over CBA is that this method is able to incorporate factors which cannot be easily expressed in monetary values. This is also in line with the exploratory nature of this study, since at this stage financial proxies for some criteria might not be feasible. Other criteria may turn out to be inherently qualitative and cannot be quantified at all. An additional benefit of MCDM cited by the decision-makers consulted by Annema et al. (2015), is that it provides a structured way of incorporating the opinion of different stakeholders.

1.3. Research Objective and research questions

The goal of this study was to create an MCDM framework to assist airline decision-makers in the effective selection of process innovation projects that match the firm's sustainability agenda while also considering existing corporate objectives along the dimensions, safety, passenger experience, on-time performance, employee experience, and finance. The main research question (RQ) was defined as follows:

How can MCDM methodology support the selection of sustainability-oriented process innovations in airlines?

In order to give focus to this broader question, the following sub-questions were defined:

- Sub-RQ I: *"Which MCDM method should be used to support this process?"*
- Sub-RQ II: *"Which indicators should be used as criteria to support decision-making towards sustainability, while also considering the five other dimensions vital to the firm's operations?"*
- Sub-RQ III: *"What weights should these criteria have in the decision-making process?"*
- Sub RQ IV: *"What are the requirements for implementing a sustainability-oriented decision support system that provides decision aid to practitioners in airline operations?"*

1.4. Report structure

Table 1.1 (p.6) gives a brief overview of the data that was used to address each sub-question. It also shows how the research questions divide the remainder of this report into 4 chapters (Chapters 2 to 5). Chapter 2 provides a review of the literature on multi-criteria decision-making. In this chapter sub-RQ I is answered by selecting an MCDM technique that best fits the research objective and current understanding of the decision problem. In Chapter 3, a detailed description of the research methodology is presented based on the technique chosen in Chapter 2. Chapter 4 presents the results of the interviews, desk research, and MCDM-questionnaires and discusses their implications. This chapter answers sub-questions II and III. Finally, sub-RQ IV is answered by providing conclusions, suggestions for future research, and managerial implications related to the decision support framework (Chapter 5).

Table 1.1: Overview of the data collection methods used to address the sub-questions

Sub-RQ	Data	Type of data	Methods of data collection	Chapter
I	Primary and secondary	Qualitative	Initial discussions with HV management, desk research	2. Theoretical Background, 3. Methodology
II	Primary and secondary	Qualitative	Interviews, desk research	4. Results and Discussion
III	Primary	Quantitative	MCDM-based questionnaires	4. Results and Discussion
IV	Primary	Qualitative, quantitative	Interviews, reflection on results from RQ I-III	5. Conclusion and Recommendations

2

THEORETICAL BACKGROUND

A Literature review on MCDM should start off with the disclaimer that searching for the best MCDM method is in itself a multiple-criteria decision-making problem and therefore a paradox (Triantaphyllou and Mann, 1989). However, the purpose of this review is to find a method that best suits the research objective and current understanding of the problem, rather than identifying the optimal model (which does not exist). As economist Thomas Sowell famously said, in complex problems there are “no solutions, only trade-offs”.

MCDM is the process of determining the best feasible solution according to established criteria and problems. MCDM applications are categorized either as multi attribute decision-making (MADM), when related to decision problems that have discrete solutions, or multi objective decision-making (MODM), for problems with continuous/ infinite solutions. Thus, the evaluation of projects is technically speaking an MADM problem. However, in the literature both subsets are commonly referred to by the umbrella term ’MCDM’. The same will be done in this report.

The goal of MCDM is to improve the quality of the decision-making process by recognizing the many conflicting aspects that are to be handled simultaneously, hereby accepting that the outcome is the decision most suitable to the context, as opposed to an optimal one. MCDM generally involves four main steps; (i) structuring the decision problem, (ii) articulating and modelling the preferences, (iii) aggregating the alternative evaluations or preferences, and (iv) providing recommendations (Guitouni and Martel, 1998).

2

Although the choice for one of the many methods that exist within MCDM is highly context dependent, some generalizations can be made about the way the overall score for an alternative is calculated. The process generally takes the following form:

$$S = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_n \end{array} \\ \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{pmatrix} \end{array}, \tag{2.1}$$

where $A = \{a_1, a_2, ..., a_m\}$ indicates a set of alternatives that are up for consideration, $C = \{c_1, c_2, ..., c_n\}$ denotes a set of evaluation criteria, and s_{ij} represents the score of alternative i in relation to criterion j. As the criteria are usually not of equal importance to the overall goal of the decision-making process, they are assigned a weight $w_j (w_j \geq 0, \sum w_j = 1)$. The overall score V_i of an alternative a_m can then be calculated, for instance using the additive value function (AVF) by Keeney and Raiffa (1976):

$$V_i = \sum_{j=1}^{n} w_j s_{ij}, \ \forall = 1, 2, ..., m. \tag{2.2}$$

Whereas the general framework (2.1) holds for all variants of MCDM, important differences exist in the way in which the criteria weights $w = \{w_1, w_2, ..., w_n\}$ are determined and how the scores are aggregated.

2.1. Comparison of MCDM approaches

Comparisons between the techniques commonly used in MCDM have been made by several scholars in the fields of management science, operations research, and computer science. Examples include Stewart (1992), Guitouni and Martel (1998), Velasquez and Hester (2013).

In his review, Steward contrasted the most prevalent streams of thought in MCDM and focused explicitly on those approaches which are most robustly and effectively useable, especially by practitioners who are non-experts in the field. For MCDM problems involving this category of decision-makers, Steward recommends that inputs required from the decision-maker should be operationally meaningful and free from ambiguities of meaning. Furthermore, he advocates the translation of these inputs into partial or complete recommendations should be consistent with the inputs used and with reasonable behavioral assumptions and should be as far as possible transparent to the decision-maker. Implicit assumptions made in the process should be both justifiable and easily understood by both the DMs and the public concerned. Most of all, the method should be simple and efficient to use. While this might sound straightforward to those unfamiliar with MCDM, it will become clear by the end of this review that many approaches do not satisfy Steward's desiderata, making them less likely to prove successful outside the research setting, i.e., in context of the daily operations of the firm.

A more general set of guidelines can be found for example in Velasquez and Hester (2013). The authors examine the advantages and disadvantages of 11 identified methods and explains how their common applications relate to their relative strengths and weaknesses. Additionally, they attempt to provide a clear guide on how MCDM methods should be used in particular situations. Similarly, Guitouni and Martel (1998) outline seven guidelines to support the selection of a suitable approach relating to:

1. the number of the stakeholders involved in the MCDM process (if there are many decision-makers, group decision dynamics should be considered),
2. the preference and experience of the decision-maker with regards to criteria weight determination methods,
3. the decision problematic pursued by the decision-maker, ranking of alternatives or otherwise,
4. the degree to which the MCDM can properly handle the input information available and for which the decision-maker can easily provide the required information (quality and the quantity of the information are major factors in the choice of the method),
5. the degree of compensation that is acceptable between different dimensions or criteria, i.e., how a good evaluation on one criterion can compensate a bad one on another,
6. the degree to which the fundamental assumptions of the method can be verified in the particular situation,
7. the decision support system coming with the method.

While it seems reasonable to apply these guidelines as a general strategy, it must be noted that there is no consensus or generally accepted framework for choosing MCDM methods. It is perhaps for this reason that MCDM techniques are often selected arbitrarily and developed in an ad-hoc manner: because an analyst involved in the process is already familiar with the procedure or the technique is chosen because the software available to apply it is more accessible than software for other methods (Kornyshova and Salinesi, 2007). To address this, Kornyshova and Salinesi (2007) performed a comparative analysis of 9 different MCDM selection approaches in attempt to come up with a collection of requirements for a selection approach. While they point out that each selection technique has its advantages and disadvantages and can be more or less useful depending on the situation, they do claim based on their analysis framework that a better approach than existing ones (mind the wording here) should satisfy a number of requirements including:

- take into account the problem situation,
- allow a typology of problem characteristics,
- consider MCDM techniques specificities,
- take into account data diversity (types, scales etc.),

- consider all main groups of MCDM techniques and be able to deal with a new one,
- present a more precise estimation for parameters as alternatives number and ease of use,
- allow selecting of MCDM technique and its adaptation to a concrete case,
- take into account interaction between goals,
- be universal with regards to an application domain,
- suggest a tool facilitating MCDM technique selection.

At this point it should be stressed that – considering MCDM is a decision-problem and the selection of an MCDM-method is a meta-decision-problem – the requirements for MCDM method selection protocols deal with a third level of abstraction that is outside the scope of this report. Besides, that level is still characterized by the same inherent paradox as the original problem. It is therefore more worthwhile for both illustrative purposes as well as the general understanding of the methodology, to explore the pros and cons of some of the more popular techniques that are actually being used to solve real life decision-making problems.

2.1.1. AHP AND ANP

The Analytic Hierarchy Process (AHP) first introduced by R. Saaty (1987), is by far the most commonly applied MCDM method in the academic literature (Marttunen et al., 2017). It can be used to derive ratio scales from both discrete and continuously paired comparisons, allowing diverse elements that normally lack a basis for comparison to be contrasted in a systematic way. This distinguishes AHP from traditional decision-making methods. The comparisons can be taken from actual measurements as well as scales which reflect the relative strength of preferences and feelings from stakeholders (R. Saaty, 1987). This is done in 5 steps:

1. Decomposing the MCDM problem into more tangible sub-problems which can be analyzed independently.

2. Evaluating the elements of the hierarchy through pairwise comparison with respect to their impact on the higher-level element. If two elements contribute equally to the higher-level objective, they get a score of 1, if expert judgment slightly favors one element over another it gets a score 3, if expert judgment clearly favors one over another it gets a score 5, if one element is strongly favored for example because its dominance is demonstrated in practice it gets a 7, if one element is backed by evidence of the "highest possible order of affirmation" it gets a score of 9. Intermediate scores are allowed when compromise is needed. This results in a matrix which is then normalized and averaged, yielding a criteria weight vector w.

3. Checking for consistency. For example, if for 3 given criteria A, B and C, $A > B$ and $B > C$, then follows that C cannot be more important than A. For a large number of criteria this process is less straightforward and has to be done mathematically through the procedure described by Saaty.

4. Creating another matrix which scores the available decision alternatives with respect to the criteria established in step 1 using pairwise comparison. This matrix

is also normalized, and the rows are averaged resulting in matrix S.

5. Creating a matrix V of the overall scores using $V = Sw$. The numbers in this matrix represent the alternatives' relative ability to achieve the decision goal.

AHP has many advantages, as well as some disadvantages. The advantages of AHP include its ease of use, scalability, and moderate data intensity (Velasquez and Hester, 2013). Due to its hierarchical structure, it can easily be adjusted in size to accommodate larger decision-making problems. However, the number of expert judgments needed for a particular decision matrix of order n is $n(n-1)/2$, making it relatively time-consuming for large numbers of criteria and alternatives. Other disadvantages of AHP include the assumed independence between criteria and that it does not allow decision-makers to score the strengths and weaknesses of elements in isolation, but only in comparison to other elements.

AHP is also susceptible to rank reversal. However, this is more of a problem inherent to MCDM then one specific to AHP. Rank reversal can occur within a method but also when applying the same data to different methods. Due to the use of comparisons as a ranking method, it is possible that the final rankings can flip or reverse when a new alternative is added to the same process or alternatively, when a different MCDM technique is used on the same source data. This highlights once more that it is technically impossible to decide what the 'best' alternative is. Because it is for similar reasons impossible to find the best method from a set of alternative methods (see the MCDM paradox by Triantaphyllou and Mann (1989)). It is however considerd to be less of a problem if one only has a limited number of alternatives to begin with (Velasquez and Hester, 2013).

To address some of the shortcomings of AHP, T. Saaty (1996) proposed another framework, the analytical network process (ANP). Compared to AHP, ANP is better equipped to handle interdependencies and feedback loops between decision elements. However comes at the cost of added complexity. In addition to the steps of AHP, interdependencies among criteria of a cluster must also be examined pairwise in ANP, with the influence of the elements on one another represented by an Eigenfactor. This results in a much larger supermatrix which is more mathematically complex and time consuming to create.

Figure 2.1 (p.12) illustrates the added complexity of ANP compared to AHP. Despite the convoluted operations involved in ANP, it has often been utilized in project selection, product planning, supply chain management, and scheduling problems. Many of these problems have interdependencies that AHP does normally not handle well (Velasquez and Hester, 2013). Still, reflecting on Figure 2.1 and the mathematical operations involved in ANP poses the question whether it is realistic to assume that an ANP process can be recreated in the context of the daily operations of the firm, to a similar but different decision-making problem by practitioners who do not necessarily have a lot of time to familiarize themselves with the process.

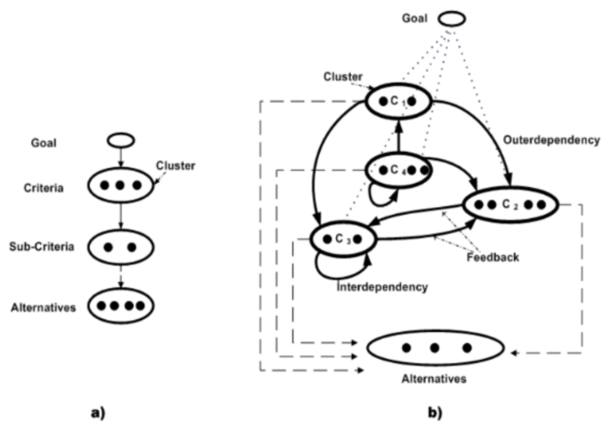

Figure 2.1: Structural difference between hierarchy (a) and network (b) processes (Görener, 2012, p.197).

2.1.2. ELECTRE AND PROMETHEE

ELimination Et Choix Traduisant la REalité (ELimination and Choice Expressing the REality) better known as ELECTRE is a family of outranking methods based on concordance analysis. Although multiple iterations on the ELECTRE method have been developed over the years, they can all be broadly divided into two main phases: (i) the construction of several outranking relations, aimed at comparing in a detailed way each pair of alternatives, and (ii) an exploitation phase which is used to elaborate on recommendations from the results obtained in the first phase. Taking into account uncertainty and vagueness in the evaluation of criteria and alternatives, ELECTRE methods use of two distinct sets of parameters for each criterion, importance coefficients and veto thresholds, in order to determine the relative performance of alternatives. The exact structure of the method depends on the iteration and the problematic of interest (choosing, ranking or sorting). A comprehensive overview of ELECTRE methods can be found in the work by Figueira et al. (2016).

Similar to ANP, ELECTRE can be relatively hard to explain in layman's terms. Furthermore by using the outranking method, which is a form of pairwise comparison, the strength and weaknesses of alternatives cannot be directly identified but only in relation to other alternatives (Velasquez and Hester, 2013).

Preference Ranking Organization METHod for Enrichment of Evaluations is similar to ELECTRE in that it also has several iterations and is also an outranking method. Among the PROMETHEE iterations, PROMETHEE II is the one most commonly referred to by researchers (Behzadian et al., 2010). The procedure of PROMETHEE II has five main steps: (i) determination of deviations based on pairwise comparisons of alternatives along the recognized criteria, (ii) application of a preference function to determine the degree of relative preference between criteria, (iii) calculation an overall global index for preference intensity by summing criteria weights, (iv) partial ranking by calculation of outranking flows, and (v) calculation of net outranking flows by subtracting leaving from entering flows.

Compared to ANP and ELECTRE, PROMETHEE it is relatively straightforward to apply, while it does not require the assumption that the criteria are proportionate like some simplistic methods. Nevertheless, the operations may still be not be easily understood by non-expert decision-makers. Furthermore, like AHP, ANP, and ELECTRE, its use of pairwise comparison does not allow alternatives to be scored directly based on a fixed benchmark, but only in relation to other available alternatives.

2.1.3. VIKOR AND TOPSIS

The basic principle of Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is that the chosen alternative should have the shortest distance from the ideal solution and the farthest distance from the negative-ideal solution in a multi-dimensional computing space. It consists of the following six steps: (i) calculate the normalized decision matrix, (ii) calculate the weighted normalized decision matrix, (iii) determine the ideal and negative-ideal solution, (iv) calculate the separation of each alternative from the ideal solution and the negative ideal solution using the n-dimensional Euclidean distance, (v) calculate the relative closeness to the ideal solution, and (vi) rank the preference order of alternatives based on relative closeness from the positive and negative ideals (Opricovic and Tzeng, 2004).

According to Velasquez and Hester (2013) TOPIS has numerous advantages. It is programmable, relatively easy to use, and the number of steps remain the same, regardless of the number of attributes under consideration. Due to its ease of use it has gained broad popularity in the scientific community and acquired a proven track record over the years, both as standalone method and as verification tool for other MCDM methods. A disadvantage is that its use of Euclidean distance does not consider the correlation of attributes. It is difficult to weigh attributes and keep consistency of judgment, especially with a large number of attributes. Although TOPSIS does not use pairwise comparison, the attributes of alternatives can still not be scored in isolation because evaluations rely on the ideal-best and ideal-worst alternatives, which are based on the characteristics of the actual available alternatives.

VIKOR is similar to TOPSIS in that it is also based on an aggregating function that represents closeness to an ideal. There are however two important differences between the two methods: (i) the VIKOR method of compromise ranking determines a compromise solution, providing a maximum group utility for the majority of criteria and a minimum regret for individual items, whereas TOPSIS determines a solution with the shortest distance to the ideal solution and the greatest distance from the negative-ideal solution without considering the relative importance of these distances, and (ii) VIKOR uses linear normalization which does not depend on the evaluation unit of a criterion function, whereas the normalized values by vector normalization in TOPSIS may depend on the evaluation unit (Opricovic and Tzeng, 2004).

2.2. GROUP DECISION PROCEDURES

Whereas on paper decisions are made by a single decision-maker, e.g., a high-ranked executive, in practice they are often the result of the views of multiple individuals brought together. Yet, from the original work on the various MCDM methods it would appear that they were written with a single decision-maker in mind (Vincke, 1992). But while decision-makers' value systems may not always coincide, it does not mean that MCMD is not useful to the group decision process. As long as the aggregation method, the alternatives, and the criteria are common to the group, they are still applicable. With a shared vision on these elements as a prerequisite, Belton and Pictet (1997) developed a framework for group decision procedures which use MCDM. The framework allows the different ways of working with MCDM methods to be contrasted from a sociotechnical perspective.

It is based on three elementary procedures - sharing, aggregating and comparing - which define the way in which the views of individuals are brought together with the aim of achieving a group decision. They are defined as follows (Figure 2.2):

- **Sharing** aims to obtain a common element by consensus, through a discussion of the views and the negotiation of an agreement; it addresses the differences and tries to reduce them by explicitly discussing their cause.
- **Aggregating** aims to obtain a common element by compromise, through a vote or calculation of a representative value; it acknowledges the differences and tries to reduce them without explicitly discussing their cause; aggregation of individual values may be obtained using the mean or through more sophisticated formulas.
- **Comparing** aims to obtain an individual element (to reach an eventual consensus based on negotiation of independent individual results); it acknowledges the differences without necessarily trying to reduce them.

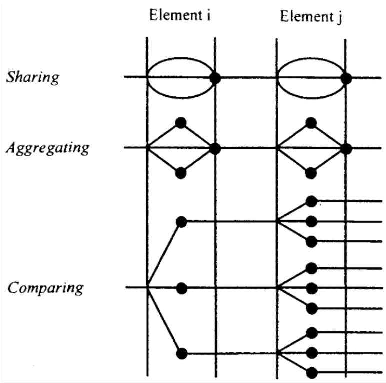

Figure 2.2: Elementary procedures in group MCDM (Belton and Pictet, 1997, p.288).

In contrast to aggregating and comparing, the act of sharing is likely to highlight differing interpretations of criteria at an early stage, thus ensuring a shared understanding of the decision-making process and an increased likelihood of a consensual outcome. Aggregating procedures may never detect such differences and when using a comparing procedure they may not emerge until a late stage. However, the cost of a sharing procedure in terms of time commitment and demands on the facilitator are significantly higher. Procedures that aggregate individual results devote less time to discussion and thus permit a more rigid regulation of the time schedule.

Assuming common alternatives, criteria, and aggregation method, the three elementary procedures can be applied to weights and evaluations. There are three main causes for different views on these elements: (i) uncertainty, (ii) conflict, and (iii) misunderstanding. Uncertainty and misunderstanding can be a consequence of inaccurate or incomplete information, while conflict results from different values or priorities regardless of the level of ambiguity of the problem context.

Since weights are considered as the most direct expression of the individual's value system, inter-criteria information is expected to be different for each decision-maker. Evaluations can be either:

- **objective:** the evaluations are based on observed performances, can be measured and are thus independent from the actors involved in the process,
- **subjective:** the evaluations are specified by each decision-maker, without compulsory reference to "reality" or to the other actors,
- **constructive:** although it is accepted that there is no objective measurement of "reality", an effort is made to adopt an explainable and justifiable framework.

In the case of subjective and constructive evaluations it is possible to use an interval of possible values rather than an aggregation or compromise value.

