

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Player performance evaluation in rugby using
Stochastic Multi-Criteria Acceptability Analysis
with simplified uncertainty formats

University of Cape Town
Department of Statistical Science

Jon CALDER

Supervisor:

Ian DURBACH



This dissertation is prepared in fulfilment of
the requirements for the degree of Master of Science
in Operations Research at the University of Cape Town.

Sunday 19th May, 2013

Acknowledgements

This dissertation would not have been possible without the involvement of a number of individuals who have each contributed in their own way.

Firstly I would like to thank my supervisor, Dr Ian Durbach, for his continued support and guidance throughout this project. You have introduced me to a challenging and fascinating area of research and I am thrilled to have been able to explore this field of study with your assistance.

To David Wessels and Rassie Erasmus who commissioned this project and provided financial support during my first year of research - thank you for your passion for rugby, your commitment to learning, and ultimately for creating this fantastic opportunity for me.

To the staff of the Statistics Department at UCT, and to all my teachers and lecturers over the years - thank you all for your investment in me and for your service to education in this country.

Lastly, to my friends, family and any others who have supported and encouraged me throughout this process - whether by your words, deeds, or finances - my sincerest thanks for helping me see this through to completion.

It is only by God's grace - to Him be all the glory.

Plagiarism Declaration

I know the meaning of plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Signature: _____

Abstract

This dissertation considers problem contexts in which decision makers are unable or unwilling to assess trade-off information precisely. The primary aim is to investigate the use of Stochastic Multi-criteria Acceptability Analysis (SMAA) with simplified representations of uncertainty as a decision support tool for prescriptive modelling in ‘low involvement’ contexts such as these.

A simulation experiment is used to assess (a) how closely a rank order of alternatives based on partial information and SMAA can approximate results obtained using full-information multi-attribute utility theory (MAUT), (b) whether a number of ‘simplified’ SMAA models which make use of summarised measures of uncertainty instead of a full probability distribution might also be suitable in certain contexts, and (c) which characteristics of the decision problem influence the accuracy of this approximation.

Within the range of simulated cases, the main findings are the following:

1. Fairly good accuracy can be achieved by utilizing only limited preference information.
2. Avoiding assessment errors in the application of a decision model is far more crucial than the choice of a particular type of model.
3. Some of the simplified SMAA models can in fact provide a suitable degree of accuracy provided that the two aforementioned issues are addressed i.e. a small amount of preference information should be elicited in order to limit the explored preference space and assessment errors should be carefully avoided.

4. Over a range of decision problems, the best performing model uses a small number of quantiles to represent uncertainty in the attribute evaluations. The worst performance is consistently produced by a ‘risk’ model which utilises both the mean and variance of the criteria attributes.

A practical application of SMAA models with simplified uncertainty formats is then investigated by considering these models as tools for short-listing of the best performing rugby players based on position-specific game data taken from the Super 14 in 2008 and 2009. Analysis for two of the positional clusters - centres and loose forwards - is presented by looking at the resulting acceptability indices and central weight vectors. This approach shows a great deal of potential and the results provide evidence that the technique is effectively able to pick up the nature of a player’s game as well as reveal some interesting insights. The central weights in particular, allow for presentation of the results in a manner that is easy to understand and interpret, and in so doing aid in the identification of strengths and weaknesses for each player within these positional profiles.

By allowing coaches to evaluate player performance on a scale which is relative to the spread of performances by other players in a similar position, SMAA can be used to inform coaches as they make selections. In particular, this ‘low involvement’ application suggests that the technique might be of greatest value in situations where there are a lot of players to choose from and the coaching staff are not already entirely familiar with all of the players. A prime example of this would be in a trials situation preceding squad selection, when coaches may be able to use historical match data to begin to evaluate the strengths and weaknesses of some of the less well known players from smaller franchises or junior teams.

Contents

1	Introduction	1
1.1	Background	1
1.2	Objectives	5
1.3	Scope and Limitations	6
1.4	Plan Of Development	8
2	Literature Survey	9
2.1	Multiple Criteria Decision Analysis	9
2.1.1	Introduction to MCDA	9
2.1.2	Problem Structuring	12
2.1.3	Overview of MCDA Methods	15
2.1.4	Value Function Methods	19
2.1.5	Treatment of Uncertainty in MCDA	22
2.2	Stochastic Multi-Criteria Acceptability Analysis	24
2.2.1	Motivation	24
2.2.2	Methods	25
2.2.3	Applications	26
2.3	The SMAA-2 Model	29
2.3.1	Preference Information	29
2.3.2	Acceptability Analysis	31
2.3.3	Computation	35
2.4	Treatment of Uncertainty	36

2.4.1	Model using expected values	37
2.4.2	Model using expected values and variances	37
2.4.3	Model using quantiles with fixed weights	38
2.4.4	Model using quantiles with variable weights	38
2.4.5	Model using a fully fitted distribution	39
2.5	Science and Sport	41
2.6	Analysis of Rugby Union	45
2.6.1	Description of the Game of Rugby	45
2.6.2	Recent History of the Game	46
2.6.3	Analysis of Match Data	48
2.6.4	Analysis of Player Performance	48
3	Simulation	54
3.1	Objectives	54
3.2	Overview of Simulation Process	55
3.3	Generating Data for Problem Context	55
3.3.1	Problem Size	57
3.3.2	Attribute Evaluations	57
3.4	Preference Modelling	58
3.4.1	Utility Function	59
3.4.2	Inter-Criterion Weights	61
3.5	Output Measures	66
4	Simulation Results	67
4.1	Weight Range Restrictions	67
4.2	Overview of Simulation Results	71
4.3	Comparison of Different Uncertainty Models under Error	73
4.4	Changes in Problem Size and Attribute Distribution	76
4.5	Changes in Utility Function	79
4.6	k -Best Ranks Acceptability Indices	82
4.7	Discussion	85

5	Evaluation of Rugby Players	89
5.1	Discussion of Data	90
5.2	Problem Structuring	91
5.2.1	Research Objectives	91
5.2.2	Stakeholder Involvement	91
5.2.3	Grouping of Alternatives	92
5.2.4	Constructing Attributes	93
5.2.5	Data Cleaning and Transformation	97
5.3	Performance evaluation models	98
5.3.1	Treatment of uncertain attribute evaluations	98
5.3.2	Implementation details	101
5.4	Results	102
5.4.1	Centers	102
5.4.2	Loose forwards	108
5.4.3	Feedback from rugby coaches	111
6	Conclusions	112
	Bibliography	117

University of Cape Town

List of Tables

2.1	Summary of match analysis research highlighting predictors of success in rugby	47
3.1	Summary of all simulation parameters used	62
3.2	Models used in simulation	65
4.1	Factorial-ANOVA results showing first and second order effects of model type, weight range, and case on utility loss	68
4.2	Summary of ANOVA results showing the main effects of all simulation parameters	72
4.3	Table showing the direction of the average change in utility loss when each of the simulation parameters is increased	73
4.4	Factorial-ANOVA results showing first and second order interaction effects for the different simulation parameters	74
4.5	Summary statistics of utility loss by model type for different assessment errors	75
5.1	Positional clusters for players	93
5.2	Performance attributes for player positions (part 1)	94
5.3	Performance attributes for player positions (part 2)	96

List of Figures

3.1	Outline of a simulation run	56
3.2	Graphs of utility function (3.4) showing the effect of different values for τ , λ , α & β	60
4.1	Mean utility loss by model for different weight restrictions	69
4.2	Mean utility loss by case for different weight restrictions	69
4.3	Plot of mean utility loss for different models and assessment errors	77
4.4	Box-plots of utility loss for the different models and assessment errors	77
4.5	Mean utility loss broken down according to the variance and skewness of the performance attributes, with 10 and 20 criteria	78
4.6	Mean utility loss broken down according to attribute variance and skewness for the different models and assessment errors	80
4.7	Mean utility loss for different levels of τ , λ and β , at the different error levels	81
4.8	Mean utility loss broken down according to the three utility function parameters, for the different models and assessment errors	83
4.9	Mean utility loss at different error levels when using different values of k for the k -best ranks acceptability indices	84
5.1	Log transformation used to scale frequency of events for players with less than 80 min of match time	98
5.2	Value Tree for Quantile Model of Player Ability	100
5.3	Possible simulated utility functions	102

5.4 Rank acceptability indices for the centres based on the empirical and average performance models 103

5.5 Central weight vectors for some of the centres based on the empirical and average performance models 106

5.6 Rank acceptability indices for the loose forwards based on the empirical and average performance models 109

5.7 Central weight vectors for some of the loose forwards based on the empirical and average performance models 110

University of Cape Town

Chapter 1

Introduction

1.1 Background

The Oxford Dictionary defines a decision as a ‘settlement of conclusion, formal judgement; making up one’s mind’. It is clear that decisions must involve a human element, the ‘decision maker(s)’, whose judgement and perception is key to the decision itself. The human element cannot be removed as it provides much of the context for the decision problem. For this reason it is important to highlight that the goal of decision modelling is to support the human decision making process and not to replace it. It is hoped that our understanding of complex decisions can be improved by approaching them in a structured way, and by providing suitable ‘decision support’ tools.

There are three recognised types of decision theory:

- normative theory,
- descriptive theory,
- and prescriptive theory.

According to Gilboa [20], a descriptive theory aims to describe reality, but makes no value judgement and takes no stand on whether or not this ‘reality’ is good or bad.

i.e. it describes what people actually do, not what they ought to do. A normative theory describes what a rational decision maker would do, and is based on sets of axioms and assumptions which form the economic foundation for rational decision making. A prescriptive theory tries to take account of the difficulties involved in decision making, and to detail procedures aimed at aiding coherent decision making consistent with reasonable goals of the decision maker(s) [4]. The chapters which follow, focus on the area of prescriptive theory.