Futhermore, Belton and Pictet (1997) provide guidance on how to deal with situations in which not all members are experts in the full range of aspects related to the problem. They argue that in this case one should take either a sharing or an aggregating approach. Sharing allows the group to combine partial expertise, while aggregating could limit the evaluations to expressed judgments, i.e., individuals could abstain from contributing when expertise is lacking. Comparing however copes less well with this situation, as its rationale is not useful if there are no evaluations to be compared. Where the knowledge gaps lie and what the required follow-up needs to be would then only become available until after the evaluation, making it a sub-optimal strategy to adopt when faced with differing levels of expertise.

2.3. MCDM APPLICATION IN THE AIR TRANSPORT INDUSTRY

Air transport is an area in which MCDM approaches are becoming more and more prevailing. The aviation industry brings about many positive effects such as global connectivity and a wide range of quality jobs, but also some harmful externalities. Stakeholders in the field face a decision environment that involves a variety of alternatives that have to be evaluated based on different, often opposing criteria. This created the need for an MCDM approach. MCDM applications in aviation range from the evaluation of airport performance and airline service quality, the analysis of airport risk and air traffic controllers stress factors, through fleet management and assignment problems, selection of aircraft types and aviation fuels, to operational maintenance as well as optimization of aircraft routing and hub allocation (Rezaei and Kadziński, 2018).

Recognizing the lack of systematic reviews on the application of different MCDM techniques to a specific problem area in general, and the gap related to the application of MCDM to problems arising in the aviation industry in particular, Dožić (2019) analyzed 166 papers in which aviation-specific problems are solved using MCDM. The papers spread across 69 journals, with by far the largest share (34%) coming from the Journal of Air Transport Management, the only journal dedicated exclusively to air transport issues. AHP was the most popular among the selected papers followed by TOPIS and ANP. Dožić also showed that the authors who employ ANP for solving MCDM problems in the aviation industry often combine it with DEMATEL.

DEMATEL, which has not yet been discussed in this review, is similar to ANP to the extent that they both provide ways to deal with interdependence and feedback between criteria and alternatives. In contrast to ANP however, the DEMATEL method aims to determine the degrees of influence between criteria, rather than using the assumption of equal weight for each cluster, hereby aiming to more accurately reflect reality. A general hybrid DEMATEL-ANP model uses DEMATEL to clarify interrelations of components before going through the traditional steps of ANP. An example of a general hybrid MCDM model combining DEMATEL and ANP can be found in Ou Yang et al. (2008).

Dožić (2019) also indicates that fuzzy logic is very commonly used in MCDM problems within air transport research (50% of the reviewed papers). She suggests that the reason for such intensive use can be found in the fact that in the aviation industry, the collection of quantitative data is often very expensive or for other reasons unfeasible. Consequently, decision-makers will frequently have to rely on linguistic assessment. In order to compensate the lack of some precise data and the vagueness of human thinking that is related to such evaluations, membership functions can be developed that transform the linguistic variables into fuzzy numbers.

The concept of fuzzy logic can be explained using the example of Bruno et al. (2015) who propose an AHP model combined with fuzzy set theory for the selection of a regional aircraft type. One of the criteria included in their analysis is the cruise speed of the aircraft.

The linguistic expression "Good" comes from natural language input and can tell something about the cruise speed. There are however no defined objective boundaries for what would be considered a "good" cruise speed. It depends on which airline manager you talk to. The statement "this aircraft has a good cruise speed" is according to Boolean logic either true or false. When the speed of a regional jet is Mach .75, no one will claim this statement is false. It gets however problematic when a regional aircraft has a cruise speed of Mach .55. One airline decision-maker will call this "Good", while others will think it would be more accurately described as "Medium" or even "Poor". Rather than assigning a binary value, fuzzy logic allows for the indication of a degree of truth, for example 0.6 on a 0 to 1 scale. Since there is overlap between one decision-makers' "Medium" range and another's "Good" range, a speed of Mach 0.55 can for example have a 0.8 membership in the fuzzy set "Medium" and a 0.2 membership in the fuzzy set "Good".

Figure 2.3 (p.17) shows the membership functions of the linguistic variables "Very Poor", "Poor", "Medium", "Good", and "Very Good" that describe the criterion "Speed" as well as an example of the estimation process for the qualitative performance level "Good". The horizontal axis shows the speed as a Mach number and the vertical axis displays the degree of membership to the five linguistic descriptors as a value between 0 and 1.

Similarly, linguistic words could be used to represent preference between criteria. Fuzzy logic can then be employed to convert these linguistic terms into criteria weights.

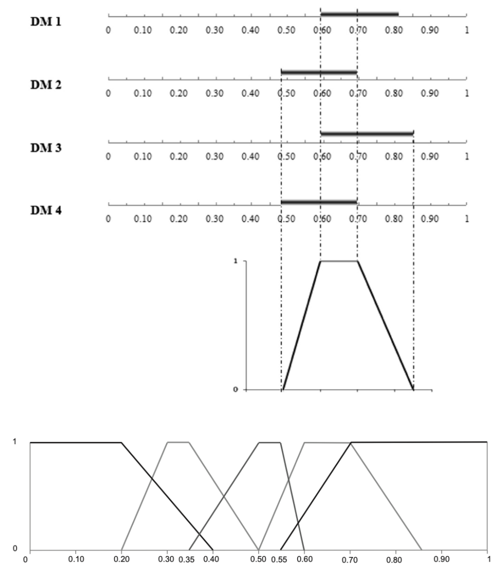

Figure 2.3: Fuzzy set of the criterion "speed" defined as linguistic variables on five qualitative levels (Bruno et al., 2015, p.5586).

An example can be found in Ma et al. (2020). As such, all MCDM methods could be used in either a crisp or fuzzy environment. Fuzzy set theory can be especially useful in situations where limited information and imprecise input is given. Determining the fuzzy sets however, can require numerous simulations or choice experiments if sets from the literature are not applicable. In the context of the case study fuzzy MCDM is therefore not suitable as a weight estimation method. Besides, airline managers are expected to have no problems expressing preference data between high-level strategic objectives on a crisp n-point scale. In future iterations of the method however, fuzzy logic may be used to quantify linguistic expressions for scores on individual sub-criteria.

2.4. REFLECTION ON THE LITERATURE

An online search returned no published scientific literature directly related to an MCDM approach for the evaluation of eco-oriented process innovations in the flight operations domain of an airline. From the special issue in the Journal of Air Transport Operations by Rezaei and Kadziński (2018) as well as the 166 papers reviewed by Dožić (2019) can be concluded that MCDM is widespread in air transport industry research. Furthermore, the approach has been applied to sustainability driven project selection problems before. This was done for example by Ma et al. (2020), who used a Fuzzy TOPSIS approach for project portfolio selection from the perspective of sustainability in an uncertain decision-making environment. However, there is a gap in the literature when it comes to the combination of the three, i.e., project evaluation, sustainability, and airline operations.

A review of the more commonly used MCDM techniques revealed that due to the decision-making paradox there is no consensus on the optimal MCDM method, nor on the process of selecting a suitable method for a given decision problem. While some methods are used more frequently than others, each method has its own strengths and weaknesses. The applicability of any MCDM method is highly dependent on the problem, the availability of data, the level of expertise of the decision-makers, and the amount of time allowed to complete the process. Since this thesis project deals with the application of an MCDM method in the context of the firm, the assessment of the methodologies should be based, to a large degree, on its operational usefulness. This requires that ease of use by non-experts, transparency of the decision-logic to the stakeholders, and freedom from ambiguity regarding the interpretation of inputs needed, have a high priority in the evaluation of the methods. For this reason, it might be wise to delay the choice for the full MCDM procedure until all relevant stakeholders have been consulted and criteria hierarchies have been defined. The assessment should include but not necessarily be limited to the popular MCDM methods listed in the review. It should also be highlighted that MCDM in general offers a large degree of flexibility. The disadvantages of methods might be overcome by combining them with other MCDM methods or other concepts such as fuzzy set theory.

2.4.1. PSYCHOLOGICAL BACKGROUND OF MCDM METHODOLOGY

In the meantime however, it is useful to reflect on the general concept of weight estimation of the various MCDM methods. Aside from TOPSIS and VIKOR which use direct rating based on geometric closeness to an ideal, all reviewed methodologies rely on pairwise comparisons. This concept is based on insights from psychometrics, in particular the work of Thurstone (1927). His 'Law of Comparative Judgment' found that human decision-making is more reliable when comparing two stimuli, rather than passing absolute judgment on a single stimulus in isolation.

Of course, all methodologies based on this concept are only applicable when various alternatives are available that are somewhat comparable. This may pose a risk to the research proposed in this document, as the candidate projects are yet to be developed, both internally and by the AI prototyping studio Transavia recently partnered with. Yet, when dealing with high-level strategic objectives, as is the case with the decision problem described in this report, the requirement of comparability is fulfilled relatively easily. The items that are being compared can differ in nature as long as they have a common objective and can be scored on the same high-level strategic dimensions. For example, Tsai et al. (2009) performed a hybrid MCDM-based ranking of companies in terms of their CSR efforts, ranging from a semiconductor manufacturer to a company involved in the operation of convenience stores.

2.4.2. THE PROBLEM WITH PAIRWISE COMPARISON METHODS

On paper, traditional matrix-based MCDM methods which use pairwise comparisons, e.g., AHP, provide an effective way to deal with relative preferences in situations where objective scoring is meaningless or unfeasible. Unfortunately however, they face significant quality control issues. Pairwise comparison requires the decision-maker to articulate the direction as well as the strength of the preference of one element over another.

Usually, indicating the direction is a relatively easy task. However, the strength of the preference is often harder to express, which leads to recurring inconsistencies (Herman and Koczkodaj, 1996; Forman and Selly, 2001).

Most methods provide procedures to check for matrix consistency. However, simply getting the decision-makers to revise their assessments until a desired level of consistency is reached, is not very efficient. Whereas an alternative approach using statistical quality control such as recommended by Karapetrovic and Rosenbloom (1999) could be used, it would be better to apply a more structured comparison method that aims to prevent inconsistencies up front.

2.4.3. BEST-WORST METHOD

The Best-Worst Method (BWM) is one of the more recently introduced MCDM methods and it addresses the low reliability that characterizes pairwise comparisons (Rezaei, 2015). To that end it utilizes the reference comparisons that are often already made implicitly in traditional pairwise comparison methods, but in a more structured way. For example if one wants to assign a number to express one's strength of preference of element A over B, one also keeps in mind the relations between A, B, and some other reference elements. Logically, these reference points would be the best and the worst alternatives. BWM not only makes this process explicit, but more interestingly; it shows how the relative preference of all criteria and alternatives can be derived from just the reference comparisons (Figure 2.4, p.20), without carrying out the secondary comparisons. This can be done in 5 steps:

1. Determine a set of decision criteria.
2. Determine the best (e.g. most desirable, most important) and the worst (e.g. least desirable, least important) criteria.
3. Determine the preference of the best criterion over all the other criteria using a number between 1 and 9, which results in a Best-to-Others vector $A_B = (a_{B1}, a_{B2}, ..., a_{Bn})$, where a_{Bj} indicates the preference of the best criterion B over criterion j.
4. Determine the preference of all the criteria over the worst criterion using a number between 1 and 9, which results in a Others-to-Worst vector $A_W = (a_{1W}, a_{2W}, ..., a_{nW})^T$, where a_{jW} indicates the preference of the criterion j over the worst criterion W.
5. Find the optimal criteria weights $(w_1^*, w_2^*, ..., w_n^*)$ by minimizing the maximum absolute differences $\left|\frac{w_B}{w_j} - a_{Bj}\right|$ and $\left|\frac{w_j}{w_W} - a_{jW}\right|$ for all j, considering the non-negativity and sum condition for w_j.

Relative to established pairwise comparison methods, the structured approach of BWM results in higher reliability, while limiting the input needed from decision-makers. Rezaei (2015) provides several reliability metrics measuring the consistency of output, violations of ordinal preferences, Euclidean distance between ratios of weights and their comparisons, and conformity to intuitive scores. In a real-world application considering a number of evaluation criteria related to mobile phone selection, BWM outperformed AHP on all metrics.

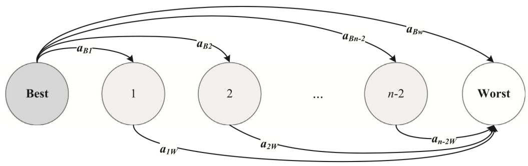

Figure 2.4: Reference comparisons used in BWM (Rezaei, 2015, p.51).

Furthermore, BWM only requires $2n-3$ comparisons, which is a significant improvement over the amount of decision-maker information required in matrix-based MCDM methods, for instance the $n(n-1)/2$ comparisons required in AHP. Emprical data from Rezaei shows that filling in a BWM questionnaire takes significantly less time than filling in a AHP questionnaire. Given the characteristics of BWM versus matrix-based methods, one can reasonably assume that the difference in time demanded from evaluators is even more pronounced when problem dimensions increase.

Finally, whereas original BWM as introduced by Rezaei (2015) finds the optimal weights of a set of criteria based on the preferences of a single decision-maker, the process can be extended to amalgamate the preferences of multiple decision-makers in a group decision-making problem. Typically, common weights for multiple decision-makers are obtained using the average operator (e.g. arithmetic or geometric mean). However, averages are sensitive to outliers and by reducing the different inputs into a single representative weight, important information regarding overall preferences is potentially lost. Recognizing this issue, Mohammadi and Rezaei (2020) propose a probabilistic group decision-making model based on BWM. Accordingly, they use a Bayesian hierarchical model, tailored to compute the weights for a group of decision-makers. The paper further introduces a new ranking scheme for decision criteria. It allows a confidence level to be assigned to measure the extent to which a group of decision-makers prefers one criterion over another, and provides a way to visualize the interrelation of criteria and confidences. Thus, Bayesian BWM yields much more information than the original method, without requiring additional time-consuming inputs from the decision-makers. By checking the weights in a probabilistic sense, the group will be better informed about where they can be certain about the relation between two criteria, and where it should be interpreted more carefully.

However, the mathematical model behind Bayesian BWM is relatively complex. Hence, striving for complete transparency in the process of translating inputs to recommendations, as advocated for by Stewart (1992), will no longer be meaningful to non-expert decision-makers. This requires another trade off: between transparency and depth of analysis.

Furthermore, the fact that in Bayesian BWM criteria weights of multiple decision-makers are instantly combined might not be beneficial for this particular application. Whereas limiting the effects of outliers is normally a good thing, the input from influential managers that deviate from the consensus might in this case provide valuable information. Thus, skipping the step of computing and reflecting on the criteria weights of each individual decision-maker is not recommended. To preserve the benefits of both iterations it might be useful to apply them in parallel and compare the results.

2.4.4. Conclusion to the theoretical background

It can be concluded that MCDM can assist decision-makers in selecting the most desirable alternative from a set of feasible alternatives based on their performance with respect to multiple and often conflicting decision criteria. The scores are derived either from objective data sources or qualitative approaches. As scores relating to different criteria are usually expressed in different units, the data need to be normalized using a normalization formula.

Based on the research objective presented in this report, the high-level criteria classes will be related to the corporate strategic objectives in the field of sustainability as well as the five other critical dimensions mentioned in the introduction. An MCDM method can be applied to establish preference relations between the KPIs associated with these dimensions, determine their weights, and assign a performance score to the proposed innovation projects.

This can be done either by an individual decision-maker or by a group of decision-makers; using one of the procedures outlined in section 2.2. The group approach promotes resolution of uncertainty, conflict, and misunderstanding – or at least makes them explicit. It may promote the adoption of a higher-level view in which people recognize that their own perspective is just one of many. Additionally, it allows for more advanced probabilistic analysis. On the other hand, discussions with airline managers highlighted that, in situations that concern sensitive issues such as flight safety, it is not uncommon that the outcome of such group procedures is vetoed by the accountable manager or nominated persons. In that case it may be more practically relevant to reduce the group of 'decision-makers' to the ones that indeed formally call the shots in the corporate setting. Even if that means the input is limited to that of only one or two decision-makers. This would also prevent the MCDM procedure from being just a side-project without commitment to following through on the results, rather than a decision-support tool that is actually embedded in the governance structure. Deciding on who will be the evaluator(s) of the problem requires a further examination of the corporate checks and balances surrounding sustainability.

Regardless of the composition of the group of decision-makers, it is important to adopt a method that is operationally relevant to the problem at hand. While the decision-making paradox makes it inherently impossible to select the "perfect" approach, the five requirements that a "good" strategy should comply with are quite straightforward. The dynamic environment of the decision-problem calls for a method that:

1. is flexible,
2. can deal with uncertainty in situations where it is unfeasible or meaningless to provide precise quantitative estimates, and
3. has been demonstrated to deliver high reliability when used by non-experts, while
4. placing minimal demands on their time, and
5. can be explained to management with relative ease.

Table 2.1 (p.24) summarizes the discussed MCDM methods and highlights the practical relevance of the method in the context of the decision problem. From the table follows that the popular MCDM methods are not in line with the strategy outlined above.

PROMETHEE, TOPSIS, VIKOR, and ELECTRE do not provide a clearly structured weight estimation method and will therefore be excluded. In theory these methods could be used combined with other methods that do allow the modeling of preference relations between criteria. However, since the positioning of sustainability relative to existing corporate objectives is the main topic of interest, these methods cannot be the starting point of the analysis. Besides, combining them with other methods such as AHP or BWM would further increase the difficulty of explaining the concept and operations to management.

For methods such as ANP, DEMATEL, and ELECTRE, the transparent communication of the process and results to management will be extremely challenging. Applications of these methods are done almost exclusively in academics and are too complex for implementations as a standard tool for practical decision-making for airline management. They assume familiarity with several technical parameters and require specific software to calculate results. Although the process could be assisted using free and open-source tools, this is not realistic as users have to become familiar with the interface and the mathematical operations involved. Furthermore, methods like ANP require extensive brainstorming or DELPHI sessions in order to find criteria and other relevant factors for the systematic investigation of interconnections between nodes. Since airline decision-makers simply do not have time for this, it is necessary to give up some depth of analysis for practicality.

In that regard AHP has some advantages over the more mathematically complex methods. The structuring of the decision-problem as a hierarchy should be intuitive to management as in business one is trained to break down complex goals into KPIs. The hierarchical model can be structured in groups of sub-criteria and if necessary additional levels can be introduced. This way, different levels of management can provide input at different levels of the hierarchy, while keeping a clear overview for all decision-makers involved. In general AHP input from a group of people can be relatively easily consolidated by using the geometric or arithmetic mean. Furthermore, it is possible to perform the calculations in a simple spreadsheet. To conclude, AHP can be seen as a neutral and easy to use decision-making tool that does not require a facilitator to explain to management what technical terms such as "Eigenfactor", "supermatrix", "outranking flows" or "Eucladian distance" mean. Thus it can be expected that results are more generally accepted by practitioners. Like BWM, AHP does in its standalone application not include a standard way for the normalization and aggregation of scores. However, as mentioned, the main focus in this report is weight estimation. It is therefore more relevant to choose a weight estimation method such as AHP and combine it with existing normalization and aggregation procedures at a later stage.

In that respect AHP could be a valid starting point for the analysis. However, the use of the traditional pairwise comparison of criteria is one of the major issues with AHP. To people who are not familiar with the method it can seem like an unintuitive way of comparing a set of criteria. Hence the inputs can result in a consistency index that is too high. Explaining this to the people involved in the decision and then asking them to reconsider their inputs can then be a real challenge. This is one of the reasons in psychology respondents are often asked to rank items or express agreement or disagreement with statements instead.

It can be concluded from Table 2.1 (p.24) that BWM is the MCDM method that best fits the strategy outlined above. First, it is adaptable. Not only can BWM be used to derive the weights independently, it can also be easily expanded to a full MCDM procedure by combining it with other techniques, including normalization formulas and aggregation methods. Second, it allows decision-makers to deal with uncertainty by using the law of comparative judgment. But without the practical inconsistencies that other pairwise-comparison-based methods suffer from. The final weights derived from BWM are highly reliable as it provides demonstrably more consistent comparisons than AHP. Finally, the vector based approach of BWM requires fewer comparisons than matrix-based alternatives and has been empirically proven to save time in comparison to AHP.

These features make BWM the most suitable starting point for approaching the problem at hand. However, in order to further develop the MCDM approach to sustainability-focused innovation project evaluation in airlines, more research is needed on the related sub-criteria, candidate project characteristics, data availability, and requirements covering the specific needs for the decision-support system.

Table 2.1: Summary of the key features and relevance of popular MCDM methods

Method	Technique	Advantages	Disadvantages	Weight est.	Norm. and agg.	Relevance
AHP	Matrix-based pairwise comparisons	Wide availability of related literature due to its popularity; easy to use; hierarchical model is intuitive to most managers; scalable; moderate data intensity; easy to combine multiple inputs from different decision-makers into a consolidated outcome.	Problems due to assumed independence between criteria and alternatives; inconsistencies in pairwise comparisons; relatively time-consuming; $n(n-1)/2$ comparisons.	Included; less reliable and more time-intensive compared to BWM.	Not included; relatively easy to expand to full MCDM procedure using normalization formula and AVF.	Limited; placing too high demands on time from busy managers; challenging to request corrections if consistency index is too high.
ANP & DEMA-TEL	Matrix-based pairwise comparisons	General network model can handle practically any decision problem; able to account for interrelations and feedback loops between decision elements; suitable for gaining a deep understanding of a specific problem.	Added complexity and time demands relative to AHP; managers cannot be expected to understand the concept and process; verification unfeasible due to interrelations.	Included; time-intensive process.	Included.	None; not suitable for the decision support system; cannot be applied as a standard tool for practical organizational decision-making; too complex to be replicated by non-experts outside the academic setting without a facilitator; assumes familiarity with specific software.

Table 2.1 continued from previous page

Method	Technique	Advantages	Disadvantages	Weight est.	Norm. and agg.	Relevance
ELEC-TRE	Matrix-based pairwise comparisons	Takes into account veto and indifference thresholds; allows for the elimination of compensation effects.	Algorithm can be difficult to explain in layman's terms; assumes familiarity with a number of technical parameters.	Not included; could be combined with AHP or BWM but this further increases the difficulty of explaining the concept and operations to management.	Included.	None; not suitable for the decision support system; cannot be applied as a standard tool for practical organizational decision-making; too complex to be replicated by non-experts outside the academic setting without a facilitator; assumes familiarity with specific software.
PROME-THEE	Matrix-based pairwise comparisons	Relatively straightforward to apply compared to ANP and ELECTRE.	Does not provide a clearly structured method for weight estimation; despite its relative ease of use compared to ANP and ELECTRE, the process may still be difficult to replicate by non-expert decision-makers.	Not included; could be combined with AHP or BWM but this further increases the difficulty of explaining the concept and operations to management.	Included.	None; does not provide a weight estimation method.

Table 2.1 continued from previous page

Method	Technique	Advantages	Disadvantages	Weight est.	Norm. and agg.	Relevance
TOPSIS & VIKOR	Closeness to ideal solution using n-dimensional Euclidean distance (TOPSIS); compromise ranking by maximizing group utility while minimizing regret for individual items (VIKOR)	Easy to use for facilitator; can be programmed; the number of steps remains the same regardless of the number of attributes.	Explanation of concept to management may be challenging (especially VIKOR); does not consider the correlation of attributes; difficult to keep consistency of judgment; does not provide a weight estimation method.	Not included; could be combined with AHP or BWM but this further increases the difficulty of explaining the concept and operations to management.	Included.	None; does not provide a weight estimation method; can only handle numerical data.
BWM	Vector-based pairwise comparisons	Requires fewer comparisons relative to matrix-based methods; provides more consistent comparisons than AHP; can be combined with other MCDM methods; only integers are used, making it easier to apply than matrix-based methods.	Does, in its standalone application, not consider interrelations and feedback loops between decision elements; does not provide a structured way for normalization and aggregation.	Included; more reliable and less time-intensive compared to AHP.	Not included; relatively easy to expand to full MCDM procedure using normalization formula and AVF.	Highest practical relevance; exploits the benefits of pairwise comparisons without the drawbacks; provides high reliability while keeping the difficulty of explaining the concept and operations to management at an acceptable level.

3

Methodology

This chapter continues with a description of the steps that were taken to answer the sub-questions and the main research question [1]. Figure 3.1 (p.41) summarizes the applied methodology. It shows the input and output of the research activities, and how these activities are related to each other.

Due to the lack of references available in the scientific literature on the evaluation of sustainability-oriented process innovations from the perspective of flight operations, and the exploratory nature of the research, a case study approach was chosen. According to Yin (2014) this type of approach is useful for answering "how" research questions. The research was carried out within Transavia, the low-cost arm of Europe's third-largest airline group, operating from bases in the Netherlands and France. The study was initiated following preliminary discussions with senior managers of the airline involved in sustainability, innovation management, and finance. These discussions highlighted the need for a systematic approach to project selection which supports planners in the effective alignment of the firm's new sustainability agenda with existing (and often conflicting) strategic goals. To this end, the study investigated how a multi-criteria decision-making approach could assist the airlines' innovation project evaluation process. Taking the SPOEFS dimensions as criteria and the operational improvements as alternatives, this type of analysis can be viewed as an MCDM problem and therefore provides a novel procedure for the appraisal of process innovations in flight operations.

3.1. Research Framework

Chapter 1 briefly discusses the main dimensions of the existing SPOEFS framework. SPOEFS is the collective name for the 6 indicators the company uses to monitor its most important business objectives. With these factors as a starting point, the framework was formulated using the components of MCDM. This implies that the framework contains several decision criteria that are used to evaluate a certain number of alternatives. In this

[1] Parts of this chapter were adapted from De Decker (2021)

case there are four main elements of interest; i) the SPOEFS dimensions, ii) their perceived relative importance, iii) the project alternatives, and iv) the performance score of the alternatives. Due to time constraints on this thesis project and the fact that the alternatives were being developed in parallel to the research activities, the main focus was on the first and second element. While recommendations were proposed related to element three and four, a crisp, ready-to-use MCDM framework was not achieved. Nevertheless, to put the results of this study into context, a complete (but generic) framework for estimating project performance is outlined in the subsections below. The generic steps were based on Li et al. (2020) and adapted to the airline context. However, the five steps described by Li et al. (2020) nor the adaptation differ conceptually from most MCDM applications in the literature. A more detailed description of the specific techniques that were applied within this framework: the BWM and Bayesian BWM, can be found in section 3.2 and 3.3 respectively. This chapter concludes with a detailed description of the case study that was conducted (Chapter 3.4).

3.1.1. Step 1. Determining the Objective of the Project Performance Analysis and Defining the Scope of the Problem.

The goal of the MCDM model is to predict the performance of process optimizations in flight operations. The first step is to define the goal of the analysis. In general MCDM applications can have 4 possible goals (Li et al., 2020). In the context of this study they can be defined as follows: (i) evaluation, where the aim simply is to identify the performance of each operational improvement proposal; (ii) selection, where the goal is to select certain projects from a list of candidate projects; (iii) ranking, where the aim is to rank several projects in terms of their expected performance; and (iv) classification or sorting, where the aim is to differentiate between different categories of projects.

Since a large number of smaller impact projects are necessary to cause a noticeably change in the airline's carbon intensity, and they are generally low-capex initiatives, the decision support system will likely go through all these phases until it is simply a sorting problem (Goal iv). The initiatives will then either get the classification; "pursue" or "do not pursue". However, due to the many unknowns the scope is, for the time being, limited to the goal that involves the least complex; evaluation (goal i).

3.1.2. Step 2. Determining the Evaluation Criteria for Project evaluation through the SPOEFS approach and Expert Opinion.

Together with step 4, this is one of the two focal points of this study. In this step, different attributes of the operational improvements determine the performance in terms of the SPOEFS criteria. While most of the strategic dimension broadly describe the relevant performance criteria already, some had to be adapted to the research context. Therefore the opinions of HV subject matter experts were included to define the evaluation criteria for the specific purpose of evaluating projects. More information about this process can be found in section 3.4.1. In its generic form however, the process takes the shape of matrix S:

$$S = \begin{array}{c} \\ p_1 \\ p_2 \\ \vdots \\ p_m \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_n \end{array} \\ \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{pmatrix} \end{array}, \tag{3.1}$$

where $C = \{c_1, c_2, ..., c_n\} = \{safety,\ pxi,\ operations,\ exi,\ finance,\ sustainability\}$ denotes the criteria for project evaluation, $P = \{p_1, p_2, ..., p_m\}$ indicates a list of projects, and s_{ij} represents the corresponding performance score of each initiative i with respect to criterion j.

3.1.3. STEP 3. COLLECTING PERFORMANCE SCORES OF EACH PROJECT FOR ALL CRITERIA FROM VARIOUS DATA SOURCES.

For each project, performance scores S_{ij} are collected from various data sources. These are usually a combination of objective evaluations, i.e. based on observed (historical) data that can be measured independent from the actors involved, or subjective evaluations from interviews and questionnaires. Due to the variety of data sources, the performance scores will have different units. In order to be compared to each other and aggregated into a single score, they have to be converted into the dimensionless s_{kj}^{norm} using one of the normalization techniques. This can be done for instance using the linear-max normalization (Equation 3.2). Other options include linear max-min, linear-sum, vector, or logarithmic normalization.

$$s_{kj}^{norm} = \begin{cases} \frac{s_{kj}}{max\{s_{ij}\}}, for\ a\ benefit\ criterion, \\ \\ 1 - \frac{s_{kj}}{max\{s_{ij}\}}, for\ a\ cost\ criterion. \end{cases} \tag{3.2}$$

Which technique provides the most robust results depends on the objective of the normalization and the type of data that needs to be normalized. The results from a simulation by Chakraborty and Yeh (2007) suggest for instance, that with ranking consistency and overall preference value consistency as the performance metrics, vector normalization and the linear-max method are the most suitable for Simple Additive Weighting methods (including the AVF by Keeney and Raiffa (1976)), in decision settings where the attributes measurement units are diverse in range. Similarly, Vafaei et al. (2016) assessed the suitability of the 5 most popular normalization techniques for pairwise comparison data from AHP. They concluded that the best technique in this context is linear-max combined with linear-sum normalization, while the application of linear-sum alone is the least optimal method.

These studies illustrate that the suitability of the normalization formulas is context dependent. Accordingly, future research might be needed to determine which technique is the most reliable in the case study situation. However, since the alternatives that have to be scored are not yet available, the normalization step will not deserve much attention in the remainder of this exploratory study. Yet, the step is included here for the sake of completeness.

3

3.1.4. STEP 4. FINDING THE OPTIMAL WEIGHTS OF THE IDENTIFIED CRITERIA.

Finding the perceived relative importance $w^* = \{w_1, w_2, ..., w_j\}$ of all criteria C defined in step 2 is an essential part of this study. It was argued in the introduction that knowing which attributes are viewed as the most important can be beneficial to the integration of sustainability into the innovation management decision-making process. To incorporate the opinion of airline leadership a multi-criteria weighting method should be applied.

Chapter 2 showed that BWM is the most appropriate method to solve this problem. Since weights are considered to be a close representation of the individual's value system, inter-criteria information can be expected to differ between decision-makers. In order to reflect on the individual weights as well as the aggregated final weights of the group, both the original and novel Bayesian method were applied.

3.1.5. STEP 5. FINDING AN OVERALL LEVEL OF ESTIMATED PROJECT PERFORMANCE BY AGGREGATING THE SCORES.

The final part of the generic framework is the calculation of the overall performance for each of the proposed innovations. After calculating the criteria weights and normalized performance scores, the global score v_i for each alternative i can be calculated. This is usually achieved by entering the values w_j and s_{ij}^{norm} in the additive value function (Equation 3.3). Then for each project the performance level of alternative i is superior to alternative i', if $v_i > v_{i'}$.

$$v_i = \sum_{j=1}^{n} w_j s_{ij}^{norm}, \forall = 1, 2, ..., m. \tag{3.3}$$

The additive value function, also known as the Weighted Sum Model (WSM) is the most straightforward and most frequently applied aggregation method. One should however be careful with the liberal application of this method. While it is usually a good first-cut approximation even for complex problems, its results might be misleading if certain implicit assumptions regarding the decision-makers' preferences do not hold. These assumptions include:

1. *Completeness:* A preference is complete if the decision-maker has a preference for any pair of alternatives, i.e. the DM must be able to determine preferences between any set of project attributes and cannot be indifferent or unsure about the comparison (Vainio, 2019).
2. *Transitivity:* A preference is transitive if for any three projects a, b, and c, for which $(a \succ b)$ and $(b \succ c)$ with respect to a certain criterion, also holds $(a \succ c)$ (Vainio, 2019).

3. *Mutual preference independence:* The preference with respect to the performance on criterion c_n must not depend on the scores for other criteria C, e.g. an on-time performance (OTP) of 91% is always preferred to a 83% OTP, regardless of the scores on other decision criteria. Criteria are mutually preferentially independent, if all possible subsets of C are preferentially independent from the rest of the criteria (Eisenführ et al., 2010).
4. *Difference independence:* The preference over a change in local scores on criterion c_n does not depend on the scores related to the other criteria (Eisenführ et al., 2010). For instance a change in current operating income (COI) from 5M to 12M is equally preferred, regardless of the scores on other criteria as long as the scores with respect to these other criteria stay the same.

3

While assumptions 1, 2, and 4 are usually justified, the application of the additive value function may not be appropriate if the scores on the attributes are not mutually preferentially independent (assumption 3). In other words, in order to apply the AVF, one needs to be able to assume that no meaningful interaction exists between the attributes. If some attributes are not preference independent from each other, the model may not accurately reflect the DM's preferences.

Goodwin and Wright (2004) point out that *mutual* preference independence does not automatically follow if independence is established in one direction. Sometimes these interrelations can be obvious, as is the case in the practical example they give in their book; when choosing a holiday destination, customers may prefer a warmer climate to a cooler one, irrespective of whether or not the hotel has an open-air or indoor swimming pool, however their preference between hotels with open-air or indoor swimming pools will probably depend on whether the local climate is warm or cool. In more complex situations such as the decision-making problem studied in this report, the presence of dependencies between attributes can be harder to identify.

Consequently, striving for full consistency in representing the DMs preferences can require additional questionnaires or choice experiments. However, the data intensity involved in these types of analyses is not compatible with the research context. According to Goodwin and Wright (2004) one should at least take note of the use of phrases like 'this depends on ...' when the DM is asked to assign a value to a pairwise comparison. They further describe two possible courses of actions in case these comments show that mutual preference independence does not exist:

1. Return to the value tree and redefine the attributes so that a set of attributes which are mutually preference independent can be identified.
2. Use a multiplicative model which can handle the interaction between the attributes, instead of the additive model.

Both present practical disadvantages in light of the current problem. Redefining the value tree means redefining the structure of the airline's balanced scorecard. Since the SPOEFS model was just recently introduced, this approach cannot count on broad support from airline leadership. The second proposal is characterized by a high-data intensity and is more suitable for esoteric academic discussions within operations research than in a practical airline planning framework. Because the multiplicative model is not widely used in the literature nor by practitioners (Goodwin and Wright, 2004), it will not be considered in detail in this report. Accordingly, it was hypothesized that offering clear qualitative guidance based on the defined criteria and estimated weights, is more meaningful than attempting to aggregate the data in an additive value function at the risk of misleading decision-makers. For that reason the case study will specifically focus on step 2 and 4 of the generic framework.

3.2. LINEAR BEST-WORST METHOD

Chapter 2 provided a comprehensive discussion on why BWM is the most appropriate method to implement step 4 of the framework. In addition to the advantages listed there, the structured pairwise comparison approach of BWM can be seen as a debiasing strategy against the psychological phenomenon known as the *anchoring effect*. This cognitive biases discovered by Tversky and Kahneman (1974) occurs when people's judgments are skewed because they consider a particular reference value for an unknown quantity before estimating that quantity. By using two opposing reference points BWM aims to mitigate this effect (Rezaei, 2020).

Section 2.4.3 described the general steps of BWM. Although there are several versions of BWM, they all share the same input data, which are obtained by performing steps 1 to 4 (p.19). The output of these steps consists of the vector $A_B = (a_{B1}, a_{B2}, ..., a_{Bn})$, where a_{Bj} denotes the preference of the best criterion c_B over other criteria $c_j \in C$, and $A_W = (a_{1W}, a_{2W}, ..., a_{nW})^T$, where a_{jW} indicates the preference of the criterion $c_j \in C$ over the worst criterion c_W. In all iterations A_B and A_W are constructed using the same 9-point scale (Table 3.1, p.33). As such, it should be self-evident that $a_{BB} = a_{WW} = 1$.

The framework presented in this paper implements both the original method, as well as Bayesian BWM. Since step 5 as presented in section 2.4.3 yields an interval of multiple optimal solutions, the linear approximation was used for the original (non-probabilistic) BWM. Accordingly, step 5 was converted into model 3.4.

$$
\begin{aligned}
& min\, \xi^L \\
& s.t. \\
& |w_B - a_{Bj} w_j| \leq \xi^L, \; for\ all\ j \\
& |w_j - a_{jW} w_W| \leq \xi^L, \; for\ all\ j \\
& \textstyle\sum_j w_j = 1 \\
& w_j \geq 0, \; for\ all\ j
\end{aligned}
\tag{3.4}
$$

By solving 3.4, the unique optimal weights $w^* = \{w_1, w_2, ..., w_j\}$ and the optimal objective function value ξ^{L*} were obtained for each individual DM. For this model ξ^{L*} can be considered as an indicator of the consistency of comparisons. Values of ξ^{L*} close to zero show a high level of consistency of the pairwise comparisons provided by the decision-makers (Rezaei, 2016b).

Table 3.1: Definition of 9-point scale used in BWM

	Meaning of the numbers 1 to 9
1	Equal importance
2	Somewhat between Equal and Moderate
3	Moderately more important than
4	Somewhat between Moderate and Strong
5	Strongly more important than
6	Somewhat between Strong and Very strong
7	Very strongly important than
8	Somewhat between Very strong and Absolute
9	Absolutely more important than

3.3. Bayesian Best-Worst Method

The second version of BWM that was applied in this study is Bayesian BWM. Since the case study involved a group of airline managers, the probabilistic approach yielded useful additional information without increasing the time commitment from the DMs. In fact, the first 4 steps are identical to the original method (as well as the linear approximation described in Section 3.2). The difference however is in step 5. While original BWM yields concrete values for $w^* = \{w_1, w_2, ..., w_j\}$, Bayesian BWM is based on a probability distribution of each individual optimal weight $w^{1:K}$ and the overall optimal weight w^{agg}, given the vectors $A_B^{1:K}$ and $A_W^{1:K}$, where k represents the individual DM and $k = 1, ..., K$ (Li et al., 2020). The probability distribution was estimated using the procedure proposed by Mohammadi and Rezaei (2020).
Subsequently, the following joint probability distribution was used:

$$P(w^{agg}, w^{1:K} | A_B^{1:K}, A_W^{1:K}). \tag{3.5}$$

Based on Equation 3.5, the probability of each variable was computed with the sum rule:

$$P(x) = \sum_y P(x, y), \tag{3.6}$$

where x and y represent two arbitrary random variables.

To build the Bayesian model, a probabilistic hierarchical model was applied, where the optimal weight for each individual decision-maker w^k depends on the pairwise-comparison output A_B^k and A_W^k. In turn, the aggregated weight for the group w^{agg} depends on w^k, while either A_B^k or A_W^k is independent of w^{agg}:

$$P(A_W^K || w^{agg}, w^K) = P(A_W^K || w^K). \tag{3.7}$$

Combining Equation 3.5 with Bayes' theorem provides the following equations:

$$P(w^{agg}, w^{1:K} || A_B^{1:K}, A_W^{1:K}) \propto P(A_B^{1:K}, A_W^{1:K} || w^{agg}, w^{1:K}) P(w^{agg}, w^{1:K}), \tag{3.8}$$

$$= P(w^{agg}) \prod_{1}^{K} P(A_W^K || w^K) P(A_B^K || w^K) P(w^K || w^{agg}). \tag{3.9}$$

To further compute the posterior distribution, the variables in Equation 3.9 had to be specified. As the pairwise comparison vectors A_B^k and A_W^k consist of integers, the distribution could be modeled as a multinomial distribution:

$$A_W^k | w^k \sim multinomial(w^k), \forall k = 1, ..., K. \tag{3.10}$$

The same applies to vector A_B^k, with the only difference being that the weight is inverted:

$$A_B^k | w^k \sim multinomial(\frac{1}{w^k}), \forall k = 1, ..., K. \tag{3.11}$$

The next step was to determine w in the posterior multinomial distribution. The Dirichlet distibution acted as the prior distribution for modeling w because the non-negativity and sum constraints inherent to BWM:

$$Dir(w||\alpha) \sim \frac{1}{B(\alpha)} \prod_{j=1}^{n} w_j^{\alpha_j - 1}, \alpha \in \mathbb{R}^n \tag{3.12}$$

Equation 3.12 depicts the probability density function of the continuous random variable w when it obeys the Dirichlet distribution (Mohammadi and Rezaei, 2020). $B(\alpha)$ is a multivariate beta function, and $B(\alpha) = \prod_{j=1}^{n} \Gamma(\alpha_j)/\Gamma(\alpha_0)$. $\Gamma(\alpha)$ is the gamma distribution; α_j is the dimensionless distribution parameter, and $\alpha_j > 0 = 1, 2, .., n$. α_0 Is the sum of the distribution parameters, i.e. $\alpha_0 = \sum_{j=1}^{n} \alpha_j$ (Forbes et al., 2011).

Next, for every individual weight w^k in Equation 3.10 or 3.11, when w^{agg} is given, it is expected to be in the proximity of w^{agg}. Therefore the mean and concentration parameter of the Dirichlet distribution had to be updated accordingly:

$$w^k || w^{agg} \sim Dir(\gamma \times w^{agg}), \forall k = 1, ..., K, \tag{3.13}$$

where w^{agg} indicates the mean of the distribution and the non-negative parameter γ denotes the concentration parameter, i.e., the closeness between w^k and w^{agg} (Mohammadi and Rezaei, 2020). Also, γ needed to be modeled and the gamma distribution was adopted:

$$\gamma \sim \Gamma(a, b), \tag{3.14}$$

where a and b are the shape parameters of the distribution. Both values a and b where set to 0.1, as with such settings, the gamma distribution is similar to the uniform distribution, thus it has minimum effect on the posterior distribution (Mohammadi and Rezaei, 2020). Since Bayesian BWM uses estimation to find the weights of the probability distribution, the valid values for certain parameters are unknown. Therefore uniform-like distributions were used and the data were skewed based on the input variables. This means that when information regarding the distribution is incomplete, the probability distribution that is in line with this information, but has the maximum entropy value should be selected (Jaynes, 1957; Li et al., 2020).