One of the most crucial aspects of decision making involves the way people in general deal with uncertainty. In the context of most non-trivial decision problems, it is common to identify uncertainty both internal and external to the (human) decision making process itself. External uncertainty refers to the potentially unknown consequences of a particular decision or choice, which may be determined or influenced by environmental processes, economic outcomes etc. Internal uncertainty pertains to the decision maker's strength of preference for specific consequences or combinations of consequences, and usually involves the unresolved nature of inherent trade-offs between different goals and priorities.

At the heart of decision theory lies the problem of choice under uncertainty, and though one can encounter decision problems from anywhere across the uncertainty spectrum in practice, to some extent the theory seems to have evolved around two broad 'categories' of decision problem: (1) those where there are well defined goals and decision parameters but a large amount of external uncertainty, and (2) problems where the key issues relate to internal uncertainty (e.g. multiple conflicting goals which may or may not be clearly defined). Some of the most challenging decision problems often arise from contexts which incorporate both of the above aspects of uncertainty. One such problem, which will be considered here, is the selection of rugby players to make up a squad or team.

Squad selection in rugby can be prone to a range of different biases by selectors and coaches, as the process often relies heavily on subjective perceptions of the play-

ers' abilities and skills based on observations of their performance. This subjectivity is of two forms. Firstly, there is 'perceptual' subjectivity about how a player has actually performed. For example, did a player make many tackles in a game? Were those tackles effective? Then, there is also 'preferential' subjectivity about what aspects are most important. For example, is a tight forward who secures scrum and line-out ball more valuable than one who performs better in general play?

From a pool of available players, groups are formed on the basis of positional profiles. Coaching staff must then build a squad by choosing a small number of players from each positional group, based on assessments of the players' levels of skill and ability in key aspects of the game. In order to decide which players might be considered superior to all or some of the other players, the coaches need to weigh up and compare the strengths and weaknesses of each player.

There are a few difficulties to overcome in this regard. Firstly, underlying the different performance criteria on which players can be evaluated, are physical traits and abilities which are commonly conflicting. For example, strength and size usually come at the cost of speed and agility. Furthermore, since true talent and ability can never be measured directly and must be inferred from match performances varying over time, there will always be some degree of (external) uncertainty about the true abilities of players on each of the criteria.

Many professional rugby teams collect extensive performance data, in part to provide a clear record to coaches of player performance and hence resolve perceptual subjectivity. However, the interpretation of this data, and the selection of players is in many cases still a relatively unassisted, unstructured process relying on the intuitive judgement of the coach and selectors. We wish to investigate the use of a decision support tool to assist in the evaluation of players and the resulting selection process by integrating available match data with preference information from the coaches. It is hoped that by considering the process of selecting a rugby squad in the context of multi-criteria decision analysis (MCDA) we can explore how historic match data can

be used to inform assessments of different players ability on a number of measurable criteria.

Stochastic Multi-criteria Acceptability Analysis (SMAA) offers a generalised framework which to some extent attempts to handle both of the previously mentioned aspects of uncertainty by allowing both criteria measurements (i.e. ‘decision consequences’) and preference information to be expressed as arbitrarily distributed stochastic variables [34]. As a relatively unique family of inverse-preference models, SMAA seeks to provide information about the volume and types of preferences (if any) that would lead to the selection of each player. That is, instead of seeking out the best player given a particular set of preferences, one sets out to explore how many different preferences result in a particular player being considered the best. More specifically, what typical set of preferences might make this player the preferred one?

Since the approach can be implemented with little or no preference information, a major advantage is that it is especially well suited to provide limited decision support in ‘low involvement’ contexts where decision makers are unable or unwilling to explicitly or precisely state their preferences. In many situations this can be due to availability and/or time constraints which limit direct interaction with the decision makers themselves. Another reason – and such is the case here – can be that a group of decision makers or stakeholders are involved and there isn’t consensus on the relative importance of different criteria. Despite the flexibility of this approach however, in practice it can still be a difficult and time consuming process to apply since the conventional SMAA model requires that one fully assess the distribution of criteria measurements for all of the different alternatives present in a given decision. For this reason, special consideration will be given to the application and suitability of some simplified SMAA models which make use of summarised formats for representing uncertainty in the attribute evaluations. The key advantages are ease of use and greater transparency, but this comes at the cost of some accuracy.

By outlining the theoretical underpinnings of the SMAA methodology and exploring

the suitability of some simplified SMAA models by comparison to the conventional ‘distributional’ SMAA model, motivation for their use in low involvement decision contexts will be given. In particular, it is hoped that their usefulness can be illustrated both for the short-listing of the best performing rugby players, and for identifying the trade-offs that are involved in selecting one player over another.

1.2 Objectives

The primary aim of this project is to investigate the use of Stochastic Multi-criteria Acceptability Analysis (SMAA) with simplified representations of uncertainty as a decision support tool for prescriptive modelling in low information contexts. The methodology is first explored by means of a simulation study which will allow for the comparison of different simplified approaches to the summary of uncertainty. Then a specific application to the short-listing of best performers within positional profiles in rugby is considered.

The main intention with the simulation experiment is to evaluate the ability of SMAA to approximate results obtained using multi-attribute utility theory (MAUT). In particular, we ask how closely a rank ordering of alternatives using a key output from SMAA (the acceptability index) can approximate a rank ordering of alternatives using MAUT. In doing so we hope to provide a broad indication of the losses that are possible if facilitators choose to use a low-involvement decision aid such as SMAA rather than compelling decision makers to be more precise in their assessment of certain types of preference information – using, for example, more comprehensive problem structuring. In the process, we also wish to test the robustness of the SMAA approach to various aspects of the decision process: the size of the decision problem, the way attribute evaluations are distributed, the underlying preference functions, the accuracy of assessed information, and the way in which the acceptability index is constructed.

In addition to the conventional SMAA model, which uses probability distributions

to represent uncertainty in the attribute evaluations, we also introduce and evaluate a number of ‘simplified’ SMAA models which make use of summarised measures of uncertainty instead of a full probability distribution. By assessing the accuracy of both conventional and simplified SMAA models under a range of different conditions we hope to provide motivation for the use of these simplified decision models in appropriate circumstances. A similar approach has recently been used in [13] to assess the effect of using simplified uncertainty formats in general decision-making, and we employ a very similar simulation structure here. Indeed, one aim of the current paper is to assess whether the results obtained in [13] also apply to SMAA models.

The second broad objective is to illustrate the use of SMAA in a practical sense by applying it to the problem of player selection in rugby. It is entirely feasible that squad selection in rugby can be prone to a range of different biases by selectors and coaches, as the process relies heavily on subjective perceptions of the players’ abilities and skills based on the (subjectively) observed performance of the players. This constitutes the primary motivation for investigating the use of SMAA as a decision support tool for squad selection in rugby.

1.3 Scope and Limitations

The SMAA methodology has been devised as a decision aid, which helps deal with uncertainty in the context of decision making, and cannot in any way replace the proper decision making process. Given that it can be utilised with very little preference information to assist in the description of the problem at hand, the model output should be accepted only as a starting point for further and more thorough analysis of the problem. In most practical cases it is best used in an iterative and interactive manner in which the model is updated as the problem and underlying preferences become better understood.

In using a simulation experiment, we acknowledge that we can only evaluate the extent to which using SMAA rather than a full MAUT approach might impact re-

sults. We cannot evaluate critical issues like whether the reduced time spent on problem structuring in SMAA is “worth” the reduction in decision quality, or the degree to which the problem structuring process, through the insight it generates for the decision maker, is useful as an end in itself. The results of the simulation do not allow us to provide general conclusions on the viability of different methods, but they do provide insight for such discussion by identifying the potential trade-offs in accuracy that are implied when using a simplified model. Ultimately decision makers’ requirements for accuracy need to be weighed up against these other factors in a more practical context to determine which decision models may be most appropriate for a given problem.

In practice, a crucial aspect of any decision modelling is the way that criteria are identified, selected and evaluated. In our current application of the SMAA model to assess individual player performance in rugby, this problem is of particular importance. Since rugby is a team game, all obvious output (or criteria) measures in the data apply to the team as a whole in most cases and often cannot be easily or meaningfully ascribed to particular individuals. Instead one has to focus on individual player metrics which are not necessarily proven and direct performance criteria measurements but rather proxy measures for player ability in different areas of the game. If the performance criteria that are used aren’t reliable indicators of a player’s ability in each of the areas under consideration, then no model or methodology (SMAA included) will be able to accurately assess which players might be better than others.

Though some aspects of this problem will be discussed in the review of literature which follows in the next chapter, we do acknowledge here that there has been a limited history of (documented) academic involvement in the analysis of rugby data, especially in a predictive context. This presents a clear opportunity for further work to support the selection of relevant and meaningful performance criteria.

1.4 Plan Of Development

An outline of the chapters that follow:

Chapter 2 contains a review of two broad areas of literature pertinent to this research. Firstly, that of decision modelling, where the primary focus is an overview of the field of Multiple Criteria Decision Analysis (MCDA), and in particular of Stochastic Multi-Criteria Acceptability Analysis (SMAA). The SMAA-2 method which is utilised in Chapter 3 and Chapter 5 will also be outlined in detail. Secondly, we briefly outline of the application of statistical disciplines in sport, provide an overview of performance analysis in rugby union, and offer some contextual motivation for the application of SMAA to the problem of player selection in rugby.