In the last step, the uninformative Dirichlet distribution provided a prior distribution of w^{agg}. Conform Mohammadi and Rezaei (2020) α was set to 1. The final weights could then be described as:

$$w^{agg} \sim Dir(1). \tag{3.15}$$

Since the Bayesian model does not yield a closed-form solution, Markov Chain Monte Carlo (MCMC) was required to determine the posterior distribution 3.9. This was done using the MATLAB implementation of Bayesian BWM provided by Mohammadi and Rezaei (2020). In the MCMC simulation the JAGS program by Plummer (2004) was employed for the generation of the random sample.

3.4. CASE STUDY DETAILS

The MCDM approach outlined in section 3.1 provides a new paradigm for estimating the performance of an operational improvement proposal by incorporating the total impact of the project's attributes on the firm's operations. As such, the derived performance estimation is a more complete assessment of how well the process innovation will perform in terms of the multiple conflicting goals, relative to the ad-hoc planning and DCF analysis. More important still is the identification of the relative importance of the criteria, as perceived by airline leadership. Through the application of BWM and Bayesian BWM, the most important requirements for the evaluation of process innovations in flight operations could be determined. While the general steps can be applied to a wide range of operational improvements, this study particularly focused on the implementation of step 2 and 4 in the context of sustainable process innovation projects. In order to further develop these steps a case study was conducted at Transavia Airlines. The aim of this case study was to provide guidance on the evaluation of sustainability-oriented process innovations for innovation managers within the airline industry. To this end interviews were conducted and the original and Bayesian BWM were adopted to respectively define the criteria and estimate the criteria weights. The research activities of the case study are illustrated in Figure 3.1 (p.41).

3.4.1. CRITERIA DEFINITION

Project performance estimation within Transavia comprises the assessment of project attributes against the predetermined SPOEFS criteria. Accordingly the main attributes involved in the analysis are Safety, Passenger Experience, On-Time Performance, Employee Experience, Finance, and Sustainability. These criteria are based on the firm's balanced scorecard. What these universal terms constitute in the context of sustainable process innovations in flight operations is not yet addressed in the academic literature nor in internal company documentation. Therefore HV experts were consulted to include their opinions on how these dimensions should be defined at the unit of analysis of the innovation project.

Nine experts were interviewed to find their view on how the criteria should be defined. Seven internal SMEs (2 for sustainability and one for each of the other SPOEFS dimensions) were identified by their company peers as the foremost experts in their respective domains. Their positions within the airline were respectively, Safety Engineer, Data Analyst Consumer Insights, Director of Operational Development, Director of Organizational Development, Chief Financial Officer, Sustainability Developer, and Flight Operations Engineer. In addition, two outside experts in the field of innovation portfolio management were consulted; the Investment Director of the firm's corporate venture capital subsidiary, and the Managing Director of the firm's Artificial Intelligence partner. Although they were not directly involved in flight operations, they were able to provide general advise on the project selection process.

In theory the six criteria can be further expanded into an additional level of multidimensional attributes. However this will increase the difficulty of the evaluation. Since the goal was to address the strategic implications of the criteria and their perceived weights for higher level management, a comprehensive list of sub-dimensions was not included. Instead experts were asked to identify the *most important* consideration for strategic decision-making in their respective domain. The aim of these nine preliminary interviews was to set the boundaries of the conceptual framework as well as collecting advise on the decision-aid requirements, e.g. describing the necessary level of operational relevance and ease of use.

The semi-structured interviews were conducted using video-conferencing software and guided by a standardized topic list. The duration of the interviews varied between 30 to 60 minutes. The interviews were not recorded and notes were taken instead. This allowed participants to speak more freely, and because triangulation between multiple experts on the same topic was not relevant in this case, it did not come at the cost of rigor. Since the only objective of the interviews was the definition of the SPOEFS criteria at the level of the innovation project, the analysis of the interviews was quite straightforward as well. The summarized notes and all relevant comments from the experts are presented in Chapter 4. Where appropriate, the criteria definitions from the experts were elaborated using internal documentation or relevant literature from the academic and practitioner communities.

3.4.2. Pairwise Comparison Data collection

As mentioned before, to provide guidance regarding the project evaluation framework, two sets of data are required: the qualitative description of the criteria which were determined based on the procedure outlined in 3.4.1, and the pairwise comparison vectors A_B and A_W. For the determination of criteria weights input was required from a group of relevant decision-makers. In real-life situations, the exact composition of the group named 'decision-makers' will depend on the airline's new corporate governance structure for sustainability issues. However, since this structure is not yet implemented, 19 of the most high-level DMs were selected based on their influence on key strategic decisions related to flight operations. Of the 19 selected managers, 17 agreed to participate by individually providing comparison data and answering additional questions about the characteristics of the decision problem.

With the exception of the Chief Financial Officer – who was selected both as subject matter expert for the definition of the finance dimension as well as a key decision-maker for the purpose of weight estimation – there was no overlap between the 17 managers and the group of eight experts consulted in the criteria definition phase. This closely models the actual corporate decision-making process in which directors, VPs, and executives take decisions based on background information gathered from knowledgeable experts. Of the other managers involved, eight are from the Management Team Flight Operations: the Director of Flight Operations Support & Innovation, Project Manager of the Airline Pilot Program, Director of the Airline Training Organization, Chief Pilot, Director of the Flight Technical Department, two Directors of Cockpit Crew Planning, and the Flight Operations HR Business Partner. Six are from the Safety Review Board: [2] the Chief Operations Officer, Vice President of Flight Operations, Vice President of Cabin Operations, Vice President Technical Services, Vice President Operations Control, and the Vice President Ground Services. In addition, two managers were consulted from commercial domains that were deemed essential to the decision-making process: the Vice President Network and Resource Planning, and the Vice President Customer Proposition.

The pairwise comparison vectors were determined based on step 1 to 4 of BWM and input from the 17 key decision-makers. To collect the vectors A_B and A_W, the pairwise comparison inputs from each individual decision-maker were collected over a video call and entered in the BWM Solver. This solver provided by Rezaei (2016a) is spreadsheet-based and has the linear approximation for calculating the optimal weight (Model 3.4) pre-programmed. To ensure that all DMs had adequate information to conduct the comparisons, a brief summary was provided on the results from the criteria definition phase. In addition to the rating on the 9-point scale, the DMs were asked about their opinion with respect to the decision-making characteristics, i.e., how they perceived the degree of compensation, correlation, and preference (in)dependence between the main criteria. The loosely-structured approach of the interviews allowed for evaluation of other important issues that emerged, as well as reflection on the individually obtained weights, which the solver automatically visualizes in a bar chart.

Decision-makers were also requested to perform the pairwise comparisons with the Safety dimension excluded, i.e., only between the *"POEFS"* criteria, and asked about their opinion towards the role of Safety in the project evaluation process. There were three main reasons for adding this step to the case study:

1. Laws and regulations already prescribe that a safety risk analysis should be performed for all procedural changes in flight operations. This specific analysis will take place in parallel to the integrated evaluation of process innovation, regardless of the outcome of this study.
2. Preliminary discussions showed that there is currently no consensus among airline managers on whether Safety should be solely a precondition for project selection, or also be included in the trade-off between the other business objectives. By asking for the pairwise comparison input on both SPOEFS and POEFS, the results of this research are valid in both scenarios.
3. It was assumed that some managers place such high importance on Safety that – despite the structured approach of BWM – the anchoring effect skews the evalua-

[2] The Safety Review Board (SRB) is a high level committee which considers strategic safety functions (UK Civil Aviation Authority, 2008). The board is chaired by the Accountable Manager (in the case of Transavia the Chief Operations Officer) and includes the Senior Vice Presidents of the operational domains.

tion of the other dimensions. By asking them to also evaluate the dimensions with the Safety indicators excluded, the decision-makers were encouraged to use the whole range of the 1-9 scale.

3.4.3. DATA ANALYSIS

After collecting the vectors A_B and A_W from each of the 17 decision-makers, two parallel data analysis approaches were applied (see Figure 3.1).

REFLECTION ON PRIORITIES OF INDIVIDUAL DECISION-MAKERS

The first approach was based on the linear approximation of the original BWM. The goal here was to find the optimal weight vector separately for each manager. The individual weight estimations were visualized in an interactive strip plot using Python. The way in which the weights were depicted shares some characteristics with the *"dot plot"* published by the Federal Reserve. Whereas the Federal Reserve publishes a chart which maps the individual policymakers' projections for future interest rates, the HV Dot Plot maps out each manager's estimation for the weight of key attributes in the evaluation of process innovations in flight operations. In the absence of a broad-based consensus on inter-criteria preference, the HV Dot Plot serves as a guide for airline decision-makers. Unlike the interactive version, the dots in this published report are not labeled with the names or positions of the managers who submitted the estimation. Mainly for confidentiality reasons. For internal HV use, the unredacted interactive plot can be requested from the author. However, as such the Dot Plot already shows the most important information to fuel productive debate within airline management teams. The opacity of the individual markers was set up in such a way that the visibility of both spread and density was maximized. Thus, decision-makers can get a grasp of where consensus exists and where opinions diverge.

The Dot Plot also displays some summary statistics, including the arithmetic mean and standard deviation. However, when individual decision-makers want to compare their weights to the central tendency, it would be more statistically relevant to compare themselves to the median weight, which is not skewed by outliers. Correspondingly, the median weight is highlighted in the Dot Plot as well, since it represent a possible central path for airline policy-making in regard to sustainable projects. For quick analysis it might also be useful for airline planners to compare the relative distance between the mean and median weights. If they are close together it suggests that the weight estimations are more homogeneous and if there is a large distance between them there will typically be less consensus among decision-makers.

This might still be misleading however, as both the arithmetic mean and the median are no good indicators of central tendency for type of data concerned in this study. The median is more robust with respect to outliers compared to the arithmetic mean, for the simple fact that it completely ignores them. However, in small data sets such as the pairwise comparison data from 17 DMs, those outliers might not be as irrelevant as initially assumed. One may be inclined in this case to classify some data points as outliers, for example the head of flight crew training might be expected to assigning a significantly higher weight to safety than his peers. Still, just because some of the observations might be anecdotally explained, it does not mean that they can be disregarded from a *statistical* standpoint. For this reason, the geometric mean is used as well. This measure of central tendency is calculated using Equation 3.16, where $\bar{w}_g$ indicates the geometric mean of the weight and $w_i = \{w_1, w_2, \cdots, w_n\}$ are the individual weight estimations. Aside from the practical considerations – while it is not uncommon in the literature to use the arithmetic mean for weight aggregation – it is more theoretically sound to use the geometric mean for ratio data, regardless of the sample size.

$$\bar{w}_g = \left(\prod_{i=1}^{n} w_i\right)^{\frac{1}{n}} = \sqrt[n]{w_1 w_2 \cdots w_n} \tag{3.16}$$

Reflection on the probabilistic approach

In the second approach, the input from the 17 managers was analyzed using the Bayesian model proposed by Mohammadi and Rezaei (2020). The difference between the two approaches is that in the first approach the weight of each individual decision-makers was first obtained using the linear model and then summarized using different measures of central tendency, while in Bayesian BWM the Dirichlet distribution for w^{agg} is computed by aggregating all individual data at once in a probabilistic model. Although the first approach allows for reflection on the individual weights, speculating whether one criterion is preferred to another based on averages is a rather binary approach to such a complex decision-making problem. Instead, Bayesian BWM allows for comparison between criteria in a more "colorfull" way (both metaphorically and literally, as we will see in the next chapter) (Mohammadi and Rezaei, 2020).

To that end, the Bayesian approach applies the notion of credal ranking, which can assess the degree to which one criterion is superior to another. From the modeled Dirichlet distribution of w^{agg}, a confidence level was assigned to the preference relation between the criteria. Based on Mohammadi and Rezaei (2020) the following definitions were used:

1. For a pair of criteria c_i and c_j, the credal ordering O is defined as: $O = (c_i, c_j, R, d)$, where R is the relation between c_i and c_j and $d \in [0,1]$ represents the confidence of the relation.
2. For a set of criteria $C = (c_1, c_2, ..., c_n)$ the credal ranking is a set of credal orderings which includes all pairs (c_i, c_j). for all $c_i, c_j \in C$.

Mohammadi and Rezaei (2020) devise a Bayesian test to find the confidence of each credal ordering. The test is predicated on the posterior distribution of w^{agg}. The confidence that c_i is superior to c_j is computed using:

$$P(c_i > c_j) = \int I(w_i^{agg} > w_j^{agg}) P(w^{agg}) \tag{3.17}$$

where $P(w^{agg})$ is the posterior distribution of w^{agg} and $I = 1$ if the condition in the subscript holds, otherwise $I = 0$. The integration can be approximated by the samples obtained via the Markov Chain Monte Carlo. With Q samples from the posterior distribution, the confidence level can be calculated using:

3

$$P(c_i > c_j) = \frac{1}{Q}\sum_{q=1}^{Q} I(w_i^{agg_q} > w_j^{agg_q}) \tag{3.18}$$

$$P(c_j > c_i) = \frac{1}{Q}\sum_{q=1}^{Q} I(w_j^{agg_q} > w_i^{agg_q}) \tag{3.19}$$

where W^{agg_q} is the q^{th} sample of w^{agg} from the MCMC samples. Using these notions, the confidence that one criteria is superior to another was computed. By adding the confidence dimension, airline managers were provided with more nuanced information which could improve their judgment on complex project selection problems. To facilitate this process the outcome of the credal ranking was visualized in a weight directed graph, a feature that was programmed into the MATLAB application by Mohammadi and Rezaei (2020). The nodes in this graph represent the criteria and each edge $c_i \xrightarrow{d} c_j$ denotes that criterion c_i is more important than c_j with confidence level d. Additionally, the mean of the Dirichlet distribution of w^{agg} was compared to the averages from the traditional BWM.

Finally, the total output from the research activities was synthesized to advise the airline on the selection of the right set of projects that combine their corporate sustainability strategy with other important operational and commercial requirements. Sub-RQ IV was answered by providing recommendations to Transavia related to the decision support system for the selection of sustainability-oriented process innovations. This was not limited to the evaluation of the current projects. Rather it described how lessons learned from this study can be used to replicate the decision process when alternatives, criteria, and preferences change.

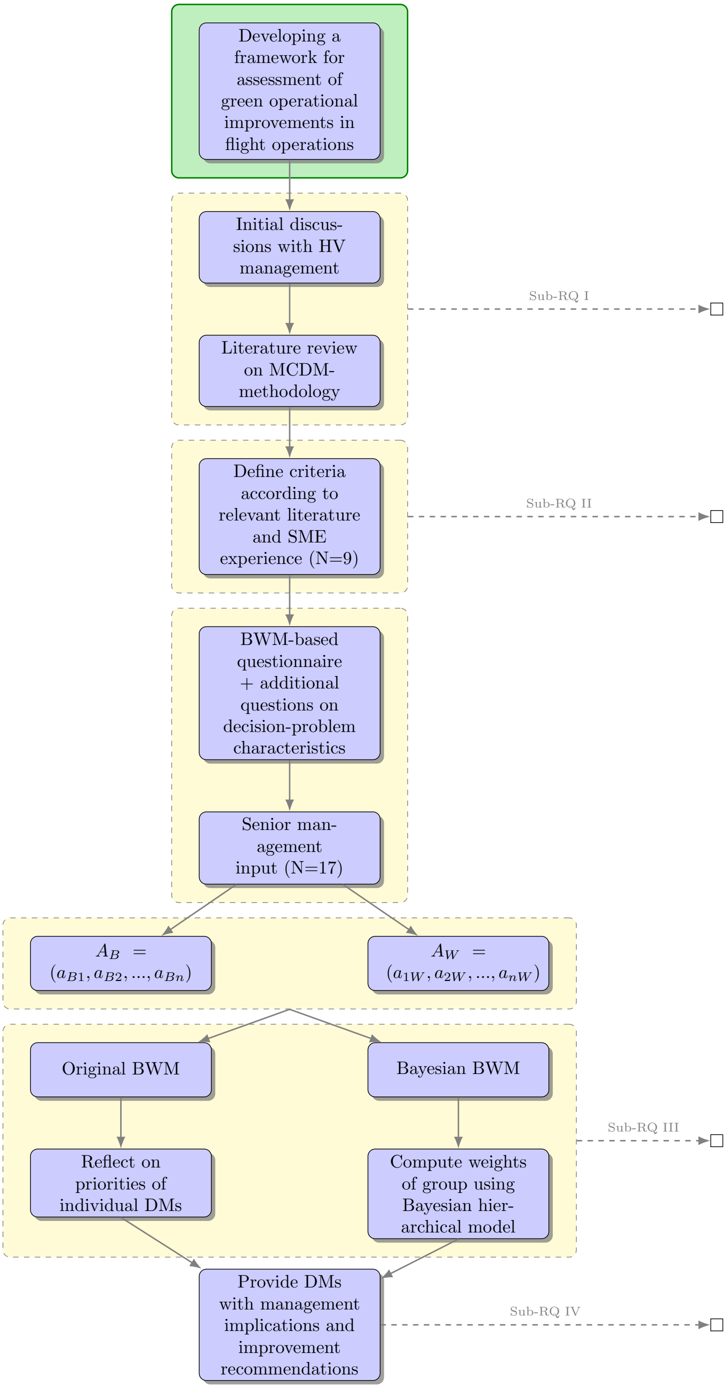

Figure 3.1: Overview of the research design.

4

Results and discussion

This chapter presents and reflects on the output from the research activities discussed in Chapter 3. First, the evaluation criteria are defined based on expert opinion, after which their relative preference according to airline management is determined. The weight estimation includes results from both the aggregation of individual preferences (AIP) as well as the novel Bayesian approach. While AIP will be more intuitive and can be explained to management with relative ease, Bayesian BWM allows to reflect on the confidence level related to the ranking. Next, the results from both methods are compared and discussed. Finally, the additional considerations related to the development of the decision support framework that came up during the interviews are presented.

4.1. Evaluation Criteria

This section first presents the conceptual framework synthesized from different experts within the firm to predict innovation project success in flight operations. Figure 4.1 (p.44) visualizes the most important considerations related to the six perspectives for predicting project performance. It should be emphasized that in this case the elements below the main criteria are no sub-criteria. Instead they are a clarification of what the criteria mean at the level of the innovation project according to relevant experts. As such, they are intended to provide guidance to the 17 key decision-makers, in order to facilitate the pairwise comparisons between the strategic goals.

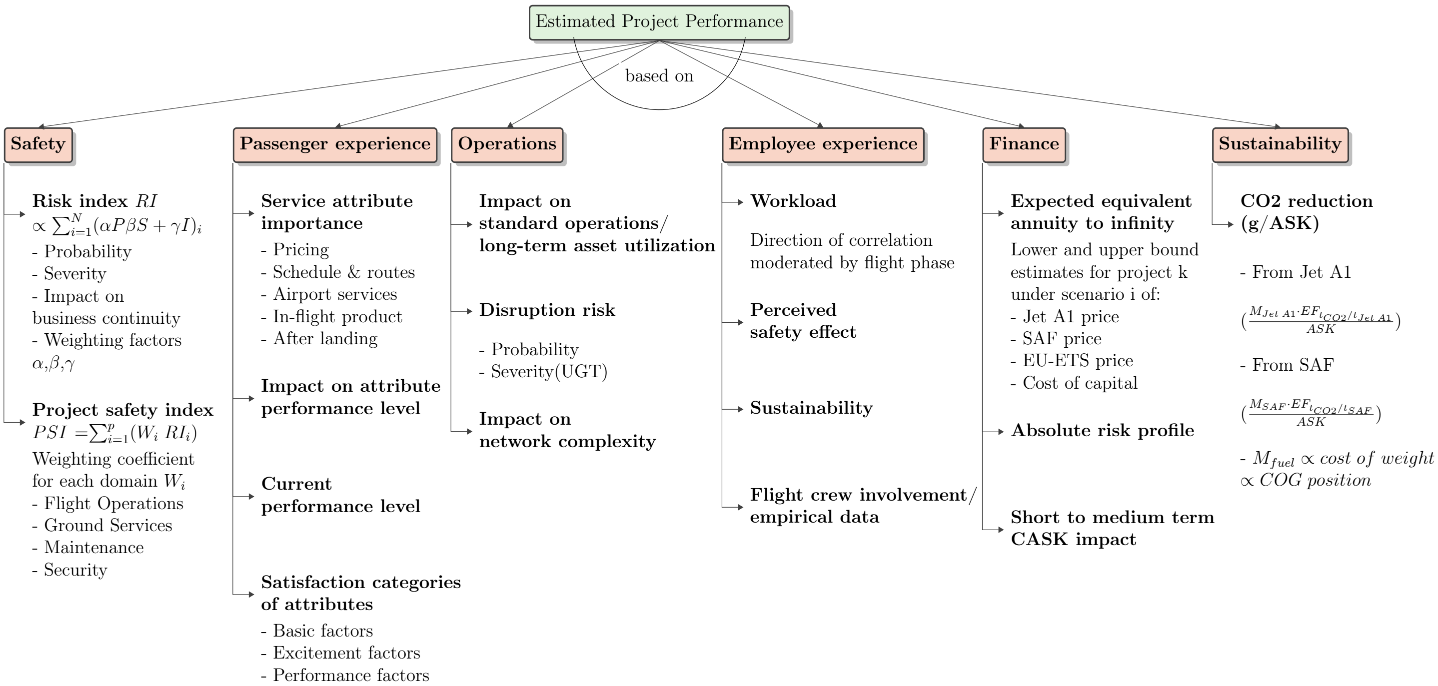

Figure 4.1: Hierarchy of the project performance estimation problem.