Chapter 3 describes a simulation study which aims to compare the performance of different simplified approaches to summarizing uncertainty, using the SMAA-2 methodology within the context of multi-attribute utility theory to short-list from a larger set of alternatives. This is tested under a broad range of different conditions, by varying simulation parameters which control the hypothetical problem size and context, assessment error and decision maker preferences.

Chapter 4 details the results of the simulation study outlined in the previous chapter.

Chapter 5 is a case study which examines a ‘real world’ application of the SMAA methodology. Using the same models outlined in the simulation study, SMAA is utilised to analyse the performance of rugby players using data from the Super14 rugby competition in 2008 and 2009.

Chapter 6 draws together some findings and conclusions based on the results and analysis contained in the previous chapters and presents these as practical recommendations for decision support in the context of uncertain attribute evaluations.

Chapter 2

Literature Survey

2.1 Multiple Criteria Decision Analysis

2.1.1 Introduction to MCDA

Both individuals and organizations regularly face situations in which there are complex decisions to be made. These decisions may involve choosing between multiple well-defined courses of action, or they could start out as loosely defined problems which need to be investigated further to identify and compare potential courses of action [4]. In most complex decision problems there are two main potential sources of complexity which may also be present simultaneously.

Firstly, there may be multiple conflicting values and objectives which need to be held in tension in order to arrive at a sensible and acceptable (compromise) decision. These different objectives could be well defined or they may be vague. Secondly, values and preferences associated with the objectives may need to be balanced with incorporated assessments of uncertainty regarding the outcome of particular choices or actions [33].

Clearly, the extent of these complexities will distinguish serious decision problems which require thorough and careful analysis, from the more trivial decision problems

in which the ‘better’ (or even ‘best’) actions are easy to discern by means of simple investigation. For example, consider a business looking to choose a delivery vehicle for short term lease. For arguments sake assume that a list of vehicles available for lease is easily narrowed down to those which satisfy certain basic requirements of the courier business. If their objective is clearly and solely defined as the minimization of cost, and the vehicles costs are explicitly known (i.e. no uncertainty), then their ‘best’ choice is likely to be quite simple to identify.

By contrast, another courier business aiming to piece together a large fleet of different delivery vehicles over a long term period faces a far more complex decision. In addition to cost, they may want to consider other conflicting criteria such as carrying capacity and space, vehicle service plans, safety and security, and likely trade-in values for the different vehicles under consideration. They may also have to deal with uncertainty about the availability, cost, and different vehicle specifications of future models which could better suit their needs. In this case, different strategies must be considered which will involve complex trade-offs, and it is quite likely that no ‘best’ strategy can be identified. However, careful analysis may at least allow certain ‘bad’ strategies to be eliminated, which adds value to the decision making process.

Belton and Stewart [4] loosely specify broad areas of decision making based on the level of precision with which the objectives can be specified and the degree of uncertainty present. When the goals and parameters of a decision are well defined, and there is limited or no uncertainty (like the vehicle leasing example above), the decision is usually referred to as programmable. Well defined decision problems with high levels of uncertainty are generally the subject of statistical decision theory, where the focus is more on dealing with the uncertainty than with modelling the multi-faceted preferences of the decision maker(s). Conventional Multi-Criteria Decision Analysis or Aiding (MCDA), usually deals with decision problems where there are low levels of uncertainty, but the objectives and parameters of the decision are more vaguely specified. Highly complex decisions which involve both high levels of uncertainty and fuzzily defined decision parameters usually require the application of a variety

of decision support methods.

The general focus for MCDA is thus on decision making problems characterised by being composed of a set of alternatives which can each be evaluated on the basis of several measurable criteria. By investigating decision maker preferences on the different criteria, and based on different levels of importance assigned to these criteria, the alternatives can be evaluated and usually compared with each other to some extent using observed assessments on each of the criteria.

There are a number of categories of problem for which MCDA can be useful:

- The choice problem - make a choice from a group of alternatives.
- The sorting problem - sort actions into categories, for example: 'acceptable', 'possibly acceptable' and 'unacceptable'.
- The ranking problem - place actions into some form of preference ordering.
- The description or learning problem - describe actions and their consequences in a systematic manner so that they can be better understood and evaluated by decision makers.
- The design problem - search for, identify or create new decision alternatives to meet the goals revealed by means of the MCDA process.
- The portfolio problem - choose a subset of alternatives from a larger set of possibilities, taking account of interactions and synergies between alternatives in addition to their individual characteristics.

Decision making applied to these multiple criteria choosing, ranking and sorting problems is known as Multiple Criteria Decision Making (MCDM).

The structuring of a decision problem is a crucial part of any decision analysis or decision aid. An overview of the process of problem structuring is provided in Section 2.1.2 which follows, but it is important to note here that problem structuring

does not simply precede the investigation of a decision problem. Rather it is an integral part of the process of decision aid itself.

This is highlighted by Bernard Roy in his forward to Philippe Vincke's book *Multi-criteria Decision-aid* [63]:

As far as decision aid is concerned, it may be advantageous not to separate the two stages of formulation and investigation. The multi-criteria paradigm invites us to progress simultaneously on these two fronts. The results one obtains will then necessarily depend on the process chosen to find them.

2.1.2 Problem Structuring

Before any problem from one of the categories referred to in the previous section can be addressed, it is essential to outline the structure of the problem itself. Within the context of MCDA, there are four key component areas which need to be identified and specified: alternatives, criteria, stakeholders & uncertainty. In addition to these, there may be particular environmental factors and constraints which need to be taken into account while structuring the problem.

Alternatives

The traditional focus of MCDA is on the evaluation of different alternatives (i.e. 'options'). However it is also possible for MCDA to be used in more of a design context where part of the study involves discovering or defining different potential alternatives. The challenge may be managing the wealth and complexity of options available, or it could be to find any suitable alternatives at all.

Alternatives may range from simple choices or actions to complex strategic plans. They could be explicitly defined and finite (e.g. which person to employ from a short-list of job applicants) or implicitly defined and infinite (e.g. how to best allocate resources like time and money to competing projects).

We will use the index i for alternatives and denote the i^{th} alternative by a_i .

Criteria

The fundamental basis of MCDA is that multiple key factors are identified which will form the basis of an evaluation of the different alternatives. These key factors or ‘criteria’ will commonly reflect specific values, objectives or points of view. Though criteria are incorporated differently into the modelling process based on the MCDA approach used (see Section 2.1.3), there are a number of universally relevant considerations when identifying and specifying different criteria to be used in a decision problem.

The concept on which a criterion is based should be unambiguously linked to the values on which it is measured, and should be understood by all decision makers involved in the analysis so that preferences can be specified which relate directly to the concept itself. The performance of an alternative on a particular criterion should always be measurable in a consistent manner and each criterion should be measuring only one factor. If the criteria overlap then concepts which are accounted for in more than one criterion will take on higher levels of importance in the analysis.

Criteria should also be specified such that preferences with respect to a single criterion, or trade-offs between two criteria, do not depend on the level of another criterion. The theoretical validity of the value function methods (which will be discussed in Section 2.1.4) is dependent on this principle, which is referred to as “judgemental independence”. Though it is not a specific requirement for other MCDA approaches it is still considered desirable.

The criteria identified should ideally be both complete, in the sense that collectively they cover all the important factors of the problem under consideration, and concise. These properties are potentially conflicting, and need to be balanced as best

as possible. Dependent on the context, it is also important to ensure that the level of detail and amount of information required for the criteria is within the operational parameters for the analysis - time in particular may be an important consideration.

As convention, we will adopt the use of the index j for criteria.

Stakeholders

With any multi-criteria analysis, it is always important to identify the different stakeholders. Stakeholders are simply people or groups of people who have a shared interest or concern in the decision under consideration and the consequences of this decision. The decision making group may not necessarily wish to take account of the different views of other stakeholders, but enforcing the consideration of different perspectives can encourage more open thinking about a problem and also allows for anticipation of the potential reactions others may have to different alternatives.

A key factor can be the amount of influence that certain stakeholders may have. If they are able to interfere with or disrupt the outcome of a decision, then often their perspective needs to be taken into account. Different stakeholders can be specifically involved in the decision making process, or sometimes it is sufficient to simply involve them by means of role play within the existing group of decision makers.

Stakeholders can then be incorporated into the modelling process in some way. Different approaches have been developed within different modelling contexts, and one key way in which these approaches differ can be whether the same model is shared by all stakeholders (but factoring in different preferences), or whether differently structured models are developed for different stakeholders.

Uncertainty

In almost all situations where a decision needs to be made, there is usually an inherent degree of uncertainty which adds to the complexity of the decision. Careful

structuring of the decision problem and analysis of data which has been gathered can aid better understanding of the nature of uncertainty and may also reduce it to some extent.

Uncertainty may present itself with respect to almost anything that relates to the decision context, whether that be other decision areas, the environment, the problem structuring process, or the use of resulting models and the interpretation of results. It is common to distinguish between internal uncertainty and external uncertainty.

Internal uncertainty refers to both the structuring of the model and the subjective inputs this requires. Some aspects of model structuring relate to imprecision or ambiguity of meaning and can thus be resolved. Other aspects may be unresolvable (e.g. which alternatives to include in the analysis, what level of detail is required). For this reason, the structuring process usually involves a number of iterations considering the problem from different angles. Uncertainty regarding the judgemental inputs to models relates mostly to the more quantitative assessments required for preference modelling.

External uncertainty pertains to the lack of information about the consequences of a particular choice. This may be because the outcome of other interconnected decisions is not yet known, or it could relate to inherent randomness in ‘environmental processes’ which are beyond control.

The handling of uncertainty in MCDA will be dealt with in more detail in Section 2.1.5.