The framework presented in 4.1 does not include additional dimensions from the 2 outside experts. The main criterion they highlighted, which was not already mentioned by the company experts, was technological maturity. For the problem concerned here however, this is a qualification rule rather than an evaluation criterion. Such rules describe dimensions to which decision-makers are indifferent as long as a certain threshold value is met, but which can disqualify alternatives if it is not met. In the case of large scale operational projects, only the highest maturity levels are acceptable. While the technological maturity level allows alternatives to be screened and shortlisted before they are evaluated more comprehensively against the main criteria, this is not something that will be part of the high-level strategic analysis by the 17 key decision-makers. Projects with a sub-optimal maturity level will ideally not reach the stage in which they are evaluated through the framework presented in figure 4.1.

It is also important to highlight the fundamental differences between the kind of projects relevant to the airline, and the type of innovations the Venture Capital (VC) subsidiary is interested in (Table 4.1). As such, it is not surprising that the VC has some additional evaluation requirements. In addition to the solution's performance on SPOEFS dimensions, they are for instance equally interested in the attributes of the involved founders, their funding, and their business models.

Criteria for projects of the high-risk, high-reward type that have yet to cross the valley of death (as described e.g. by Klitsie et al. (2018)) will not be considered in the remainder of this section. This is also conform the actual situation at the company where there is, generally a clear separation between the activities of the airline and the VC. In accordance with the framework in Figure 4.1 the evaluation criteria are the six SPOEFS dimensions. Their definitions based on the interpretation of the relevant company experts are presented in the following subsections.

Table 4.1: Differences between Transavia Airlines CV and Transavia Ventures BV with respect to innovation projects.

	Transavia Airlines CV	Transavia Ventures BV
Goal	Immediate reduction of the carbon footprint; improving the efficiency of current operations	Strengthen future ability to reduce the carbon footprint; tap into new revenue streams
Scope	Process innovation	1) Product innovation; 2) process innovation
Source	In-house development; long term partnerships; trusted suppliers	Equity in startups; strategic partnerships with startups; ecosystem venturing
Horizon	Short to medium-term	Long-term

4.1.1. Safety

For most airline decision-makers, the definition of the safety dimension requires little explanation. According to Regulation (EU) No 965/2012 on Air Operations, Annex III (Part-ORO) of the European Parliament and Council (2012), senior airline management is required to continually promote the safety policy to all personnel and demonstrate their commitment to it (European Aviation Safety Agency, 2016). Furthermore, it prescribes that airline operators should manage safety risks related to a change in operations. This should be a formal documented process to identify the changes that may have an adverse effect on safety. It should make use of the operator's existing hazard identification, risk assessment, and mitigation processes laid down in the airline's operational manuals.

Accordingly, all involved decision-makers are assumed to be familiar with the considerations related to proactive risk management and management of change. However, the Safety Engineer who participated in the study did highlight the importance of making the distinction between *safety* and *compliance*. While compliance is a necessary condition for safety, it is not sufficient. Thus, like technological readiness, compliance can be treated as another qualification rule. The quality dimension of new projects, i.e., compliance to legal norms, is in essence binary and non-negotiable. Consequently, it is no MCDM trade-off and should be assessed lower in the organization. Including compliance would also be considered double-counting, i.e., non-compliance would be accounted for in the scores for other SPOEFS dimensions, including but not limited to the main safety criterion.

As opposed to compliance, a safety component that does present a trade-off is operational risk management. According to Hubbard (2014), risk can be defined as "a state of uncertainty where some of the possibilities involve a loss, catastrophe, or other undesirable outcome". While risks need to be mitigated to prevent adverse outcomes, it is impossible to account for all risks, as this would most certainly mean that the company loses its competitiveness. Reason (1997) illustrates this dilemma between protection and production as a broad acceptable operating zone demarcated by two extremes: bankruptcy and catastrophe. Bankruptcy occurs when resources spent on protection exceed the level of what is reasonably practicable. The other extreme describes a scenario in which protection falls short to such an extent that it results in a catastrophic accident.

At Transavia, and every airline for that matter, the navigation along the center-line of the production-protection space is done using formalized risk management matrices, resulting in a small, medium, high, or substantial risk. These linguistic risk levels could – in a further iteration of the framework – easily be quantified using fuzzy logic or a simple 1 to 5 scale. For the current conceptual framework however, the risk index RI in its general form can be described using:

$$RI \propto \sum_{i=1}^{N} (\alpha P \beta S + \gamma I)_i, \tag{4.1}$$

where P, S, and I are respectively the estimated probability, severity, and impact on business continuity of a potential adverse consequence i related to a process innovation; α, β, and γ denotate the weighting factors of the respective variables.

Furthermore, two adverse events with identical risk indices might not be evaluated equally by decision makers if they occur in different domains. In addition to events related to flight operations, cabin operations, maintenance and ground services, which were already listed prior to the interview, the consulted HV safety expert called attention to the security component of the risk assessment. Subsequently, the total project safety index PSI can be determined using:

$$PSI = \sum_{i=1}^{p} (W_i \, RI_i), \tag{4.2}$$

where W_i is the weight assigned to domain i and *i={flight operations, cabin operations, maintenance, ground services, security}*. Weights α, β, γ, and W_i can be estimated for example using linear BWM or fuzzy sets. The safety expert indicated that under the current guidelines $\gamma = 0$ and $\alpha = \beta$. Reasons for this are that for most commercial operators, the business continuity I is strongly if not perfectly correlated to risk level PS, and the importance placed on probability is assumed to be equal to the importance placed on severity. As we know, the latter assumption does not always accurately reflect human thinking. Low probability, high severity events are sometimes perceived as less risky than high probability, low severity events with similar objective risk profiles. Examples include the case of the 2008 Financial Crisis or COVID-19. The same can be true vice versa as well, e.g, the risk of coal relative to nuclear power.

4

It should be noted however, that currently Formulas 4.1 and 4.2 are for guidance only. Developing them further would not be very meaningful, as there is currently no consensus on whether safety should even be included in the evaluation. Moreover, the Safety Engineer emphasized that the seemingly unambiguous risk management templates laid down in internal documentation and guidance material from the regulator, do not always correctly describe the actual situation. The normative concept *"As Low As Reasonably Practicable"* (ALARP) will always give rise to discussion. And however detailed the risk analysis, the expert pointed out that it is not uncommon that it gets overruled based on the experience from the Accountable Manager or Nominated Persons.

As discussed, the definition of the safety dimension is relatively clear to the managers involved. Thus, assigning relative preference data to it should not be a problem. However, its fundamental role in the decision framework is a controversial topic. This will be further discussed in Section 4.5.1.

4.1.2. PASSENGER EXPERIENCE

At the level of the firm, passenger experience is measured using the Passenger Experience Index (PXI). The PXI is very broad, multivariate concept. It asks passengers to rate on a scale from 1 to 10 statements such as: "Looking back on the experience, Transavia was the best available option". The company's PXI expert mentioned it is currently not possible to say how the tracker correlates with a change in certain service attributes, especially when the changes are subtle. The fact that the firm's communication strategy strongly moderates the impact of the change, makes the PXI even harder to predict.

The company's data expert on consumer insights showed internal research that suggest 40% of the passengers are willing to pay more, if the premium is invested in sustainability efforts. It is however unknown how much the ticket prices can increase before this figure starts to drop. This is also rarely included in the academic literature. Results from Rice et al. (2020) indicate the increased willingness to pay in hypothetical scenarios is limited to 15% of the ticket price. The question remains whether hypothetical willingness to pay can provide useful statistical information regarding the actual willingness to pay in this context. And of course, pricing is only one aspect of the customer proposition. No data exists on whether passengers are also willing to accept a lower service level.

As a consequence of the unpredictable nature of the PXI, it is not suitable to use for estimating the effect of innovation projects on customer satisfaction. Recognizing the lack of data, the relevant expert saw benefit in taking a broader view on passenger experience, rather than referencing the complex PXI indicator.

A simple but effective tool to assess whether a decrease in service attribute performance is acceptable, is the importance – performance analysis (IPA) (Matzler et al., 2004). It analyses quality attributes on two dimensions: their performance level (satisfaction) and their importance to the customer, and combines them into a 2x2 matrix (Figure 4.2).

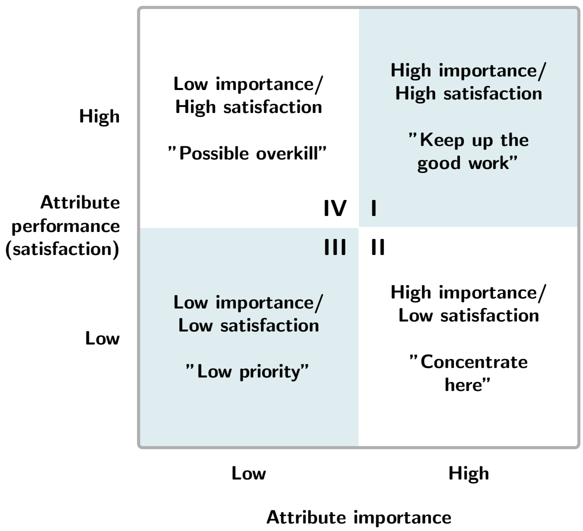

Figure 4.2: Importance performance analysis quadrants - by author, based on Matzler et al. (2004).

IPA has been applied in air transport research as a mechanism for multi-criteria passenger satisfaction measurement, e.g. by Tsafarakis et al. (2018). However, IPA relies on two implicit assumptions that are not in line with the latest paradigms in marketing management: (i) attribute performance and attribute importance are independent variables, and (ii) the relationship between quality attribute performance and overall performance is linear and symmetrical. To assess the full impact of the projects on the service level, one should also consider the inherent asymmetries. The three-factor model for customer satisfaction developed by Kano (1984) aims to take these into account by dividing service elements into three categories with differently shaped utility curves (Figure 4.3, p.49):

- **Performance factors** have a linear utility curve and lead to satisfaction if performance is high and to dissatisfaction if performance is low. Note that the IPA model assumes all attributes fall into this category.
- **Basic factors** (dissatisfiers) have a concave utility curve. They cause dissatisfaction if not fulfilled but do not lead to customer satisfaction if fulfilled or exceeded. Factors in this category are characterized by loss aversion bias (discussed by Kahneman (2011)).
- **Excitement factors** (satisfiers) have a convex utility curve. They increase customer satisfaction if delivered but do not cause dissatisfaction if they are not delivered.

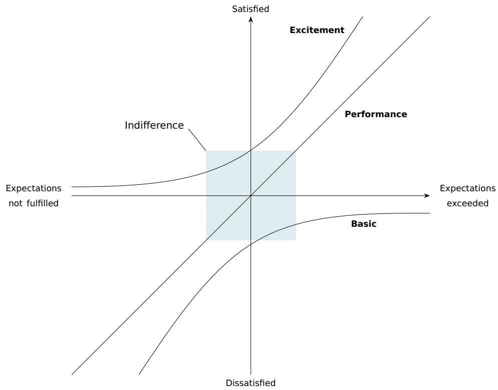

Figure 4.3: Three-factor theory for customer satisfaction - by author, based on Kano (1984).

There is currently no scientific literature to suggest in which categories the airline service attribute fall. This will also be specific to the airline business model. For a full-service network carrier the in-flight product might be a performance factor or even a basic factor, while for LCCs it may be an excitement factor. The slopes and x-intercepts of the curves are also different in each context, which further complicates the analysis.

The modeling of the utility curves is far outside the scope of the current study. Nevertheless, it is wise to consider – at least quantitatively – the asymmetries related to the impact of new projects. Accordingly, these components were included in the definition of the PX dimension in Figure 4.1.

4.1.3. OPERATIONS

Since operational excellence is the core business of any airline, the operational dimension was relatively simple to define. At the level of the firm the operations dimension is represented by the on-time performance (OTP) indicator, i.e. the percentage of flights that arrive within 15 minutes of the target on-blocks (arrival) time.

As OTP is an unpredictable and lagging indicator, it is not directly relevant for the proactive estimation of innovation project performance. Hence, the foremost operations expert within the airline was asked to explain the main OTP-drivers applicable in this context. The expert input was synthesized into the conceptual framework illustrated in Figure 4.4.

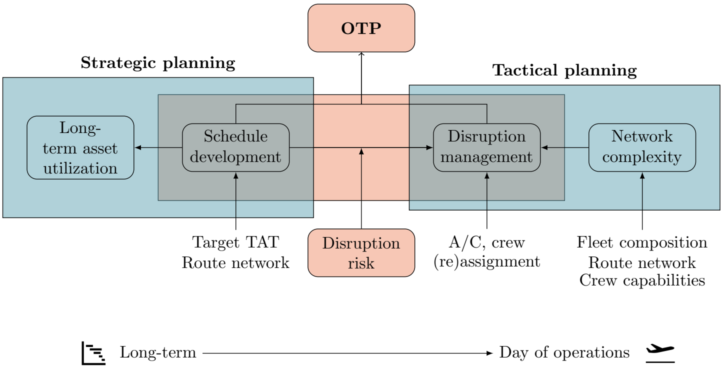

Figure 4.4: Conceptual Framework for OTP drivers - by author based on HV OTP expert input.

It shows that OTP is the result of the interplay between long-term strategic airline planning and short-term tactical airline planning. In this framework disruption risk is introduced to a long-term plan. The ability to deal with this risk is moderated by the airline's disruption management capabilities, which in turn depend on the complexity of the network. Disruption management involves the (re)assignment of aircraft (A/C) and crew to (re)scheduled flights. Network complexity is a function of the composition of the fleet, the route network, and the capabilities of the crew. More variation within these inputs usually leads to reduced efficacy of disruption management practices. For instance an operator could reprogram the engine mapping of certain A/C in the fleet to burn less fuel and emit less carbon. However, this could limit the peak output needed to take-off from small airfields, putting constraints on which aircraft could be used for which mission.

On the left side of the framework, the inputs for the long-term schedule development are the route network composition and the target turnaround time (TAT), i.e., the time allowed for ground handling operations between arrival and the departure of the next flight. Staying on the left side of the framework, potential long-term changes related to the innovations can call for the need to change the schedule, leading to a sub-optimal asset utilization. Alternatively, it could be fitted in the existing schedule and possibly affect the disruption risk related to standard operations. This would subsequently lead to a higher standard deviation of the OTP.

Based on the OTP framework, and together with the relevant expert, the three defining elements that should be included in the problem hierarchy were determined:

1. Impact on standard operations/long term asset utilization
2. Disruption risk
3. Impact on network complexity

While some of these considerations could also fall within the finance dimension, it was decided to include them here and limit the definition of the finance dimension to the more traditional metrics, for which it is easier to develop monetary estimates. This has two reasons.

First, the operational considerations have many second and third order effects which are hard to quantify. We can estimate for example the direct cost of delays in terms of compensation for passengers and loss of revenue. However, instability in the network also appeals to the contingency management capacity and expensive tactical buffers, such as spare aircraft and standby crew that would otherwise not be necessary. In a complex network it is impossible to quantify which disruption lays claims on which part of the redundant capacity.

Second, the consideration of the operational dimension as a separate element is in accordance with the high-level SPOEFS framework that is already used by airline's decision-makers.

4.1.4. EMPLOYEE EXPERIENCE

At the level of the firm, employee experience is measured using the Employee Experience index (EXI). The EXI is a very broad and multivariate concept that is determined in a way comparable to the process for the PXI. Trying to establish a comprehensive definition of PX at the project level was characterized by similar challenges as the ones described in Section 4.1.2.

Fortunately, this MSc thesis coincided with an internal psychology research project on how flight crew attitude towards new innovation projects varies based on several predictors. The most important findings relevant to this report are summarized in Figure 4.5.

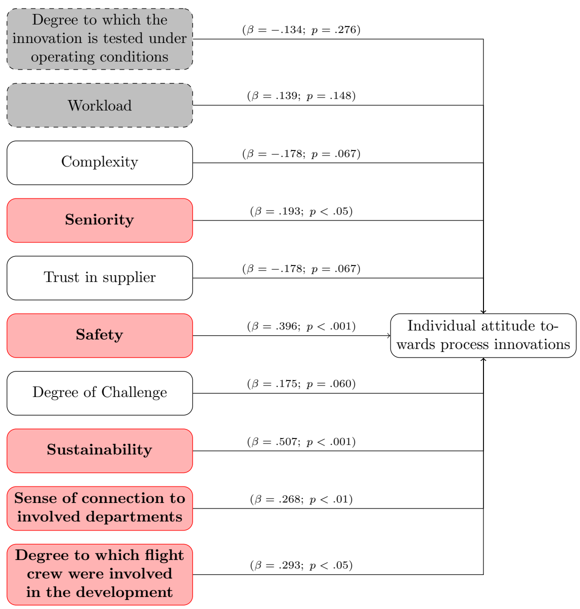

Figure 4.5: Predictors of individual pilot attitude towards innovation projects, highlighting statistically significant predictors (red) and additional important considerations that emerged from quantitative analysis (grey) - by author, based on Transavia Airlines (2021).

4

It shows the totality of factors that were investigated and highlights the statistically significant predictors (red) and the factors that were not statistically significant but emerged from additional interviews as important considerations (grey). The factors highlighted in red that are related to project attributes and were not already included under a different dimension (including safety and sustainability) were added to the problem hierarchy (Figure 4.1) to provide context for the pairwise comparisons. The non-significant predictor *"Degree to which innovation is tested under operating conditions"* was combined with the degree of flight crew involvement. Finally, workload was included as well. The fact that the correlation between workload and individual attitude was not significant can be easily explained by the fact that the relationships can go in both directions, depending on the flight phase considered by the respondents. When workload is low, additional workload is looked upon as favorable, but when workload is already high, additional tasks can be undesirable and even dangerous. Figure 4.6 and Figure 4.7 (p.53) show the flight crew workload across the different phases of flight and the cognitive explanation for the decrease in safety margin respectively.

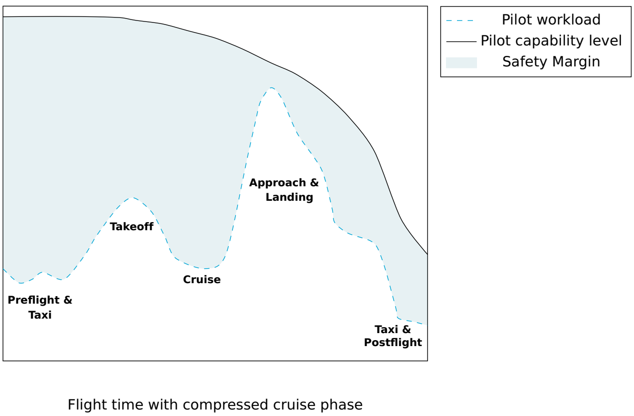

Figure 4.6: Pilot workload versus flight phases - by author, based on Transport Canada (2017).

4.1.5. FINANCE

Traditional financial project appraisal is based on the expected net present value (NPV), absolute risk, and relative risk related to the project. The decision-maker also has to assume different economic scenarios and their respective probabilities. In the context of this research it would involve the lower and upper bounds of the estimates for the price of traditional jet fuel (Jet A1), sustainable aviation fuel (SAF), and the CO2 price under the EU Emission Trading System (EU ETS), as well as estimates for the cost of capital. Then, the expected NPV, absolute risk, and relative risk for each individual project can be calculated using Equations 4.3, 4.4, 4.5, and 4.6 respectively.