2.1.3 Overview of MCDA Methods

Though MCDA as a scientific field is fairly young, there is already a wealth of literature on the topic. Different methodologies have been rapidly developed due to the large and varied number of real-life problems that lend themselves to the application

of this science. Many of these methodologies ‘borrow’ from each other and are constantly being adapted to handle different challenges, so it can be difficult to classify all available methods into clearly distinct groups, and there are thus a few variations in classification. Since most of these methods differ largely in their approach to preference modelling, this forms the main basis for classification. According to Vincke [63], in the past MCDA methods were generally loosely divided into three large families which in very general terms made use of complete aggregation, partial aggregation, or local aggregation of preference criteria. More recently, MCDA methods have been better classified by Belton and Stewart [4] into:

1. Value Measurement Methods
2. Satisficing and Aspiration based Methods
3. Outranking Methods

Value measurement methods

The value measurement approach aims to develop a way of associating a real number value with each alternative, in order to produce a preference order on the alternatives consistent with the judgements of the decision maker. A value function $V(a_i)$ is created which takes all criteria into account, such that alternative a_1 is considered preferable to a_2 if and only if $V(a_1) > V(a_2)$. Indifference between a_1 and a_2 is then implied if and only if $V(a_1) = V(a_2)$.

The value measurement approach assumes that preferences are both complete and transitive. Completeness means that for any pair of alternatives, either one is preferred to the other or there is indifference between them. If preferences are transitive, the implication is that if a_1 is preferred to a_2 and a_2 is preferred to a_3 , then a_1 must also be preferred to a_3 , and similarly for indifference between alternatives.

These preference assumptions seem simple and sensible (i.e. ‘rational’), but their implications are surprisingly strong, and as a result are often not rigidly adhered to

in reality. When implementing the value measurement approach, minor violations of the preference assumptions tend to be overlooked, but the results are generally always subjected to extensive sensitivity analysis as a safeguard.

In order to measure the relative importance of different performance levels for each identified criterion, the first aspect of the approach is to construct marginal or partial value functions $v_j(a_i)$, for each criterion. These partial value functions must satisfy the previously mentioned properties of completeness and transitivity on each criterion, and in addition need to model the strength of preference in some sense. The argument of the partial value function $v_j()$ could be the alternative a_i itself or some performance attribute $z_j(a_i)$ associated with alternative a_i on criterion j . The notation is used interchangeably so that $v_j(a_i)$ and $v_j(z_j(a_i))$ are assumed to be equivalent.

Secondly, scores obtained on each of these partial value functions for the different criteria are aggregated to obtain a single overall value score for each alternative. This aggregation process takes into account the relative importance levels of different criteria by weighting scores on the criteria differently. Value measurement methods (also known as value function methods) will be dealt with in more detail in Section 2.1.4.

Satisficing and aspiration based methods

The satisficing model was suggested by Nobel prize winning economist Herbert Simon [49, 50] as a descriptive model of a heuristic which people use as a result of what he termed 'bounded rationality'. Within the context of MCDA the general assertion is that decision makers would initially focus on what they perceive to be the most important criterion, and eliminate potential courses of action which would not provide a satisfactory level of performance on this criterion, before moving on to evaluate the performance of the remaining alternatives on the next most important criterion, again eliminating more alternatives with unsatisfactory performance

on this criterion. This process then continues for all the criteria in a dynamic manner as the decision maker may if necessary either backtrack to include previously eliminated alternatives (i.e. ‘lower the bar’) or re-cycle through criteria with more stringent performance requirements (i.e. ‘raise the bar’) until coming to a decision. One example of an operational implementation of the satisficing heuristic is the goal programming approach to MCDA.

Outranking methods

The French inspired outranking methods have traditionally been particularly popular in Europe and focus on building an outranking relation which represents strongly established preferences on the part of the decision maker(s). As with the satisficing models, it is assumed that partial preference functions $z_i(a)$ have been defined for each criterion. In contrast to the strict properties required for value functions, these preference functions need only satisfy ordinal preferential independence (i.e. it must be possible to rank order alternatives on one criterion independently of the other criteria).

The main principle in outranking is a generalization of the concept of dominance in the context of the value function methods. For two alternatives a_1 and a_2 , if $z_i(a_1) \geq z_i(a_2)$ for all criteria i , and there is strict inequality $z_i(a_1) > z_i(a_2)$ for at least one criterion, then alternative a_1 is said to dominate a_2 . More generally, a_1 can be said to outrank a_2 if there is ‘sufficient’ evidence to justify the statement that a_1 is at least as good as a_2 when all criteria are taken into account.

By contrast to preference relationships based on value functions, the outranking definition places emphasis on the strength of evidence for the assertion that a_1 is at least as good as a_2 , and not on the strength of preference itself. This means that even if neither alternative outranks the other, the decision maker is not necessarily indifferent between the two alternatives - if there is insufficient evidence they may be deemed incomparable in some sense. Strict adherence to a value function would

require that either a_1 is preferred to a_2 , a_2 is preferred to a_1 , or they are equally preferable.

The outranking approach treats the preference functions as imprecise measures, so that alternative a_1 can only be conclusively preferred to a_2 in terms of criterion i if $z_i(a_1) - z_i(a_2)$ is greater than some ‘indifference threshold’. This indifference threshold is allowed to depend on the value of $z_i(a_1)$ since equal increments in the preference function values may not be of equal importance. A threshold function $q_i[z]$ is defined for criterion i such that alternative a_1 is preferred to a_2 if $z_i(a_1) - z_i(a_2) > q_i[z_i(a_2)]$.

This evidence is then combined across criteria in order to summarise information discriminating between alternatives. This process is dynamic and varies widely for different varieties of outranking methods.

2.1.4 Value Function Methods

Value function methods and Multi-Attribute Value Theory (MAVT) relate to the practical use of the value measurement approach outlined in the previous section. It is important to note that value functions refer to representations of preferences under certainty, while utility functions refer to representations of preferences under uncertainty [16]. Multi-Attribute Utility Theory (MAUT) will be introduced later in this section, as an extension of MAVT.

Value function methods combine two types of preference information in order to provide an overall evaluation of each of the alternatives based on the judgements of the decision maker(s). Intra-criterion information involves assessments of an alternatives performance on a particular criterion, usually by means of a partial value function which is constructed based on elicited preferences from the decision maker(s). Inter-criterion information reflects the relative importance attached to each different criterion by the decision maker(s), most often through the use of weights.

With value measurement, the properties required of the partial value functions are critically related to the form of aggregation that is used. The simplest and most widely used form of value function (and the primary focus here) is the additive model:

$$V(a) = \sum_{j=1}^m w_j v_j(a) \quad (2.1)$$

where:

$V(a)$ is the overall value of alternative a

$v_j(a)$ is a value score measuring alternative a 's performance on criterion j

w_j is a weight assigned to indicate the importance of criterion j

The reason this additive form is so widely used is because it is seen to be the easiest to understand for decision makers from a variety of backgrounds, and places relatively modest constraints on the preference structures being modelled.

The basic conditions for the validity of the additive value function model are outlined below, highlighting according to Belton and Stewart [4] the requirements of *preferential independence*, the *interval scale property*, and the *trade-off property*:

Preferential Independence For two alternatives a and b which differ on only r out of $m > 2$ criteria, let D be the set of criteria on which the alternatives differ. Thus $v_i(a) = v_i(b)$ for $i \notin D$, and it follows that a is preferred to b if and only if:

$$\sum_{i \in D} w_i v_i(a) > \sum_{i \in D} w_i v_i(b) \quad (2.2)$$

This means that a decision maker should be able to express meaningful preferences and trade-offs between levels of achievement on a subset of criteria, without needing to be concerned about the levels of achievement on the other criteria, provided that they remain fixed. For the case of only two criteria ($m = 2$), the required independence property takes a slightly different form, known as the *corresponding trade-offs* or *Thomsen* condition. This is not of

importance here, but details can be found on pg. 90-91 of Keeney and Raiffa [33].

Interval Scale Property The partial values indicate a level of performance which is fairly arbitrary in absolute terms, unless for some reason there is a natural and unambiguous zero value point. As a result, ratios of partial values will usually not be meaningful in general. Only the ratios of differences between the $v_i(a)$ will have absolute meaning independent of the choice of zero-point and scaling of partial values. An interval scale of preferences is defined by this property.

Trade-off Property The weights in an additive value function are scaling constants which balance the proportions of the criteria on different value scales. For example, if two alternatives a and b are judged to equally preferred, and differ only on two criteria r and s , then since $V(a) = V(b)$:

$$\begin{aligned} w_r v_r(a) + w_s v_s(a) &= w_r v_r(b) + w_s v_s(b) \\ w_r (v_r(a) - v_r(b)) &= w_s (v_s(b) - v_s(a)) \\ \frac{w_r}{w_s} &= \frac{v_s(b) - v_s(a)}{v_r(a) - v_r(b)} \end{aligned}$$

The assessed weights in an additive value function must thus satisfy this scaling requirement in line with the relevant trade-offs.

So far it has only been assumed that the value function preserves preference ordering, so that a is preferred to b if and only if $V(a) > V(b)$. In this case the conditions described above ensure that an additive aggregation model of preferences is suitable. If stronger properties of the value function are desired, then the above conditions may not be sufficient.

An alternative to the additive form is the multiplicative form: $\prod_{j=1}^m [v_j(a)]^{w_j}$, and the two are closely related. By taking logarithms this can be converted into an additive form - the above properties would then apply to $\log v_i(a)$. The multiplicative form

is suitable when the partial value functions have a ratio scale interpretation, due to the existence of a natural zero point. More complicated forms of aggregation exist but are rarely used in practice and won't be mentioned here.

2.1.5 Treatment of Uncertainty in MCDA

As highlighted in the introduction in Section 2.1.1, there are two possible areas of complexity to consider in the field of MCDA - balancing conflicting or competing objectives, and taking risk and uncertainty into consideration. Figueira et al. [18] identifies these as internal uncertainties which are related to decision maker values and judgements about the potentially conflicting objectives; and external uncertainties which relate to imperfect knowledge concerning the consequences of a chosen action.

Much of conventional MCDA modelling is essentially focused on situations where the primary source of decision making complexity is the multi-criteria nature of the problem and not the (potentially) uncertain nature of the decision's consequences. As a result, many MCDA models are based on deterministic evaluations of the consequences of each action in terms of each criterion, and then simply subject final results and recommendations to some sensitivity analysis.

However, in situations where risks and uncertainties are as critical as the issue of conflicting goals, it becomes necessary to model these uncertainties formally. Uncertainty about the environment is usually incorporated into MCDA by drawing from a combination of three broad approaches. These three approaches are summarised below, but will be elaborated upon further in Section 2.4 where the treatment of uncertainty for the simplified SMAA models is outlined:

Decision Theory makes use of probability to describe the likelihood of specific events, while using utility theory to model the decision maker's attitude to risk. This formal approach to modelling risk is known as Multi-attribute Utility Theory (MAUT), and provides the core foundational framework on which much of the

simulation study in Chapter 3 will be based.

Risk as a Criterion: Here the level of risk is included as one of the criteria in the analysis. Risk can either be represented by a probability of success or failure for a certain outcome, or as the variance associated with an attribute. A slight variation on this kind of approach is exemplified in the risk model used in the simulation study in Chapter 3 (see Section 2.4.2).

Scenario Planning requires decision makers to identify a few scenarios relevant to the decision context. The likelihood of different scenarios is not explicitly modelled, and not all possible scenarios are considered. The emphasis is placed on defining good strategies which will be robust over a range of potential future scenarios. This approach can be integrated with MCDA (see for example Durbach and Stewart [14]) but will not be used here.

2.2 Stochastic Multi-Criteria Acceptability Analysis

2.2.1 Motivation

Of the many methods directed at dealing with multi-criteria decision problems, a large number use weights in order to describe the preferences of the decision maker(s) with regard to the relative importance of different criteria. The additive model outlined in equation 2.1 is one example of a model which requires such weights. Though several procedures exist for eliciting these weights from decision makers, Lahdelma and Salminen [37] highlight a range of potential problems - different procedures tend to result in different weights for the same problem; it may be difficult to reach consensus about weights in a problem with multiple decision makers; and often decision makers don't want to restrict themselves to specific weights, can't do so due to the complexity of the problem, or expect that their weights will change over time.

Stochastic multi-criteria acceptability analysis (SMAA) is a recently developed family of Multiple Criteria Decision Aiding (MCDA) methods which address this problem by making use of a modified approach which allows discrete MCDA problems to be investigated despite the absence of accurate preference information from the decision makers [34]. The SMAA methods are based on exploring the weight space in order to describe the preferences that would make each alternative the most preferred one, or that would give a certain rank for a specific alternative [57].

This means that unlike most other MCDA methods, SMAA can be used in a prescriptive context with very limited preference information or to some extent even with none at all. In addition criteria measurements which are uncertain or inaccurate can also be incorporated. In fact the main advantage of the SMAA approach is that it allows both preference information and criteria measurements to be expressed as arbitrarily distributed stochastic variables.

Three types of information ignorance have been recognised and defined by Smets [51]; incomplete, imprecise, and uncertain information. Incomplete information refers to missing values. Imprecise information refers to values that are present, but not with the required precision. Uncertainty instead is a subjective form of ignorance which comes about through an agent (e.g. a decision maker) who gives information that is both complete and precise, but unreliable as it may be constructed based on a proposition that is not definitively established.

The highly flexible approach adopted by SMAA has allowed it to be applied to a range of existing MCDA methods in recent years so that they can be used with incomplete, imprecise or uncertain information.

2.2.2 Methods

Within the family of SMAA methods there are already a number of variants of SMAA which have been developed. The SMAA approach is very versatile and most of these methods extend the SMAA-type analysis to handle a range of existing methods within the general context of MCDA. Typically it is the importance weights involved in certain types of preference modelling that are allowed to vary in SMAA modelling. However, different variants of SMAA also allow other aspects of preference information to be varied.

In the original SMAA method [34], inverse weight space analysis is performed based on an additive utility or value function and stochastic criteria measurements. In SMAA-2 the analysis is generalized to a general utility or value function, in order to include various kinds of preference information and to consider holistically all ranks. SMAA-O [35] extends SMAA-2 so that mixed ordinal and cardinal criteria can be treated in a comparable manner.

SMAA-A (or Ref-SMAA) allows for modelling of decision maker preferences using reference points and achievement scalarizing functions [36]. A variant of the SMAA-A method which uses achievement functions has also been developed by Durbach

[11]. SMAA-P is a prospect theory based variant of SMAA. SMAA-3 [38] involves the application of ELECTRE III type pseudo-criteria within the existing framework of the analysis. SMAA-TRI [54] extends ELECTRE TRI [65] to allow for ignorance on the parameter values. SMAA-TRI is one of the only SMAA variants available for sorting problem statements (as opposed to those for ranking or choosing). There is also a variant of SMAA based on data envelopment analysis, which is known as SMAA-D. This is described in Lahdelma and Salminen [39].

2.2.3 Applications

As mentioned previously, MCDA is applicable in a wide number of real-life problems. Since SMAA allows for even more flexible use of the existing MCDA methods in a practical sense, it appears to have even greater potential in terms of breadth of application. SMAA provides an adaptable way to model different kinds of uncertain or inaccurate preference and criteria information through stochastic distributions. The inverse weight space approach is suitable for many group decision-making problems where decision makers are unable or unwilling to provide preference information, or may struggle to reach consensus regarding their preferences.

SMAA can be used to help the decision makers identify commonly acceptable compromise solutions, based on preference information expressed as weight intervals or some form of weight distribution that is accepted by all decision makers. In addition, SMAA computations can be implemented very efficiently through numerical methods [57], which makes it possible to use the method in interactive decision processes. For many of these reasons, in the decade or so that SMAA has been around it has already been used successfully in a number of different problem contexts, mostly in Finland where the method originated and was subsequently developed. To highlight the versatility of the SMAA methodology, we highlight and summarize a few examples of its application below.

Probably the best known and most widely documented application of the SMAA model occurred in the area of infrastructure planning in Finland. The Helsinki Harbour problem [24] was actually the catalyst in the early development of SMAA, and involved investigation of the environmental and economic impact of various alternatives under consideration for constructing a general cargo harbour in Helsinki. In addition to the construction of the harbour itself, development of the surrounding infrastructure including road and rail links needed to be considered, which led to a complex decision making context that involved a large number of stakeholders each with different underlying preferences.

The relevance and applicability of SMAA and particularly SMAA-O in forest and environmental planning processes has also been demonstrated in a number of studies. According to Kangas and Kangas [31], decision problems involving incomplete or imprecise data on both cardinal and ordinal criteria are becoming more common in the planning of forestry practice, and the already wide potential for application of SMAA-O in natural resource management will most probably increase in the future.

SMAA has been proposed and evaluated by Durbach [12] as a model of purchase behaviour, also assisting with the prediction of customer relative purchase frequency or defection. The evaluation was done using two longitudinal datasets on Fast Moving Consumer Goods (FMCG) in Europe, and the results seem to validate the usefulness of SMAA in descriptive decision making contexts in addition to the prescriptive contexts they were developed for.

Work by Alvarez-Guerra et al. [2] illustrated the suitability of SMAA for the prioritization of sediment management alternatives by applying the methodology to a case study in the bay of Santander in Spain. The least preferred alternatives were identified for each of four different hypothetical preference profiles (Idealist, Politician, Environmentalist, Balanced) by considering criteria involving the technical, economic, social and environmental aspects of the process.

Tervonen et al. [55] made use of SMAA-TRI to re-analyse a case study which concentrates on Frances Lorraine region, where iron has been mined for more than a century. The objective of the original study was to partition the land into zones and assign these zones into 4 pre-defined risk categories based on 10 available criteria. Another application of SMAA is presented in Lahdelma et al. [40] where ordinal criteria measurements were used with no information on decision maker preferences in order to inform the choice of location for a waste treatment facility near Lappeenranta in South-Eastern Finland.

An interesting application of SMAA to the elevator planning problem is dealt with by Tervonen et al. [56]. In addition to developing elevator systems in accordance with specific minimum requirements for a number of standard performance criteria, it is desirable to optimise the elevator configuration based on other non-essential performance criteria, including the economy and service level of the group of elevators. Examples of these criteria include total cost, floor area occupied, journey time and waiting time. Since these goals are clearly conflicting, and stakeholders (representatives of the elevator company, consultants or customers) have different preferences for the semi-interdependent criteria, a compromise needs to be reached. Ten feasible elevator group configurations for a 20-floor building were compared based on mixed type criteria including ordinal information and deterministic values. Results of the analysis showed that SMAA is effective in recognizing acceptable solutions in elevator planning, and can assist with determining compromise solutions or those which are favoured by different groups of decision makers.

2.3 The SMAA-2 Model

The fundamental idea of SMAA is to provide decision support by means of descriptive measures which are calculated as multidimensional integrals over stochastic parameter spaces which account for partially known preference information and uncertain criteria values. In the general case these descriptive measures must be calculated by means of numerical techniques. Generally Monte Carlo simulation is used as it is simple and able to provide sufficient accuracy.