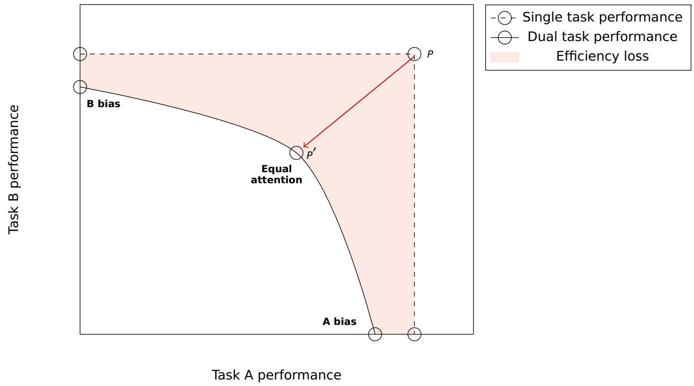

Figure 4.7: Performance operating characteristic curve, illustrating the cognitive psychology concept underlying Figure 4.6, where a move of task performance towards the origin will first result in mere discomfort for the flight crew but eventually leads to exceeding the safety margin - by author, based on Proctor and Zandt (2018).

$$V_{ki} = \sum_{j=0}^{n_k} \frac{F_{kij}}{(1+R'_{ki})^j} \tag{4.3}$$

$$\mu_{V_k} = \sum_{i=1}^{m} V_{ki} P_i \tag{4.4}$$

$$\sigma_{V_k} = \sqrt{\sum_{i=1}^{m} (V_{ki} - \mu_{V_k})^2 P_i} \tag{4.5}$$

$$C_{V_k} = \frac{\sigma_{V_k}}{\mu_{V_k}} \tag{4.6}$$

F_{kij} denotes the cash flow for project k under scenario i at period j; V_{ki} and R'_{ki} respectively denote the NPV and the cost of capital for project k under scenario i. P_i denotes the probability for scenario i; μ_{V_k}, σ_{V_k}, and C_{V_k} respectively denote the expected value, absolute risk, and relative risk of the NPV for project k.
For projects with different time horizons the equivalent annuity can be calculated using:

$$_{EA}V_{ki} = V_{ki} A_{n_k, R'_{ki}} = V_{ki} \frac{1-(1+R'_{ki})^{-n_k}}{R'_{ki}}, \tag{4.7}$$

And for projects with unequal risk levels the equivalent annuity to infinity can be determined using:

$$_{EAI}V_{ki} = \frac{_{EA}V_{ki}}{R'_{ki}} \tag{4.8}$$

While it must be noted that these specific formulas were adapted from Tsao (2012), they do not differ conceptually from the typical financial management textbook. As part of the standard business case, similar calculations will be made for each innovation project. However, with a view on the long term profitability of the firm, the airline's financial expert favored a process in which the primary concern of the key decision-makers should be *affordability*, not return on investment (ROE). Accordingly, when it comes to sustainability-oriented operational improvements, the main consideration in the finance domain should be that – while focusing on the reduction of the carbon footprint in the long term – the absolute risk σ_{V_k} is kept at an acceptable level to guarantee a competitive CASK [1] in the short to medium-term.

4.1.6. SUSTAINABILITY

The total negative environmental externalities from aviation include the climate impact (carbon and non-carbon), noise impact on local communities, impact on air quality, impact on biodiversity, and resource depletion (Delft University of Technology and Netherlands Aerospace Centre, 2021). However, based on the current understanding in the academic literature of the climate impact of aviation, and discussions with the airline's sustainability and flight-technical experts, it was decided to narrow the scope of the sustainability dimension to carbon emissions only. More specifically the sustainability dimension was defined as the efforts towards reducing the *relative* carbon emissions, i.e. the amount of carbon emitted per unit of production measured in grams per available seat kilometer (g/ASK).

For aviation, the total anthropogenic (caused by humans) effect is predominantly determined by the additional radiative forcing of CO2 and non-CO2 emissions related to burning jet fuel. The most comprehensive quantitative assessment of aviation climate forcing terms to date is presented by Lee et al. (2021). For the purpose of this report, their most important findings have been summarized in a single figure (4.8, p.55), where the diamonds show the best estimates for the respective climate forcing factors, the whiskers show the 95% confidence interval, and the blue and red shading indicate whether the term is a climate cooling or warming effect respectively.

From the confidence intervals can be concluded that the climate impact of CO2 emissions is the only major factor that is relatively well understood. This is because the way in which CO2 interacts with the climate is linear, and directly proportional to the amount of fuel burned, unlike the non-CO2 factors. Still, even when taking into account the uncertainty related to the other factors, the current best guess is that aviation emissions are warming the climate at approximately three times the rate associated with aviation CO2 emissions alone (Lee et al., 2021).

This raises the question why the sustainability dimension is only defined by the carbon footprint. There are two important reasons for this decision:

1. **The policy environment**: According to Regulation (EU) 2018/1999, recognizing the complexity of measuring non-CO2 climate terms – and the uncertainty regarding trade-offs between the various impacts – the European Commission currently has not specified hard targets for airlines regarding non-CO2 emissions (European Parliament and Council, 2018).

[1] Cost per Available Seat Kilometer: the most important measure for unit cost in the airline industry.

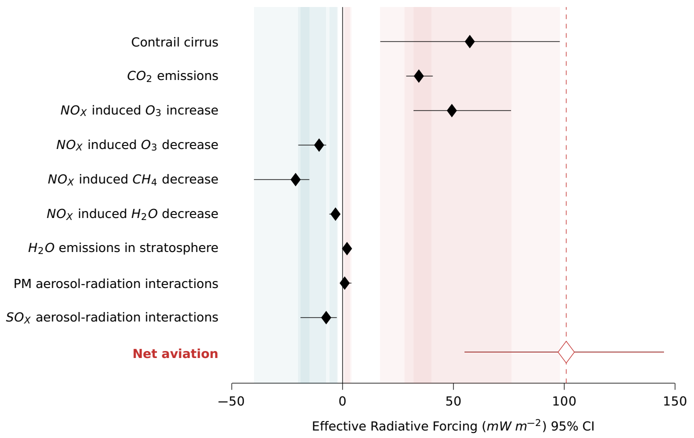

Figure 4.8: Best estimate of effective radiative forcing of global aviation from 1940 to 2018 - by author, based on Lee et al. (2021).

2. **Focus on efficiency**: The operational innovations that will be considered by the decision-makers aim to increase operational efficiency and reduce energy consumption. Unlike the development of engine technology, which can involve a trade-off between carbon and NOX emissions, streamlining operational procedures or reducing the weight carried on board are always worthwhile pursuits for both the reduction of CO2 and non-CO2 emissions. For these types of innovations CO2 will remain a useful proxy for sustainability, even with improved understanding of non-CO2 effects. While CO2 emissions are directly proportional to fuel burn, NOX for example still scales with increased kerosene consumption, be it in a non-linear way (see Figure 4.9).

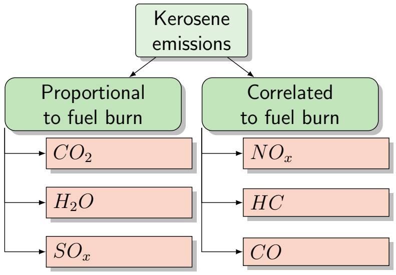

Figure 4.9: Proportional and correlated kerosene combustion emissions - by author, based on Schaefer and Bartosch (2013).

For medium range air transport, the CO_2 proxy will stay relevant until at least 2040, when the first hydrogen powered aircraft with 166-250 passenger capacity are expected to be introduced (McKinsey & Company, 2020). Until then, aircraft will be powered by a mix of conventional Jet A1 fuel and SAF. While SAF has the potential to reduce the climate impact by 30-60%, its price ranges currently from 2 to 7 times the price of Jet A1 (Delft University of Technology and Netherlands Aerospace Centre, 2021). Although company finance and sustainability experts expect this figure to drop, it is clear that – at least for time horizon considered in this research – the relative carbon footprint is an accurate predictor of project success in the sustainability dimension.

Thus, in the problem hierarchy sustainability is defined as CO_2 reduction expressed in grams per ASK. Where the weight of the fuel saved $M_f uel = \{M_{Jet\ A1}, M_{SAF}\}$ can be converted into grams of CO_2 through the respective emission factor EF, e.g., 1 kg jet A1 = 3.15 kg CO_2. For weight reduction efforts, M_{fuel} is proportional to the cost of weight factor, i.e., the amount of fuel saved per flight hour for every kg weight reduction, which in turn depends on the average center of gravity (COG) position of the aircraft. For the evaluation of individual projects or when comparing a weight reduction project with a procedure change this latter part is non-trivial. The base line cost of weight factor for the Boeing 737 is around 3%, however this may change significantly based on the operators preferences (Boeing, 2021). COG position is therefore an important additional consideration for the evaluation of sustainability-focused operational innovations.

4.2. Criteria Weights: Aggregation of Individual Priorities

To ensure that all DMs performed their comparisons based on common definitions of the SPOEFS criteria at the project level, a brief summary was provided on the results from the previous section. Subsequently, comparison data was collected from seventeen senior airline managers according to the procedure described in Chapter 3.

Figure 4.10 (p.57) shows the weights of the six SPOEFS criteria based on the aggregation of the individual preferences of the seventeen DMs. The consistency ratios are all close to zero, ranging from 0.0175 to 0.2308, indicating a high reliability of the results.

As can be seen in the SPOEFS Dot Plot (4.10), Safety (weight = 0.4187) is by far the most important aspect for evaluating operational innovations. It is followed by four criteria between which there seems to be no evident relative preference: Employee Experience (weight = 0.1399), Passenger Experience (weight = 0.1241), Finance (weight = 0.1222), and Sustainability (weight = 0.1164). Operations (weight = 0.0788) is clearly the least important criterion.

In the results, the position of safety is in line with what one expects based on the legal responsibilities of airline organizations. A large part of the managers involved in the evaluation have a legal obligation under EU 965 to constantly demonstrate their commitment to safety. It is therefore not surprising that the safety perspective is by far the most important aspect.

The Passenger and Employee criteria are characterized by a high sigma and relatively large distance between the mean and median weights, hinting at a positive skew of the data. The DMs that placed particularly high importance on PXI and EXI (N=6, three for

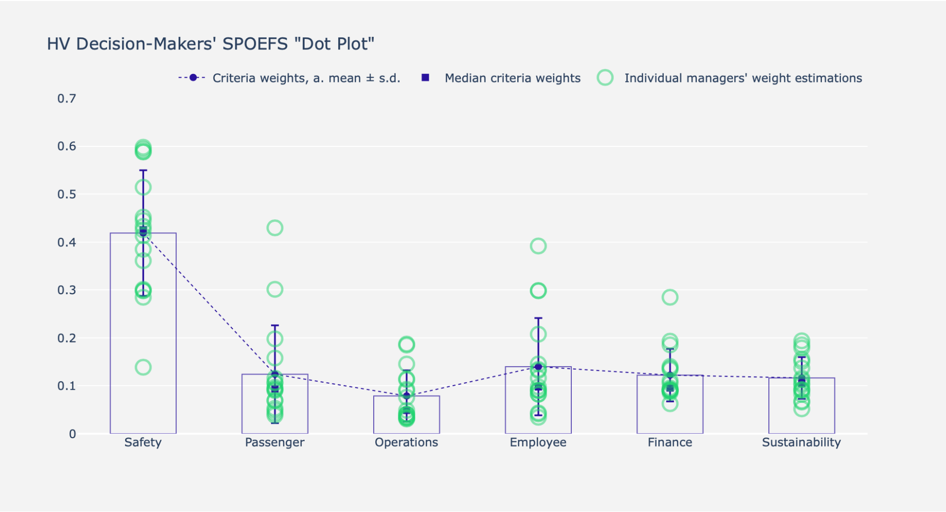

Figure 4.10: Individual SPOEFS preferences of the 17 airline decision-makers.

each dimension) were asked to clarify their position. They were all of the opinion that the satisfaction of respectively the passenger or employee was the true leading indicator of company success.

Also noteworthy is the fact that the 3 managers who expressed a high relative preference for EXI, have all been on a field trip to Southwest Airlines, an operator known for its relentless focus on employee satisfaction. Their reasoning can be illustrated as follows:

$$EX \rightarrow PX \rightarrow profitability \tag{4.9}$$

Managers who assign a high relative weight to EX argue that, through perceived correlations between the KPIs (the existence of actual correlations was not quantitatively verified in this study), a high PXI will eventually lead to better results in other dimensions as well.

For finance this is relatively well established in the scientific literature (e.g. Bernhardt et al. (2000)). As for safety, it is also reasonably to assume that satisfied employees are more likely to adhere to the safety norms that are specified in the instructions and procedures laid down in the operations manual. However, the link between employee experience and sustainability is less clear, which gives rise to the possibility that these managers overlook the importance of reducing the carbon footprint in the short to medium-term.

Managers that were according to their job description closer to the passenger dimension, including the VP of Customer Proposition and the VP of Cabin Operations, applied the same logic as illustrated in Equation 4.9, while skipping the PXI perspective.

In line with what one might expect from a low-cost carrier, the relative importance of the operational considerations are low. As a point-to-point carrier between leisure destinations, Transavia does not transport many connecting passengers. Non-performance is in that case less detrimental to the integrity of the route network than it would be for a full-service hub-and-spoke airline. This is underlined by the fact that all four managers who in a relative sense placed the highest weight on operations, previously worked for full-service network carrier KLM Royal Dutch Airlines. The input from these managers is also the source of the large relative distance between the mean and median values. If their evaluations were to be excluded, the operations dimension would score even lower.

Additionally, some managers asserted that on-time performance is unimportant as long as passengers and crew are satisfied, but deserves more attention if it is not the case[2].

Airline leadership places significantly less importance on the Finance and Sustainability dimensions compared to Safety, while their weights are about the same as those of the EX and PX perspectives. The relatively low weight of Finance may be explained by the mentality implied from the CFO's definition of the finance dimension. It is not quite the Draghi-esque *whatever it takes*, but also not far from it (presumably because operational improvements are relatively affordable compared to other investment projects). However, this does not give a reason for why the Sustainability dimension (the original objective of the operational innovations!) is rated equally low.

Interestingly, the AIP data suggest that when managers are asked to compare the reduction of carbon intensity to other corporate objectives, it appears it is not rated as high as what one might expect from asking them to rate it in isolation. Of course, everyone will agree that sustainability is important, the question is; compared to what? This will be further reflected upon after the discussion on the probabilistic approach. Noteworthy but not unexpected is the fact that the six managers who assign the highest relative preference to sustainability are 5 DMs from the Management Team Flight Operations and the CFO. While the CFO is ultimately responsible for the sustainability dimension, Flight Operations is of course the most closely involved in the business of burning kerosene. It is therefore not surprising that these stakeholders feel responsible for the reduction of carbon emissions.

Separately, several decision-makers made reference to the relationship between the finance and sustainability perspectives, which can be both positive and negative. On the one hand, operational innovations that decrease fuel burn have a positive impact on both the carbon footprint and operational expenditures. On the other hand, if the innovation involves costly physical modifications to the aircraft, the correlation between finance and sustainability might be weakened or even inverted.

As discussed in the previous chapter, decision-makers were also requested to perform the pairwise comparisons with the Safety dimension excluded, i.e., only between the *"POEFS"* criteria. The Dot Plot for the POEFS criteria is shown in Figure 4.11.

As can be seen in 4.11 (p.59), the higher standard deviations and less overlap between the dots suggest more variation in the input preferences when other factors than safety are used as a reference point. However, the relative importance of the remaining criteria does not change significantly when safety is excluded from the evaluation. In the ranking some criteria have swapped places. Yet, this is not very meaningful when the difference between criteria is not that significant in the first place.

[2]This hints at the existence of some form of preference dependence. However, the methods applied in this study did not allow to quantitatively verify the degree of interaction among attributes.

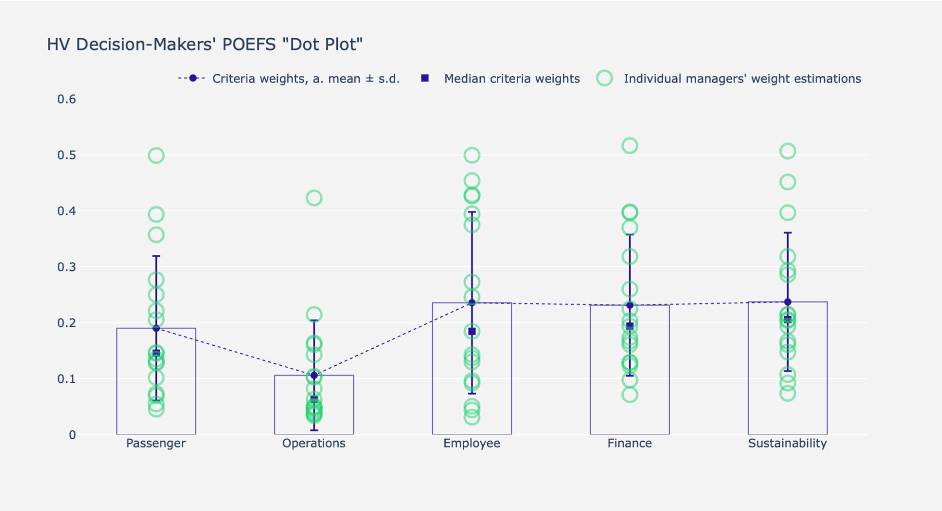

Figure 4.11: Individual POEFS preferences of the 17 airline decision-makers.

4.3. CRITERIA WEIGHTS: PROBABILISTIC APPROACH

For the assessment of the preference between one criterion over another, the credal ranking method from Bayesian BWM was applied. As opposed to the traditional way of comparing averages, which was shown in the previous section, the Bayesian approach can estimate the confidence related to the ranking.

Figure 4.12 (p.60) confirms that Safety is the most important dimension of the SPOEFS criteria, with a confidence of 100% over all other criteria. The Finance dimension ranks in second place and is absolutely more important than Operations. Although its superiority over Passenger Experience and Employee Experience is quite strong as well (94% and 82% respectively), the confidence assigned to its relationship with Sustainability of 63% percent indicates that a small subset of managers believes Sustainability should take second place instead. Based on the analysis of the individual preferences, this would presumably be favored by the MT Flight Operations and the Chief Financial Officer. Sustainability outranks Operations, PX, and EX with confidence levels of 100%, 89%, and 73% respectively. This suggest that the superiority of Sustainability over the Operations dimension and Passenger dimension is less contended, while the superiority over Employee Experience will probably be challenged by some of the decision-makers who adhere to the paradigm of Southwest Airlines (Equation 4.9). In the credal ranking Employee Experience is the fourth most important factor with a confidence of 100% against Operations. It has a confidence of 73% against Passenger Experience, most likely caused by the preference data from the managers who are inherently more customer-focused. Finally, based on the results from the AIP approach, it is not surprising that Operations is the least important dimension, with even Passenger Experience outranking it with a confidence of 98%. This is in line with what is expected from a point-to-point leisure airline with hardly any connecting traffic.

To verify the ranking resulting from the probabilistic approach, an internal team concerned with planning the future HV fleet was asked to provide an overall intuitive ranking of the SPOEFS criteria, according to their relative importance in the aircraft purchasing decision. This problem was assumed to be a good test case, as the investment in a new fleet must continue to meet the airline's strategic objectives for decades to come.

According to Schilling's (2017) definition, process innovations are innovations in the way an organization conducts its business, such as in the techniques of producing its goods or services. They are often oriented toward improving the efficiency of production. The core product of an airline is the production of available seat kilometers (ASKs). As such, the implementation of a new, more fuel-efficient machine to produce those units is the 'ultimate' process innovation. While the framework presented in this report was originally developed for less capital-intensive solutions, the high-level criteria and how they are viewed by senior airline management should not be different.

As mentioned, the team tasked with selecting the new aircraft type was asked to discuss the relative importance of the SPOEFS criteria, and come up with a unanimous overall ranking. Without having seen the results from this research, the ranking provided by the group was identical to the credal ranking in Figure 4.12. Only one of the evaluators from this group was also among the seventeen key decision-makers, from whom the pairwise comparison data was collected. The others were not consulted before in this study, which suggests intra-firm reliability of the results.

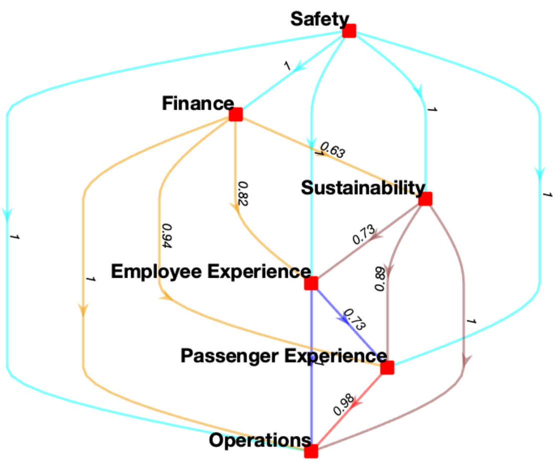

Figure 4.12: Credal ranking for the SPOEFS dimensions.

The credal ranking was also performed on the POEFS-vectors (Figure 4.13). With the Safety dimension excluded, the ranking of the factors remained unchanged, which implies the separately collected pairwise comparisons are consistent. Only the confidences are slightly changed. However, not to such an extent that there is reason to believe that the inclusion of the Safety dimension distorts the analysis. The results are analogous to the respective Dot Plots, where the distribution of the dots changed but aggregated picture stayed the same.

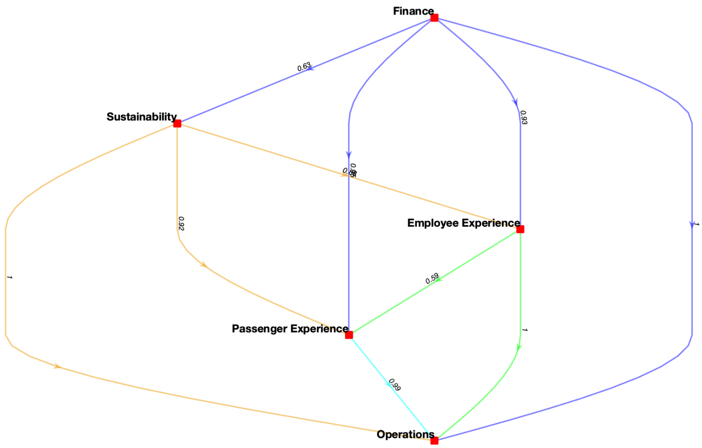

Figure 4.13: Credal ranking for the POEFS dimensions.