Section 2.2.2 highlighted many of the SMAA variants and the specific problem contexts for which they have been developed. We now turn our focus to the basic SMAA-2 method as outlined in Lahdelma and Salminen [37]. The fundamentals of this particular SMAA variant will be addressed in detail here as they form the basis for later modelling in both the simulation study in Chapter 3 and the application in Chapter 5. The key idea in SMAA-2 is the application of inverse weight space analysis in order to describe for each alternative what range of preferences make it the most preferred one, or result in it achieving a particular rank.

Consider a decision problem where a group of decision makers need to choose between m alternatives $x_1, x_2, x_3, \dots, x_m$ based on an evaluation of n criteria measurements on each alternative. Without loss of generality we assume that all criteria are to be maximized.

2.3.1 Preference Information

It is assumed that the decision makers' preference structure can be represented by a real-valued utility or value function

$$u_i(w) = u(x_i, w) \tag{2.3}$$

This value function maps each of the alternatives to a single utility value based on a weight vector \mathbf{w} for each decision maker, which quantifies their relative preferences

for different criteria. Clearly if all criteria values were fully known and the decision makers could all agree on a single weight vector then total utility could be calculated for all of the available alternatives which could subsequently be ranked from best to worst. However, as explained on page 24, many decision makers find it difficult to fully specify their own preferences, so it is even more unlikely that they would all be able to agree on a single weight vector for the problem under consideration.

Generally weights are assumed to be non-negative and standardized to sum to 1 (though this is not strictly required). We thus consider the following set of feasible weights for the decision makers' unknown or partially known preferences:

$$W = \left\{ w \in R^n : w \geq 0 \text{ and } \sum_{j=1}^n w_j = 1 \right\} \quad (2.4)$$

This forms an $(n - 1)$ -dimensional simplex. A total lack of preference information on the weights is represented in 'Bayesian spirit' by a uniform weight distribution in W , such that:

$$f_W(w) = \frac{1}{\text{vol}(W)} \quad (2.5)$$

Even with a total lack of preference information one can usually gain some insight into the decision problem, but most of the time it is possible to elicit some form of preference information from the decision makers. Although the SMAA methods allow this preference information to be represented with almost any density function, it is generally easier to represent preferences by constraining the weight space, and then apply a uniform distribution in the restricted weight space:

$$f_{W'}(w) = \left\{ \begin{array}{ll} \frac{1}{\text{vol}(W')} & \text{if } w \in W' \\ 0 & \text{if } w \in W \setminus W' \end{array} \right\} \quad (2.6)$$

According to Lahdelma and Salminen [37], there are 5 types of weight restriction which can be used:

1. Partial or complete ranking of weights - $w_j \geq w_k$ for some j, k
2. Intervals for weights - $w_j \in [w_j^{min}, w_j^{max}]$
3. Intervals for weight ratios - $w_j/w_k \in [w_{jk}^{min}, w_{jk}^{max}]$
4. Linear inequality constraints - $A\mathbf{w} \leq c$
5. Non-linear inequality constraints - $g(\mathbf{w}) \leq 0$

2.3.2 Acceptability Analysis

In addition to accommodating vague or unspecified preference information, SMAA also allows for uncertain or imprecise criteria values which are represented by stochastic variables ξ_{ij} with joint density function $f_X(\xi)$ in the space $X \subseteq R^{m \times n}$. Any distribution can be used in principle, but use of either the normal distribution or the uniform distribution is most common in practice.

SMAA-2 is then primarily based on the analysis of stochastic sets of ‘favourable rank weights’ $W_i^r(\xi)$ which are defined such that any weight vector w drawn from the set results in alternative i achieving rank r :

$$W_i^r(\xi) = \{w \in W : rank(i, \xi, w) = r\} \quad (2.7)$$

The rank of an alternative is formally defined as follows:

$$rank(i, \xi, w) = 1 + \sum_{k \neq i} \rho\left(u(\xi_k, w) > u(\xi_i, w)\right) \quad (2.8)$$

where $\rho(true) = 1$ and $\rho(false) = 0$, so that $rank(i, \xi, w) \in \{1, 2, \dots, m\}$.

We first describe three primary measures: (1) rank acceptability indices, (2) central weight vectors and (3) confidence factors. We then consider some secondary

measures which combine these primary measures on different ranks: (4) k best ranks acceptabilities and (5) holistic acceptability indices.

Rank Acceptability Index

The rank acceptability index b_i^r indicates the portion of parameter values (across both the weight and criteria distributions) which will result in alternative i obtaining rank r .

$$b_i^r = \int_{\xi \in X} f_X(\xi) \int_{w \in W_i^r(\xi)} f_W(w) dw d\xi \quad (2.9)$$

The rank acceptability indices are within the range $[0, 1]$, with 0 indicating that the alternative will never obtain the given rank and 1 indicating that the alternative will always obtain the given rank, regardless of the choice of weights. Thus the best alternatives will be those that have high acceptabilities for the smallest (best) ranks. Though the rank acceptability indices simultaneously take into account the uncertainties on both criteria values and decision maker preferences, it is important to note that they are still only applicable given the assumed preference model of utility.

The first rank acceptability index with b_i^1 is called the acceptability index a_i . The acceptability index a_i can be used to classify alternatives into stochastically efficient ($a_i \gg 0$) or inefficient (a_i near zero). The magnitude of the acceptability index also indicates the strength of the efficiency for each alternative, taking into account the the specified uncertainty in criteria evaluation and decision makers' preferences. The original SMAA method considers only the acceptability index b_i^1 , and is not intended for directly ranking the alternatives but rather for the rough classification of alternatives into those that are acceptable and those that are not.

Lahdelma and Salminen [37] highlight three main difficulties which can result in bias in the acceptability indices resulting from the original SMAA method:

1. Scaling of the criteria affects the acceptability indices
2. Low acceptability indices can be misleading if the assumed weight distribution

does not accurately represent the decision makers' preferences

3. SMAA ignores information about other ranks which may be needed to discern the following three problems:
 - (a) Extreme alternatives may obtain excessively high acceptability
 - (b) Neighbouring alternatives decrease each other's acceptability
 - (c) Solid compromise alternatives which are very likely to obtain a high rank (e.g. second or third), may hardly ever be ranked best, and will thus obtain (deceptively) small acceptability

These particular difficulties form the primary motivation for the SMAA-2 method, which extends acceptability analysis from the original SMAA method to consider all ranks, so that potential compromise alternatives can be better identified. Considering other rank acceptabilities provides additional information for the acceptability analysis. Two particular ways this information can be utilized are the k -best ranks acceptabilities and holistic acceptability indices which are described later in this section.

Note that the discussions of the central weight vector w_i^c and the confidence factor p_i^c (which follow) pertain to the best ranked alternative and are thus common to both the original SMAA method and SMAA-2.

Central Weight Vector

The central weight vector w_i^c is the expected centre of gravity (centroid) of the favourable first rank weights for alternative i . Given the assumed preference model, the central weight vector represents the 'average' or 'typical' preferences of a decision maker who supports this alternative.

$$w_i^c = \frac{1}{b_i^1} \int_{\xi \in X} f_X(\xi) \int_{w \in W_i(\xi)} f_W(w) w dw d\xi \quad (2.10)$$

The central weight vector is a key tool in the inverse approach to decision support which sets SMAA apart from many other methods. Instead of eliciting preferences

and then building a solution to the decision problem, the approach is to present the central weight vectors of different alternatives to the decision makers in order to help them learn how different weights correspond to different choices with the assumed preference model.

Confidence Factor

The confidence factor p_i^c measures the probability that alternative i will obtain rank one if preferences are represented by the central weight vectors. Confidence factors can thus also be interpreted as an indication of whether criteria data are accurate enough to discern whether the alternatives are efficient.

$$p_i^c = \int_{\xi \in X: u(\xi_i, w_i^c) \geq u(\xi_k, w_k^c) \forall k=1, \dots, m} f_X(\xi) d\xi \quad (2.11)$$

Confidence factors can be calculated for any given weight vectors and not just for the central weight vector.

k Best Ranks

The kbr-acceptability a_i^k measures the variety of different valuations that would result in alternative i being assigned any one of the k best ranks:

$$a_i^k = \sum_{r=1}^k b_i^r \quad (2.12)$$

A central kbr weight vector w_i^k can also be defined:

$$w_i^k = \frac{1}{a_i^k} \int_X f(\xi) \sum_{r=1}^k \int_{W_i^r(\xi)} f(w) w dw d\xi \quad (2.13)$$

Given the assumed weight distribution, the central kbr weight vector is simply a single vector representation for the preferences of typical decision maker who would assign alternative i a rank between 1 and k .

Similarly, the kbr confidence factor p_i^k is defined as the probability that alternative i is assigned on of the top k ranks if the central kbr weight vector is chosen. It is calculated as an integral over the criteria distribution:

$$p_i^k = \int_{\xi: \text{rank}(\xi_i, w_i^k) \leq k} f(\xi) d\xi \quad (2.14)$$

Holistic Acceptability Indices

Rank acceptabilities can also be combined in a more general form into holistic acceptability indices:

$$a_i^h = \sum_r \alpha^r b_i^r \quad (2.15)$$

The meta-weights α^r can of course take on a wide range of values. Special cases include:

- $\alpha = (1, 0, \dots, 0)$ - the acceptability index itself,
- $\alpha = (1, \dots, 1, 0, \dots, 0)$ - the kbr acceptabilities with k ones and $m - k$ zeros,
- and $\alpha = (1, 2, \dots, m)$ - which is equivalent to calculating the expected rank of alternative i .