4.4. COMPARISON OF AIP AND PROBABILISTIC APPROACH

Table 4.2 and 4.3 (p.62) compare the averages of the Dirichlet distribution from Bayesian BWM (BBWM) to the aggregated individual preferences of the traditional approach. The results are visualized in Figure 4.14 (p.63) and 4.15 (p.64). For the aggregation of the individual preferences, both the more practical arithmetic mean of the weights $\bar{w}$ and the mathematically correct geometric mean $\bar{w}_g$ of the weights were used. By definition, the ranking resulting from the arithmetic aggregation is the same as the geometric aggregation, but with lower weights. As the geometric mean is less sensitive to outliers, the distance between both averages is larger for dimensions with more extreme values, and smaller where there is more consensus.

As shown in figure 4.14 and 4.15 the average weights from the Dirichlet distribution from the probabilistic approach do not differ in a meaningful way from the weights obtained from Bayesian BWM. Certainly, the weights are computed in a fundamentally different way. In the AIP approach, the individual weights were calculated first and than aggregated with the respective mean operators, while the Bayesian BWM computes the aggregated distribution and all the individual preferences at once using probabilistic modeling (Mohammadi and Rezaei, 2020). Consequently, if the difference between two criteria is not that significant, the swaps in the ranking as shown in 4.14 and 4.15 are

Table 4.2: Comparison of the original BWM and the Bayesian BWM on the SPOEFS evaluation criteria based on the preferences of 17 airline managers.

	Saf.	PXI	Ops.	EXI	Fin.	Sus.
BWM ($\bar{w}$)	.4187	.1241	.0788	.1399	.1222	.1164
BWM ($\bar{w}_g$)	.3956	.0988	.0642	.1118	.1135	.1088
Bayesian BWM	.3004	.1335	.1001	.1449	.1642	.1569

Table 4.3: Comparison of the original BWM and the Bayesian BWM on the POEFS evaluation criteria based on the preferences of 17 airline managers.

	PXI	Ops.	EXI	Fin.	Sus.
BWM ($\bar{w}$)	.1901	.1059	.2355	.2313	.2372
BWM ($\bar{w}_g$)	.1528	.0795	.1736	.2013	.2082
Bayesian BWM	.1909	.1225	.1983	.2502	.2382

expected.

An important difference between the AIP and BBWM is that the criteria can be compared in a probabilistic way. Based on the weight directed graph we can now say that criterion A is more important than criterion B with confidence d. This yields much more information compared to the traditional approach. Nevertheless, in this case it does not lead to different conclusions regarding the *magnitudes* of the weights. For example: while it allows to conclude that Sustainability outranks Employee Experience with confidence of 89%, it does not change anything about the fact that both dimensions have a weight of about 0.15.

4.5. DECISION PROBLEM CHARACTERISTICS

In addition to the rating on the 9-point scale, the seventeen DMs were asked about their opinion with respect to the decision-making characteristics, i.e., if they perceived that correlation, compensation, and preference (in)dependence exist between the main criteria. Subsequently, the decision-makers were also requested to reflect on whether the

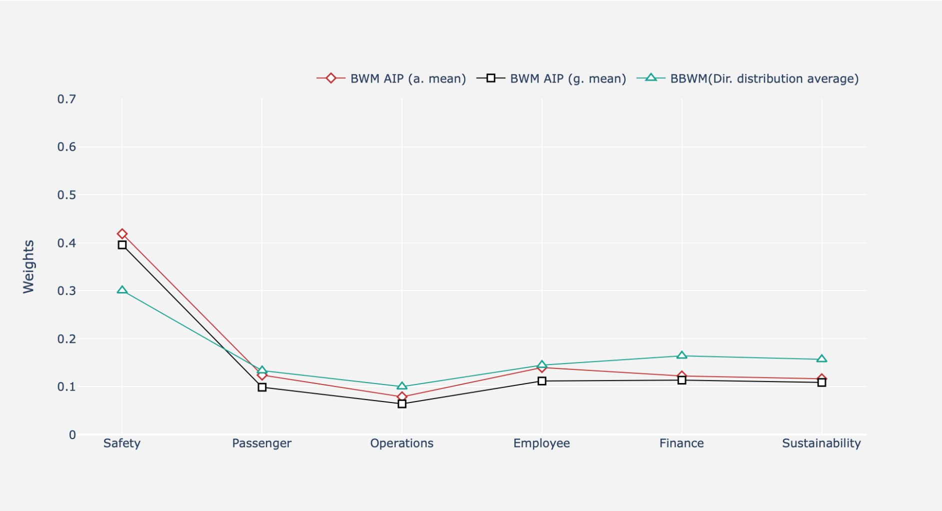

Figure 4.14: Comparison of the SPOEFS criteria weight estimations between the AIP and probabilistic approach.

Safety dimension should be included in the framework. Table A.1 in Appendix A summarizes the view of the Transavia managers with respect to the additional considerations. In this table, the zeros and ones indicate if the DMs believe correlation, compensation, and preference dependence between the criteria exist and whether the safety perspective should be integrated or not.

The questions with respect to correlation, compensation, and preference (in)dependence are definitely interesting and important. Answering them in a clear and meaningful way however, is complicated. This is not only because the respective views from airline management were scattered (as shown in Appendix A), but because it really requires a quantitative approach that did not fall within the scope of this thesis. For that reason the discussion on correlation, compensation, and preference (in)dependence is not presented among the main results of this study. An elaboration on this limitation can be found in Chapter 5. For an initial exploration of the subject and the mapping of the DMs' intuitive assessments of the characteristics, the interested reader is referred to Appendix A.

4.5.1. THE ROLE OF THE SAFETY DIMENSION

In regard to safety, the main theme that emerged during the discussions is that all DMs agree that it is the most important dimension, yet they have different opinions as to which role it should have in the evaluation of the operational innovations. As can be seen in Table A.1, ten managers would integrate safety in the comprehensive trade-off, while seven decision-makers preferred to view safety as a boundary condition.

There are two main perspectives that respectively apply the following logic: i) safety is extremely important and therefore it should be integrated, and ii) safety is extremely important and therefore it should be handled separately. Based on stakeholder input, it can be concluded that the role of safety is a topic that will continue to give rise to

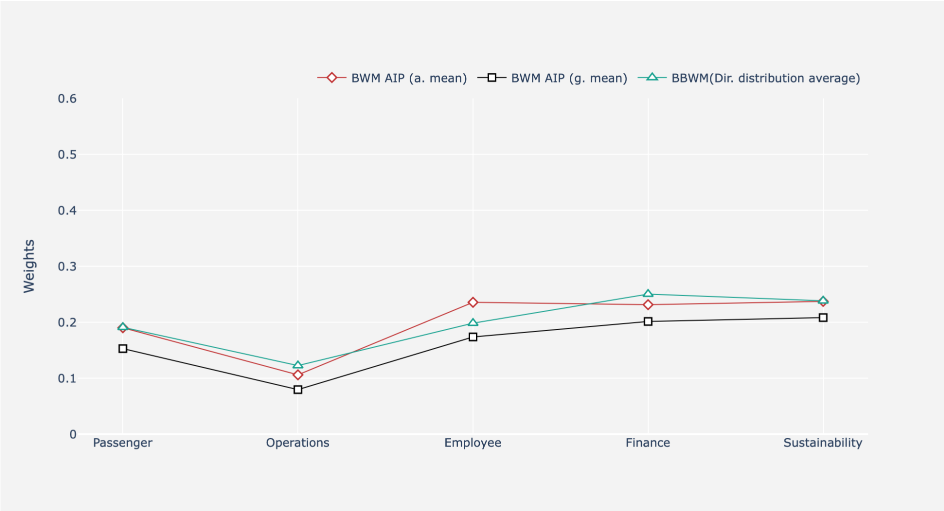

Figure 4.15: Comparison of the POEFS criteria weight estimations between the AIP and probabilistic approach.

discussions. This is illustrated in Table 4.4. It shows the divergent views of five selected airline managers who had a strong opinion on the role of the safety dimension, as well as their advise on whether or not to integrate it in the comprehensive trade-off.

In this publicly available report, the positions of the decision-makers have been redacted, and they are numbered instead. This is not because the information is highly confidential, but because through a quick online search, the positions directly trace back to the names of the senior managers involved. Besides, their positions are not relevant for the purpose of this discussion. Note that the opinions are presented in a random order and the numbers of the DMs do not correspond to the numbers in Appendix A.

Table 4.4: The view of selected senior airline managers on the inclusion of the safety dimension.

	About Safety	Advise
DM 1	Safety is simultaneously the most important and the least important dimension [see comments in third row of this table].	Decide on a case by case basis whether it should be included in the evaluation.
DM 2	Safety is often not a project attribute, but is in the DNA of the organization. Safety does not have to be expensive and is often free, so usually there is no trade-off to be made.	Integrating safety is in essence a good thing and in line with the way we do business, but be careful with the exact mechanism, as it might be misleading.

Table 4.4 continued from previous page

	About Safety	Advise
DM 3	Providing that an acceptable safety level is reached, added safety does not increase the projects desirability. A project's outstanding safety attributes should not be able to 'compensate' for dimensions on which it has sub-optimal performance.	Exclude safety from the evaluation and treat it as a qualification rule by handling it separately in the formal operational risk management system.
DM 4	The importance of safety is so ingrained into every level of the organization that a decision-maker will implicitly consider it as part of one of the other dimensions, e.g. EXI or PXI, if it is excluded.	Always include safety in the evaluation, as it is part of the organizational DNA.
DM 5	Safety is always and everywhere. It should not be in the same level of the hierarchy as the other perspectives. This also leads to a much more interesting discussion as safety does – or at least should – occupy every airline manager's foremost attention and concern. Compared to safety, the difference in relative importance of all other factors should be negligible.	Assess safety first, and then evaluate the other dimensions.

4.5.2. SPOEFS, POEFS OR BOTH: A POSSIBLE SUPPORTIVE FRAMEWORK FOR THE INTEGRATION OF THE SAFETY DIMENSION

Table 4.4 illustrates that the views regarding the safety dimension are scattered. As a consequence, we cannot clearly observe a single way forward as to how this aspect of the evaluation should be handled. This is something the firm will have to decide internally. Additionally, the views may differ depending on the scope of the evaluation; a high-level assessment or a quantitative calculation in a crisp MCDM-model.

Regardless of which of the two options is chosen, it is wise to evaluate the operational risk related to the projects pre and post reasonably achievable mitigation measures. As mentioned by one of the decision-makers, risk management does not have to be costly and is sometimes more closely related to the attributes of the organization than the attributes of the project itself. The safety expert consulted in the criteria definition phase mentioned the possibility of developing quick decision-rules for assessing the feasibility of the project from the safety perspective. This would eliminate the need for a formal and time-intensive risk assessment, before it has even been decided if the project is desirable based on the other criteria. In some instances, the hazards related to a project can have a history of similar event types of which prior data is available. In any case, the Flight Safety Engineer should be able to give a best guess estimate of the outcome of the formal assessment, based on experience and intuition.

This leaves of course a small probability that a project which performs well on the POEFS dimensions, is canceled after a more elaborate safety risk analysis later on. Therefore, in situations where a high-level assessment is concerned, it may be reasonable to consider the relative importance of all SPOEFS dimensions up front. This will help the airline to improve their strategic assessment based on the integrated set of objectives.

Due to the amount of discussion centered upon the integration of the safety dimension, definitive answers cannot be provided yet. However, it is possible to take a step back and draw some practical conclusions. On the one hand, the responses from certain airline decision-makers indicate that at the highest strategic level, everything is a trade off, i.e., a non-compensatory framework in regards to safety is not necessary. Indeed, their views indicate that at a more fundamental level – especially when pursuing a more innovative and sustainable air transport system – all factors, including safety are negotiable. After all, the safest way of flying is not flying at all.

On the other hand, some managers – predominantly those from the Flight Operations department who are closer to the sharp-end of safety risk management – take a more detailed view. Their comments imply that the potential of compensation between the scores on other attributes and safety performance, would reduce the relevance of the results. In contemplating how to best integrate the innovations into the firms operations, they push for a strategy that asks decision-makers to consider the safety aspect first.

It can be argued however, that these views are not necessarily mutually exclusive. A possible two-stream approach is illustrated in Figure 4.16 (p.67). The central idea in this model is that airline managers decide on a case-by-case basis which of the two streams is more appropriate. A fully compensatory SPOEFS approach will generally be more efficient and allows for a more strategic perspective on general high-level initiatives, for example a certain carbon reduction pathway.

For highly specific project proposals, the right side of the framework (Figure 4.16) could be applied. Here, the alternatives are first screened using either the full or shortened version of the formal risk assessment, before they are tested against the POEFS criteria. This type of analysis would require more time commitment. However, it will be more relevant compared to the first stream, when certain detailed attributes of a concrete project proposal present safety-sensitive issues. For example, based on safety data or intuition from accountable managers, lowering the landing gear at a certain specified distance from the runway could pose a safety risk that is unacceptable, regardless how much it decreases the airline's carbon intensity. In a high-level assessment, this may only become clear after the mandatory risk assessment is performed later on, which of course negates the efficiency argument that is usually connected to the first stream.

It should be noted that the current framework to support the integration of the safety dimension in the innovation decision-making process (Figure 4.16) is just a first-cut sketch. Its usefulness when fully developed will have to be tested in future case studies.

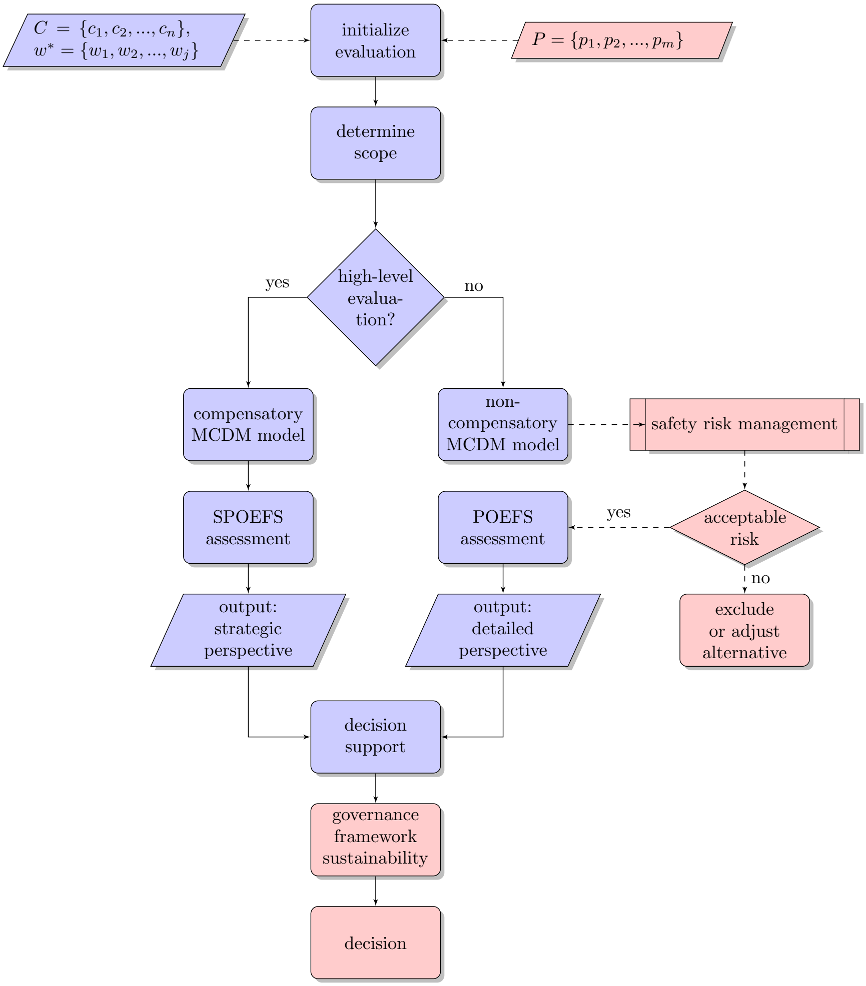

Figure 4.16: Possible supportive framework for the integration of the safety dimension, showing the steps that have been detailed in the MCDM framework of the current case study (blue) and elements that will have to be provided internally by the firm or synthesized from future research (red), where P represents a set of potential innovation projects.

5

Conclusion and Recommendations

This is a concluding chapter, which explains the scientific and managerial implications of the research findings. It also discusses the limitations of the study and suggests several interesting avenues for future research.

5.1. Theoretical Implications

The European aviation sector agrees that a mix of multiple smaller operational innovations collectively have great potential to help shortening the pathway towards net zero aviation (SEO Economics and Netherlands Aerospace Centre, 2021). Nevertheless, from the operator's point of view, hardly any insights have been provided on how the sustainability perspective aligns with other industry-specific corporate objectives with respect to the selection of innovation projects. In order to improve the effectiveness of sustainability-oriented decision-making in airlines, this study aimed to address this gap. Specifically, the goal was to view the evaluation of operational innovations as a multi-criteria decision-making (MCDM) problem. To that end, a case study was conducted at the low-cost arm of Europe's third largest airline group.

From discussions with seven company experts and relevant literature from the academic and practitioner communities, the definitions of the main criteria to support the evaluation of operational improvements in flight operations were synthesized. Using the Best-Worst Method (BWM) developed by Rezaei (2015), inter-criteria preference data from seventeen senior and executive airline managers were collected. The pairwise comparison vectors were analyzed through the aggregation of individual preferences as well as probabilistic modeling.

Results from both approaches showed that "Safety" plays by far the most important role in the evaluation of operational innovations. It is followed by the four criteria, "Employee Experience", "Passenger Experience", "Finance", and "Sustainability", between which there seems to be no evident relative preference based on the aggregation of individual results. From the probabilistic approach could be concluded with relative certainty that Finance ≻ Sustainability ≻ Employee Experience ≻ Passenger Experience. This order showed conformity to an overall intuitive ranking provided by a separate group of decision-makers.

Still, neither the aggregation of individual data, nor the probabilistic model revealed practically meaningful differences between the magnitudes of the weights assigned to the four intermediate criteria. Both approaches established however, that the criterion related to on-time performance (OTP) is clearly the least important.

The results presented in this study can be interpreted three ways:

1. As a high-level framework to assess the priority of sustainability relative to the five other dimensions vital to the firm's operations. In this interpretation, the results can be used by airline managers to develop an effective corporate strategy for the appraisal of operational innovations in flight operations;

2. As the conceptual foundation for the quantitative comparison of a set of projects with respect to the definitions of the six perspectives presented in this report. In its most basic form, the process innovations could be rated on an n-point scale, and normalized using simple linear-max normalization. Subsequently, the final scores could be obtained using the weighted sum method, based on the weights determined in this study. However, a full set of requirements for implementing a sustainability-oriented decision support system could not be provided due to limitations that will be discussed in Section 5.3;

3. Separately, knowing the relative preference from key decision-makers can help other practitioners to improve the performance of their innovation proposals with respect to the six dimensions. Concretely, if an internal or external party wants to introduce a new sustainability-oriented operational innovation to airline management, it should also explicitly highlight the alternative's attribute performance from the safety, employee, and passenger perspectives.

The fact that the Sustainability dimension takes second place to safety in the appraisal of eco-focused innovation projects, may seem strange to some outsiders. However, given the legal responsibilities of some of the managers involved, and ongoing campaigns that promote the firm's safety policy on a daily basis, the results are in fact not unexpected. What is more interesting, is the question of how the firm will balance these two dimensions going forward. As discussed, the corporate interests from the sustainability and safety perspectives are sometimes diametrically opposed. It is expected that through increased global focus on sustainability, the dichotomy will only become more relevant in the coming years. To this end, the supportive framework provided in this report gives the opportunity for future case studies to connect the safety policy of the firm with sustainability-driven initiatives in a more detailed way.

Surprisingly, the findings from the judgment provided by airline leadership also reveal that the relative preference for passenger and employee dimensions is almost equal to the importance placed on finance and sustainability. Several of the consulted decision-makers argue that through perceived correlations, all other dimensions will eventually benefit from increased satisfaction among passengers and crew. The long-term relationship between satisfaction and profitability is partly supported by both the scientific literature and company data. Eventually, passengers and crew might demand sustainable business practices as well. But it is doubtful whether relentless focus on customer and employee also assures that enough attention is paid to sustainability in the short to medium-term. This could be an interesting direction for future research.

A final theoretical contribution from this study concerns the general application of an MCDM model for the evaluation of sustainability-driven projects in aviation, and the use of the Best-Worst Method by Rezaei (2015) in particular. It seems that the current approach provides a reliable way of defining and prioritizing the criteria related to the selection of operational innovations. However, in future research the incorporation of other MCDM methods might be needed to further develop the remaining steps of the general MCDM framework, most notably the steps related to assigning and normalizing performance scores for concrete projects, as well as aggregating the results.

5.2. Managerial Implications

Due to the practical nature of this case study, the theoretical implications are highly relevant to practitioners as well. Furthermore, in the interactive Dot Plot, managers can locate their individual weights and compare them to the central tendency or the data points from other individuals. The credal ranking on the other hand, provides guidance on the extent to which the airline leadership group prefers one criterion over another. Finally, one additional recommendation to the practitioners emerged from the results of this study:

The operator is recommended to reflect on the importance placed on sustainability – which seems not to be particularly prominent among the other corporate objectives that vie for attention – and decide whether priorities need to be reevaluated.

This should be done both in light of the recent report by the Intergovernmental Panel on Climate Change (2021), as well as from a pure economic perspective. As referenced by some of the decision-makers, the sustainability dimension correlates with finance. However, this effect is currently moderate due to the relatively low CO2 price under EU ETS. Therefore, the assessment of whether the current set of priorities are in line with mid and long-term ambitions should include, but not necessarily be limited to the following considerations:

1. the future development of the CO2 price;
2. other enforcement mechanisms such as a carbon cap with no possibilities of trading emission rights;
3. the potential impact on future profitability of violating climate norms, such as losing valuable slots on capacity restricted airports.

For Transavia these considerations are particularly important, given the airline's capacity expansion in the recent decades. During the firms now completed transformation from a more ad-hoc charter focused operation towards a full-fledged scheduled low-cost-carrier model, it has been steadily adding seat-kilometers, and as a consequence also increased its carbon-intensity. At the same time, one should keep in mind that relative economies of scale in comparison to the direct competition, in terms of asset-intensity and purchasing power, is – and always has been – the foremost source of strategic advantage in the airline industry. If we combine these two facts with the current relative importance of sustainability, it is not implausible that a CO2-ceiling with a reference year that lies before the period of accelerated capacity growth, could potentially pose a serious threat to long-term profitability. In that regard, ongoing campaigns to promote the firm's sustainability policy in a similar way as is currently done for safety may be required.

5.3. Limitations and Future Research

The scientific contribution that was addressed in this study relates to the gap in the literature identified in 2.4. While the single case study design allows to gain an in-depth understanding of the phenomenon under investigation, it usually provides limited grounds for using the knowledge outside the specific context. However, due to the relative uniformity in the air-transport industry in terms of product, operations, corporate structure, objectives, and challenges; results of this paper may be generalized to other operators. In more general terms, findings concerning the prioritization of corporate sustainable practices related to the multiple and conflicting dimensions of the firm, may apply to the broader transportation sector. Nevertheless, there are some clear limitations to this study that suggest the need for further research. They are related to: i) the sample, ii) the criteria, and iii) the decision problem.

First, the sample size from both the group of experts (N=7) as the group of decision-makers (N=17) was limited, in order to follow as closely as possible the data collection and governance structure of the firm. Moreover, the research was conducted in the form of a case study at a low-cost operator. It is therefore interesting to see for instance, what the weight placed on "Operations" would be at a full-service network carrier, and its effect on the other relative preferences. Furthermore, the input preferences were not weighted based on the place of the decision-makers in corporate hierarchy or their capabilities in a specific problem area. Therefore, another potential line of research would be to add the dimension of *"believeability-weighted"* decision-making to the MCDM analysis.

Second, the decision-criteria mentioned in this study were defined with the goal of weighting them at the strategic level. The definitions could of course be expanded or developed into sub-criteria. New criteria could also be added based on the knowledge from other experts. After all, the definitions established in this study are merely constructs that make it easier to study the preference relations, but do not necessarily represent what is right in all circumstances. The second group of decision-makers, who were asked to intuitively verify the overall ranking, mentioned for example that the definition of the Employee dimension should be extended to include cabin crew and office personnel. Criteria could also change over time. It is particularly interesting to see how the Sustainability dimension will change based on new scientific data on both CO2 and non-CO2 emissions. As long as traditional jet fuel is used at a large scale, relative carbon intensity will remain a good proxy for sustainability. Nevertheless, other aspects may have to be added in the future, when the path towards a climate neutral air transport system becomes clearer.

Third, in the problem hierarchy, some elements of the criteria are supported by mathematical expressions. However, with the current definitions there are some technical issues to take into account when interpreting the model mathematically: (i) if the share of SAF in the fuel mix increases, the climate benefits that can be achieved through operational improvements will be lower, while of course the absolute climate intensity decreases. In the final performance scores, this may or may not be compensated by the fact that SAF is more expensive (and thus operational innovations remain relevant in the finance dimension); (ii) the relative carbon footprint may be very sensitive to the average COG position of the aircraft, which is something that this study has not covered in detail.

Finally, before the results from this research can be translated into a full quantitative evaluation, a more detailed, and systematic examination is needed on the role of the safety perspective, the amount of compensation allowed between criteria, and whether or not the assumption of mutual preference independence is valid. Although these aspects were not neglected in the present study, more data is required in order to draw certain conclusions for the development of a crisp MCDM model.

In regard to the integration of the Safety dimension, a clear direction for future research was provided in Chapter 4. However, for the comments on the initial exploration of correlation, compensation, and preference (in)dependence in Appendix A, one should keep in mind that this is an initial mapping based on qualitative assessment. The inference of practically useful conclusions for the decision support system – beyond a first-cut approximation – requires further quantitative analysis. As discussed, for a clean application of the additive value function, one should ascertain that additive independence represents reasonably well the airline managers' utilities, i.e., for assessing the utility with respect to one criterion, it should not matter (too much) what the performance on other criteria is.

In future case studies this may be provided for instance through the application of a discrete choice experiment. In these experiments airline managers could be asked to make trade-offs between combinations of different concrete innovation project attributes. The results from presenting the multiple variants of the same project in different scenarios could provide insight into the extent to which the utilities related to certain attributes depend on each other. Subsequently, the information obtained on the decision-problem characteristics could be used by researchers to create a more detailed design of the decision-support system for selecting operational innovations in airlines.

Bibliography

Annema, J. A., Mouter, N., & Razaei, J. (2015). Cost-benefit analysis (cba), or multi-criteria decision-making (mcdm) or both: Politicians' perspective in transport policy appraisal. *Transportation Research Procedia*, *10*, 788–797.

Behzadian, M., Kazemzadeh, R., Albadvi, A., & M. (2010). Promethee: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, *200*(1), 198–215.

Belton, V., & Pictet, J. (1997). A framework for group decision using a mcda model: Sharing, aggregating or comparing individual information? *Journal of Decision Systems*, *6*(3), 283–303.

Bernhardt, K. L., Donthu, N., & Kennett, P. A. (2000). A longitudinal analysis of satisfaction and profitability. *Journal of Business Research*, *47*(2), 161–171.

Boeing. (2021). My boeing fleet: Fuel dashboard.

Bowman, E. H., & Hurry, D. (1993). Strategy through the option lens: An integrated view of resource investments and the incremental-choice process. *The Academy of Management Review*, *18*(4), 760–782.

Bruno, G., Esposito, E., & Genovese, A. (2015). A model for aircraft evaluation to support strategic decisions. *Expert Systems with Applications*, *42*(13), 5580–5590.

Chakraborty, S., & Yeh, C. (2007). A simulation based comparative study of normalization procedures in multiattribute decision making. *Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases*, 102–109.

Chan, T., Nickerson, J., & Owan, H. (2007). Strategic management of r&d pipelines with cospecialized investments and technology markets. *Management Science*, *53*(4), 667–682.

De Decker, G. (2021). Graduation agreement appendix i: Msc thesis research proposal. *Delft University of Technology*.

De Decker, G., Poelman, A., & van Veen, C. (2020). The growth of schiphol airport. *Delft University of Technology*.

Delft University of Technology, & Netherlands Aerospace Centre. (2021). Towards a sustainable air transport system. *Whitepaper*.

Dožić, S. (2019). Multi-criteria decision making methods: Application in the aviation industry. *Journal of Air Transport Management*, *79*.

Eisenführ, F., Weber, M., & Langer, T. (2010). *Rational decision making*. Springer-Verlag Berlin Heidelberg.

European Aviation Safety Agency. (2016). Acceptable means of compliance (amc) and guidance material (gm) to annex iii – part-oro.

European Parliament and Council. (2012). Commission regulation (eu) no 965/2012 of 5 october 2012: Laying down technical requirements and administrative procedures related to air operations pursuant to regulation (ec) no 216/2008. *Official Journal of the European Union*.

European Parliament and Council. (2018). Regulation (eu) 2018/1999 on the governance of the energy union and climate action. *Official Journal of the European Union*.

Figueira, J., Mousseau, V., & Roy, B. (2016). *Multiple criteria decision analysis: State of the art surveys*. Kluwer Academic Publishers.

Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). *Statistical distributions*. John Wiley & Sons.

Forman, E., & Selly, M. (2001). *Decision by objectives: How to convince others that you are right*. World Scientific Publishing Company Incorporated.

Goodwin, P., & Wright, G. (2004). *Decision analysis for management judgment*. John Wiley & Sons.

Görener, A. (2012). Comparing ahp and anp: An application of strategic decisions making in a manufacturing company. *International Journal of Business and Social Science*, *3*(11), 194–208.

Guitouni, A., & Martel, J.-M. (1998). Tentative guidelines to help choosing an appropriate mcda method. *European Journal of Operational Research*, *109*(2), 501–521.

Herman, M. W., & Koczkodaj, W. W. (1996). A monte carlo study of pairwise comparison. *Information Processing Letters*, *57*(1), 25–29.

Hubbard, D. W. (2014). *How to measure anything*. John Wiley & Sons.

Intergovernmental Panel on Climate Change. (2021). Climate change 2021 the physical science basis. *Journal of Business Research*.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620–630.

Kahneman, D. (2011). *Thinking, fast and slow*. Penguin Random House.

Kano, N. (1984). Attractive quality and must be quality. *Journal of the Japanese Society for Quality Control*, *14*(2), 147–156.

Karapetrovic, S., & Rosenbloom, E. (1999). A quality control approach to consistency paradoxes in ahp. *European Journal of Operational Research*, *119*(3), 704–718.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value trade-offs*. John Wiley & Sons.

Klitsie, J. B., Price, R. A., & de Lille, C. (2018). Overcoming the valley of death: A design innovation perspective. *Overcoming the Valley of Death: A Design Innovation Perspective. In Next Wave: The 21st dmi: Academic Design Management Conference Proceedings*, 958–972.

Kornyshova, E., & Salinesi, C. (2007). Mcdm techniques selection approaches: State of the art. *2007 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making*, 22–29.

Lee, D., Fahey, D., Skowron, A., Allen, M., Burkhardt, U., Chen, Q., Doherty, S., Freeman, S., Forster, P., Fuglestvedt, J., Gettelman, A., De Len, R., Lim, L., Lund, M., Millar, R., Owen, B., Penner, J., Pitari, G., Prather, M., ... Wilcox, L. (2021). The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. *Atmospheric Environment*, *244*, 117834.

Li, L., Wang, X., & Rezaei, J. (2020). A bayesian best-worst method-based multicriteria competence analysis of crowdsourcing delivery personnel. *Complexity*, *2020*, 1–17.

Ma, J., Harstvedt, J. D., Jaradat, R., & Smith, B. (2020). Sustainability driven multi-criteria project portfolio selection under uncertain decision-making environment. *Computers Industrial Engineering*, *140*, 106236.

Marttunen, M., Lienert, J., & Belton, V. (2017). Structuring problems for multi-criteria decision analysis in practice: A literature review of method combinations. *European Journal of Operational Research*, *263*(1), 1–17.

Matzler, K., Bailom, F., Hinterhuber, H. H., Renzl, B., & Pichler, J. (2004). The asymmetric relationship between attribute-level performance and overall customer satisfaction: A reconsideration of the importance-performance analysis. *Industrial Marketing Management*, *33*(4), 271–277.

McKinsey & Company. (2020). Hydrogen-powered aviation: A fact-based study of hydrogen technology, economics, and climate impact by 2050.

Ministry of Infrastructure and Water Management. (2019). Verantwoord vliegen naar 2050: Luchtvaartnota 2020-2050.

Mohammadi, M., & Rezaei, J. (2020). Bayesian best-worst method: A probabilistic group decision making model. *Omega*, *96*, 102075.

Opricovic, S., & Tzeng, G.-H. (2004). Compromise solution by mcdm methods: A comparative analysis of vikor and topsis. *European Journal of Operational Research*, *156*(2), 445–455.

Ou Yang, Y.-P., Leu, J.-D., & Tzeng, G.-H. (2008). A novel hybrid mcdm model combined with dematel and anp with applications. *International Journal of Operational Research*, *5*(3).

Perlitz, M., Peske, T., & Schrank, R. (2002). Real options valuation: The new frontier in rd project evaluation? *RD Management*, *29*(3), 255–270.

Plummer, M. (2004). Jags: Just another gibbs sampler. *American Educational Research Association*.

Proctor, R. W., & Zandt, T. V. (2018). *Human factors in simple and complex systems*. Taylor & Francis.

Raiffa, H. (1982). *The art and science of negotiation*. Harvard University Press.

Reason, J. (1997). *Managing the risks of organizational accidents*. Taylor; Francis.

Rezaei, J. (2016a). Bwm solver. *bestworstmethod.com*.

Rezaei, J., & Kadziński, M. (2018). Editorial: Special issue: Multiple criteria decision making in air transport management. *Journal of Air Transport Management*, *68*.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49–57.

Rezaei, J. (2016b). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, *64*, 126–130.

Rezaei, J. (2020). A concentration ratio for nonlinear best worst method. *International Journal of Information Technology Decision Making*, *19*(3), 891–907.

Rice, C., Ragbir, N. K., Rice, S., & Barcia, G. (2020). Willingness to pay for sustainable aviation depends on ticket price, greenhouse gas reductions and gender. *Technology in Society*, *60*, 101224.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169.

Saaty, R. (1987). The analytic hierarchy process, what it is and how it is used. *Mathematical Modeling*, *9*(3), 161–176.

Saaty, T. (1996). *Decision making with dependence and feedback: The analytic network process*. RWS publications.

Schaefer, M., & Bartosch, S. (2013). Overview on fuel flow correlation methods for the calculation of nox, co and hc emissions and their implementation into aircraft performance software. *DLR Institut für Antriebstechnik*.

Schilling, M. A. (2017). *Strategic management of technological innovation* (Fifth). McGraw-Hill Education.

Schilling, M. A. (1998). Technological lockout: An integrative model of the economic and strategic factors driving technology success and failure. *The Academy of Management Review*, *23*(2), 267–284.

SEO Economics, & Netherlands Aerospace Centre. (2021). Destination 2050 – a route to net zero european aviation.

Steffens, P., & Douglas, E. (2007). Valuing technology investments: Use real options thinking but forget real options valuation. *International Journal of Technoentrepreneurship*, *1*(1), 58–77.

Stewart, T. (1992). A critical survey on the status of multiple criteria decision making theory and practice. *Omega*, *20*(5), 569–586.

Thurstone, L. (1927). A law of comparative judgement. *Psychological Review*, *34*, 273–286.

Transavia Airlines. (2021). Vlieger attitude bij nieuwe projecten. *Internal documentation*.

Transport Canada. (2017). *Human factors for aviation - basic handbook*. Government of Canada.

Triantaphyllou, E., & Mann, S. H. (1989). An examination of the effectiveness of multi-dimensional decision-making methods: A decision-making paradox. *Decision Support Systems*, *5*(3), 303–312.

Tsafarakis, S., Kokotas, T., & Pantouvakis, A. (2018). A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement. *Journal of Air Transport Management*, *68*, 61–75.

Tsai, W.-H., Chou, W.-C., & Hsu, W. (2009). The sustainability balanced scorecard as a framework for selecting socially responsible investment: An effective mcdm model. *Journal of the Operational Research Society*, *60*(10), 1396–1410.

Tsao, C.-T. (2012). Fuzzy net present values for capital investments in an uncertain environment [Special Issue: Advances of Operations Research in Service Industry]. *Computers Operations Research*, *39*(8), 1885–1892.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Uitbeijerse, G. (2020). Co2-emissie van de luchtvaart op de lange termijn. *Planbureau voor de Leefomgeving*.

UK Civil Aviation Authority. (2008). Safety management systems – guidance to organisations. *Safety Regulation Group*.

Vafaei, N., Ribeiro, R. A., & Camarinha-Matos, L. M. (2016). Normalization techniques for multi-criteria decision making: Analytical hierarchy process case study. *Technological Innovation for Cyber-Physical Systems*, 261–269.

Vainio, T. (2019). Additive value functions. *Seminar on Operations Research*.

van de Weijer, C. (2019). De vliegschaamte voorbij. *Financieele Dagblad*.

Velasquez, M., & Hester, P. (2013). An analysis of multi-criteria decision making methods. *International Journal of Operations Research, 10*(2), 56–66.

Vincke, P. (1992). *Multi-criteria decision-aid*. John Wiley & Sons.

Yin, R. K. (2014). *Case study research: Design and methods (applied social research methods)*. Sage publications.

A

Correlation, Compensation, and Preference (In)dependence

In addition to the rating on the 9-point scale, the seventeen DMs were asked about their opinion with respect to the decision-making characteristics, i.e., if they perceived that correlation, compensation, and preference (in)dependence exist between the main criteria. Subsequently, the decision-makers were also requested to reflect on whether the Safety dimension should be included in the framework. Table A.1 summarizes the view of the Transavia managers with respect to the additional considerations. The zeros and ones indicate if the DMs believe correlation, compensation, and preference dependence between the criteria exist and whether the safety perspective should be integrated or not.

The discussion on the Safety dimension was presented in the main part of this report (Chapter 4). Some comments on the initial exploration of correlation, compensation, and preference (in)dependence are provided below. Keep in mind that this is an initial mapping based on qualitative assessment. The inference of practically useful conclusions for the decision support system – beyond a first-cut approximation – requires further quantitative analysis.

Table A.1: The view of 17 senior airline managers on the additional considerations for the decision-problem.

	Role	Domain	Corr.	Comp.	Dep.	Safety
DM 1	CFO	Finance	1	1	1	0
DM 2	Director of Innovation	MT FO	1	0	0	0
DM 3	Manager Airline Pilot Program	MT FO	1	1	1	1

Table A.1 continued from previous page

	Role	Domain	Corr.	Comp.	Dep.	Safety
DM 4	Head of Training	MT FO	1	1	1	0
DM 5	Chief Pilot	MT FO	1	1	0	0
DM 6	Director Flight Technical	MT FO	1	1	0	1
DM 7	Director Crew Planning 1	MT FO	1	1	0	1
DM 8	Director Crew Planning 2	MT FO	1	1	0	0
DM 9	FO HR Business Partner	MT FO	1	1	0	1
DM 10	COO	SRB	1	1	1	1
DM 11	VP Flight Operations	SRB	1	1	1	1
DM 12	VP Cabin Operations	SRB	1	1	1	0
DM 13	VP Technical Services	SRB	1	0	1	1
DM 14	VP Operations Control	SRB	1	1	1	1
DM 15	VP Ground Services	SRB	1	1	1	1
DM 16	VP Network	Commercial	1	1	0	0
DM 17	VP Customer Proposition	Commercial	1	1	1	1
Total			**17**	**15**	**10**	**10**

As can be seen in Table A.1, all seventeen airline managers agree that correlations exists between at least one pair of decision criteria. Several decision-makers made reference to the correlation between the finance and sustainability perspectives, which can be both positive and negative. On the one hand, operational innovations that decrease fuel burn have a positive impact on both the carbon footprint and operational expenditures. On the other hand, if the innovation involves costly physical modifications to the aircraft, the correlation between finance and sustainability might be weakened or even inverted.

Also interesting is the fact that managers who assign a high relative weight to EX argue that, through the correlations between the criteria, this will eventually lead to better results in other dimensions as well. For finance this is relatively well established in the scientific literature (e.g. Bernhardt et al. (2000)). As for safety, it is also reasonably to assume that satisfied employees are more likely to adhere to the safety norms that are specified in the instructions and procedures laid down in the operations manual. However, the link between employee experience and sustainability is less clear, which gives rise to the possibility that these managers overlook the importance of reducing the carbon footprint on the short to medium-term.

As such, correlations between the main criteria pose no implications for future iterations of the framework. They are inherently accounted for by the way most multi-criteria assessments are set up. While it might be interesting to reflect on the correlations internally, they do not lead to new insights from a decision science perspective.

This is different for compensation and mutual preference independence. Fifteen out of seventeen DMs implied that at a high level the dimensions are compensatory. This is of course on the condition that safety is defined as additional operational risk management, not compliance to basic regulations. Two managers were concerned that with the current dimensions and their respective weights, a crisp evaluation, e.g. through a weighted sum method, would be misleading.

Much discussion centered upon the degree of dependence between the criteria (keep in mind that dependence $\neq$ correlation). Seven managers recognized that on a strategic level the criteria were mutually preferentially independent (while naturally some dependencies exist on a more detailed level). They favored an approach which seeks to optimize all individual elements, regardless of the performance on other elements. Ten decision-makers (almost 60%) were less eager to accept this notion. More specifically, they asserted that on-time performance is unimportant as long as passengers and crew are satisfied, but it deserves more attention if this is not the case. While this was the only form of preference dependence that was observed, it is possible that more dependencies are unveiled when investigated in a more systematic way, i.e. through a discrete choice experiment.