In general, a complete priority order between the meta-criteria (rank acceptabilities) is well justified: $\alpha^1 \geq \alpha^2 \geq \dots \geq \alpha^m \geq 0$. In addition it is convenient to normalize $\alpha^1 = 1$ so that the best achievable holistic acceptability index is 1. Thus meta-weights can range from placing all importance on the best rank $\alpha = (1, 0, \dots, 0)$ to placing equal importance on all ranks $\alpha = (1, 1, \dots, 1)$.

2.3.3 Computation

In practice, the multidimensional integrals outlined previously (Section 2.3.2) are usually computed using numerical analysis of some kind, since in the general case they are impossible to evaluate analytically. Monte Carlo simulation is a well-established method for computing approximate values for high-dimensional integrals [23] and

since a high level of precision is not required this is commonly used in applications of SMAA. Monte Carlo simulation can also be implemented with sufficient efficiency to make it suitable for large decision problems and even discrete approximations of continuous problems [57].

With SMAA implemented by generating preferences and attribute evaluations randomly using computer simulations, the acceptability index b_i^r (2.9) is simply the relative proportion of all simulation runs in which a_i obtains rank r . Similarly, the integral representation of the central weight vector (2.10) is not evaluated directly; instead it is computed from the empirical averages of all weight vectors supporting the selection of a_i as the best alternative i.e. the j -th element of \mathbf{w}_i^c is the average of all weights for attribute c_j in W_i^1 . Note that in most applications (including ours) the utility functions are considered as quantities that are “controlled for” in a statistical sense, so that only the weight vectors are considered uncertain and allowed to vary.

2.4 Treatment of Uncertainty

The key aim of the study is to compare some of the potential approaches to the handling of uncertainty on the attribute evaluations. In conventional SMAA, the approach is to fit distributions (possibly multivariate) to the attribute evaluations on the different criteria for all the alternatives, and to then draw random realizations from these distributions in each run of the SMAA method. This approach is exemplified by the model outlined in Section 2.4.5.

By contrast, it is also possible to base the SMAA model on various summaries of uncertainty, which may include expected values, variances or different quantiles of the attribute evaluations. These summary measures are generally much simpler to obtain, and this is the primary motivation for their potential use, provided that models based on these measures can still provide decent accuracy.

In a real life decision making problem of this kind, it is likely that summaries of

the attribute evaluations on the different alternatives may be prone to error, so we attempt to account for this in the simulation by introducing a fixed multiplicative error on the summary measures. All inputs to the models (fitted distribution parameters, expected values, variances, and quantiles) are multiplied by randomly generated realizations from $U(1 - v, 1 + v)$. The error factor v is a simulation parameter which takes on one of three values : 0% (no error), 10% (small error), 20% (substantial error). The error adjusted parameter values are then denoted using the ‘hat’ symbol e.g. 5% quantile values $\hat{Q}_{0.05}(\hat{Z}_{ij})$

We now summarize the five classes of decision model that will be used [13]:

2.4.1 Model using expected values

The model using only expected values ignores any uncertainty in the attributes and is thus the simplest of all the models to be considered here.

$$U_i^{(\text{ev})} = \sum_{j=1}^n w_j u_j(\hat{E}[\hat{Z}_{ij}]) \quad (2.16)$$

2.4.2 Model using expected values and variances

The model outlined below uses both expected values and variances and is referred to here as a risk model. It is an extension of the previous model which attempts to balance the measurement of utility based on both the average level of performance and the amount of inconsistency present.

$$U_i^{(\text{risk})} = \sum_{j=1}^n w_j u_j(\hat{E}[\hat{Z}_{ij}]) - p \sum_{j=1}^n w_j \hat{\text{V}}\hat{\text{A}}\text{R}[\hat{Z}_{ij}] \quad (2.17)$$

The value p is a weight factor which controls to what extent the average performance on a criterion should be penalized for inconsistency. We set $p = 0.25$.

2.4.3 Model using quantiles with fixed weights

This fixed weight quantile model makes use of what is referred to in Keefer and Bodily [32] as the “extended Pearson- Tukey” approximation. They compared a number of approximations used to serve as substitutes for the probability distributions of continuous random variables and found this to be a widely-applicable general-purpose three-point approximation for continuous probability distributions. It is highly robust and also more accurate than other similar approximations in estimating means and variances of distributions typical of those elicited via judgemental assessments.

$$U_i^{(\text{quan-fix})} = \sum_{j=1}^n w_j \sum_{r=1}^3 x_r u_j(\hat{Q}_r[\hat{Z}_{ij}]) \quad (2.18)$$

with $x_1 = x_3 = 0.185$ and $x_2 = 0.63$ and $Q_1 = 5\%$ quantile, $Q_2 = 50\%$ quantile, $Q_3 = 95\%$ quantile

2.4.4 Model using quantiles with variable weights

Another more general model using variable weights for the different quantiles is also implemented here (particularly to allow for comparison with the extended Pearson-Tukey approximation). The same three quantiles (5%, 50% and 95%) are used, and the weights x_r are constrained to sum to 1.

$$U_i^{(\text{quan-var})} = \sum_{j=1}^n w_j \sum_{r=1}^3 x_r u_j(\hat{Q}_r[\hat{Z}_{ij}]) \quad (2.19)$$

$Q_1 = 5\%$ quantile, $Q_2 = 50\%$ quantile, $Q_3 = 95\%$ quantile

As highlighted in Wang and Zions [64], many of the more obvious approaches to the generation of uniformly distributed weights given linear restrictions do not in fact result in uniformly distributed values. The method employed here, also supported by Wang and Zions [64], is as demonstrated by Tervonen and Lahdelma [57].

We want to generate weights X such that:

$$X = \left\{ x \in R^p \text{ and } \sum_{r=1}^p x_r = 1 \right\} \quad (2.20)$$

The weights x_r are generated according to the following method:

1. Generate $p - 1$ independent random numbers from a uniform distribution in the interval $[0, 1]$.
2. Sort these in ascending order and denote as y_1, y_2, \dots, y_{p-1} .
3. Let $y_0 = 0$ and $y_p = 1$.
4. Uniformly distributed, normalised weights are then obtained as intervals between the consecutive numbers i.e. $x_r = y_r - y_{r-1}$

2.4.5 Model using a fully fitted distribution

This is the conventional stochastic model used in SMAA, and is included here as a benchmark against which the previously outlined models can be compared. For each run, the utility of each alternative is calculated based on generated attribute values from a statistical distribution fitted on each of the criteria.

$$U_i^{(\text{dist})} = \sum_{j=1}^n w_j u_j(\hat{Z}_{ij}^{\text{fitted}}) \quad (2.21)$$

Here $\hat{Z}_{ij}^{\text{fitted}}$ is a standardised attribute value taken from a pool of randomly generated values from a (potentially erroneously) fitted distribution based on the observed (unstandardised) data. Unlike for the simplified models above, in this case the error factor is applied to the parameters of the fitted distribution before generating realizations from it, and the realizations are then standardised to lie between 0 and 1.

However, it must be noted that with $nZ = 100$ the observed empirical distribution will differ slightly from the expected theoretical Gamma distribution. This means that the standardised values from the resulting fitted distribution (which is strictly Gamma) will have a ‘built-in error’, even when the error factor itself is set to zero. In addition, since the process of standardization is dependent on the range of the sample, this could introduce further error.

We avoid this problem by drawing attribute values from the empirical distribution of the standardised attribute values (i.e. sampling directly from the \hat{Z}_{ijk}) rather than from the theoretical distribution. Since the bootstrap sample replicates the ‘true’ empirical distribution it is expected to result in the best fitting SMAA model overall. This bootstrap sample will be used for the ‘no error’ distributional model, and we then designate the fitted distributional SMAA model with error factors $v = 0, 0.1$ as the ‘small error’ and ‘substantial error’ models respectively. Obviously the categories of assessment error are not directly comparable across the different model types (since errors have to be incorporated slightly differently for the different models), but should still give some idea of how the models perform when inputs are erroneously assessed. Further details on the implementation of these models will follow later in Chapter 3, and an outline of the different models and error categories is provided in Table 3.2.

2.5 Science and Sport

The scientific analysis of different sporting codes by various means has brought a widely varying degree of success over the years. Some sports are naturally more complex than others and thus much harder to approach analytically. For example, “rugby is known for its complex laws and unique game structures, which render research from other sports as largely inapplicable” [1]. The innate degree of complexity (or perhaps variability) of a sport is contributed to, amongst other things, by the number of active participants (e.g. individual & team sports), the environment in which the game is played (e.g. indoor & outdoor sports), game length (e.g. Test match & One Day Cricket), and the number of different ways of scoring.

Sports science typically includes aspects of psychology [9], physiology [48], and biomechanics [66], as well as nutrition and diet [29]. There is continual development in all of these areas, but the rapid development of information technology in recent years has been highly conducive to the advancement of notational analysis as the emerging form of investigation.

Notational analysis centers on comprehensive analysis of the behavioural aspects of sports performance by attempting to objectively record critical game events in a consistent and reliable manner [25]. This serves two purposes:

- to provide a direct and accurate feedback system for players, who can view summaries of their match statistics and performance, and watch video replays of specific events (or passages of play) in order to evaluate techniques, successes and errors etc.
- to collect detailed match information for coaches who can then use this in order to review and assess player performance, and to inform decision making, strategy and tactics.

Both forms of feedback are central to the process of improving performance, and in ideal situations could even lead to a team or franchise gaining competitive advantage.

An example of the kind of success that can be achieved through the application of scientific analysis to sport was recently described in the book *Moneyball* [42]. Due to the increasing wealth disparity in baseball at the turn of the 21st century, the prevailing sentiment was that the game was ceasing to be a sporting competition and instead becoming a financial one. The richest teams could outbid everyone else in order to buy the best players, and the poorer teams seemed destined to fail. However, the Oakland Athletics baseball team defied this perception when over a period of several years they won more regular season games than all but one of their twenty-nine opponents, despite being one of the poorest teams in Major League Baseball.

The book describes the team's success and attributes it to a re-thinking of the approach to player selection, inspired by Oakland's Athletic general manager Billy Beane:

“In what amounted to a systematic scientific investigation of their sport, the Oakland front office had re-examined everything from the market price of foot speed to the inherent difference between the average major league player and the superior Triple-A one. That's how they found their bargains. Many of the players drafted or acquired by the Oakland A's had been the victims of an unthinking prejudice rooted in baseball's traditions.”

There was a lot of criticism from baseball insiders following the release of the book denouncing Beane and Lewis as charlatans, but Hakes and Sauer [22] verify that hitters' salaries during the period didn't accurately reflect the contribution of various batting skills to winning games and exploiting this inefficiency enabled the Oakland Athletics to gain a substantial advantage over their competition. Study of baseball's labour market also revealed that market adjustments around the time the book was published took place with sufficient force that it no longer exhibits the 'Moneyball anomaly'. Economic issues aside, the book serves at least to demonstrate the huge potential there is for performance analysis, and the related statistical analysis in

particular, to be successfully applied to sport.

For most sports, as with baseball, the coaching and selection process has traditionally been almost entirely based on subjective observations of players by the coaching staff. Though more objective methods of analysis have been incorporated into some sports over time, it seems the ‘opinion’ of the coaching staff is still heavily relied upon in many professional sporting disciplines, and quite naturally, particularly in complex sports where scientific analysis is refractory [17].

However, research has shown that human memory of visual observation is both selective and constructive. As a result of this, visual recall is not only highly subjective, but often unreliable and inaccurate. According to Neisser [45]:

“Neither perception nor memory is a copying process. Perception and memory are decision-making processes affected by the totality of a person’s abilities, background, attitudes, motives and beliefs, by the environment, and by the way his recollection is eventually tested.”

Very little research has been completed in the specific area of observational accuracy within the discipline of sports science, but considerable applied research done on eyewitness testimony has highlighted the weaknesses of human perception and memory, and has led to the discovery of many sources of unreliability. Factors which have a potential effect on memory include:

1. the apparent insignificance of (some) events at the time they are witnessed
2. less than ideal conditions for observation - distance, fast movement, the presence of a crowd etc
3. the observer’s personal condition (e.g. being in a good or bad mood [21] or being under stress [10])
4. a natural tendency to see what we one wants to see
5. memory loss over time

6. conscious or unconscious persuasion from other individuals
7. propensity to conform to the majority opinion

Clearly many of these factors will have a potential effect on coaches observations of players during matches. In fact, Franks and Miller [19] considered the coach as an eyewitness in competitive sports, using methodology gained from applied memory research to show that soccer coaches at international level could only recollect 30% of the key elements determining successful soccer performance observed during one half of a televised game.

There is thus quite clearly a need for alternative forms of feedback to both coaches and then the players themselves. Within the discipline of sports science, this has been the primary incentive for the development of the notational analysis systems referred to on page 41.

It is hoped that by facilitating the observation and recording of the behaviour of players objectively, notational analysis will provide coaching staff with accurate information on which to base decisions about game tactics etc. This will also provide athletes with insights into many aspects of their own performance, which can be used for player evaluation and correctional coaching where necessary.

In team sports such as soccer and rugby, one of the hugely important ‘decisions’ which coaching staff are involved in on ongoing basis is the selection of players to fill the required positions within the squad. This is another area where the information provided by notational analysis is of use. In Chapter 5 we hope to explore application of the previously outlined SMAA methodology to this problem of player selection in rugby. To provide some context for this case study, the next section will provide an overview of rugby union and some of the related research within the field of sports science.

2.6 Analysis of Rugby Union

2.6.1 Description of the Game of Rugby

The following description of the game of rugby is based primarily on the International Rugby Board's 2008 book of complete laws of the game [27].

Rugby is a ball game played by two teams, each of fifteen players, on a rectangular field of play about 100 meters in length and 70 meters in width. On either end of the field of play, centered, are a set of goal posts 5.6 meters apart and at least 3.4 meters high, with a crossbar between them at a height of 3 meters. There is also a rectangular in-goal area on either end of the field of play (i.e. behind the goal posts).

The object of the game is to score as many points as possible, gaining territory by carrying, passing, kicking and grounding the ball, so as to obtain a good field position from which to score. There are two primary ways a team can score points:

Try A try is scored when an attacking player is first to ground the ball in the opponents in-goal area. If a player would probably have scored a try but foul play by an opponent prevented him from doing so, a penalty try may also be awarded.

Conversion/Penalty/Drop Goal A player may attempt to score points by kicking the ball between the opposition's goal posts (and over the crossbar). A conversion kick takes place after the attacking team scores a try. A penalty kick can take place when the referee grants the attacking team a penalty due to an infringement from the defending team. Time is allocated for these two types of kick so that the kicker can take time to place the ball on the ground and prepare for the kick. By contrast a drop goal takes place during general play, when the ball is deliberately dropped onto the ground and kicked as it bounces.

The game is made up of two halves, each forty minutes in length, with additional time added on to compensate for time lost to injury and other stoppages in play,

and a ten minute interval in-between the two halves. The team that has acquired the most points by the end of the game is declared the winner.

2.6.2 Recent History of the Game

In August 1995 the International Rugby Board declared that the game of rugby union would become professional, removing all restrictions on payments or benefits to those connected with the game. The game has thus undergone major change over the past 15 years as the sport has adjusted to its professional structure, and the level of competition between teams and franchises has increased drastically. However, despite this the academic study of rugby has developed fairly slowly, particularly by comparison to soccer, golf, cricket, and various racket sports [43].

Due to the ever growing amount of frequency data captured by computerized notation systems in rugby, the application and involvement of statistical analysis is now common in practice, but research of this nature is generally undertaken within the confines of the team, organization or governing body environment, and due to the fierce competition between different teams and franchises it usually remains proprietary. Consequently, new developments in analysis often don't find their way into academic publications and are thus not subjected to wider academic scrutiny [28]. This no doubt hampers the advancement of research, despite the increased academic interest and involvement in game analysis within rugby union over recent years.

While rugby has been given attention in the literature across various sub-disciplines of sports science - namely sports medicine, physiology, psychology and biomechanics [43] - we focus here on the primary area of quantitative research in rugby which is performance analysis. We first overview some of the research that has been done in match analysis and then look at the development of player performance indicators, bearing in mind that there is an overlap and that research in each of these areas informs developments in the other.

Reference	Competition	Summary
[61]	World Cup 2003	superior performance linked to possession retained, number of points scored in the 2nd half & propensity to lose possession in areas of the field from which opposition is likely to score
[62]	World Cup 2003	Clear differences between teams reaching the semi finals & SA (knocked out at in quarter-final). Differences can be summarized as a greater ability to retain the ball for longer & to move possession from defensive to attacking half of the field.
[47]	Super 12 2005	Compared to the SA teams, Crusaders have lowest defensive & offensive recycle times, indicating a higher tempo of play. Conceded similar amount of tries as SA teams, but scored more tries (i.e. attack & not defence, seems key to success). All teams obtained similar & conceded similar percentages of fast ball, indicating that obtaining quick ball from tackle situation is not a predictor of scoring success. All teams committed support players to attacking tackle situations emphasizing the importance of ball retention. At defensive tackle situations, the Crusaders & Stormers generally committed no support players so they could have greater numbers in their defensive line. This strategy thought to be advantageous to the defending team.
[60]	Super 14 2006	meters gained, kicks out of hand, line-breaks & percentage tackles made were the most important discriminators between successful & less successful teams. line breaks, kicks out of hand, rucks won, percentage tackles made & meters gained contributed the most to variance in team rankings.
[44]	Tri-nations, Six nations and Test matches 2004-2006	In order to increase success rate of ball acquisition in kick-off play with respect to kicking side, important for the kicking side to take some action against the ball earlier than opponent at an aerial contest for the ball, creating situations where receiving side cannot employ the lifting play will help.

Table 2.1: Summary of match analysis research highlighting predictors of success in rugby

2.6.3 Analysis of Match Data

Match analysis generally investigates descriptive aspects of matchplay, drawing comparisons between winning and losing teams, analysing general patterns of play, or focusing in on specific areas of game play (called set pieces in rugby) in an attempt to better understand how the game is played, and particularly how it is won or lost.

Table 2.1 summarizes recent research involving different aspects of match analysis - patterns of play, territory and possession, work rate, attacking and defensive strategy etc.

2.6.4 Analysis of Player Performance

A performance indicator, as defined by Hughes and Bartlett [26], is a selection or combination of action variables that aims to define some or all aspects of a performance, and should obviously relate to successful performance or outcomes in order to be useful. Performance indicators are used to assess aspects of individual or team performance and can be used for comparison with opposition players, teams etc, or in isolation as a measure of the performance of a team or individual alone.

Interestingly it has been noted that the performance indicators used by analysts within each of three categories of formal games are very similar [26]. The categories of formal games as classified by Read and Edwards (1992) are: net and wall games which are score dependent, invasion games which are time dependent, and striking and fielding games which are innings dependent. Since there are of course a number of different sports within each category, and a wide variety of game and rule variations around the world, it thus appears that the common primary objective is what drives the selection of similar key performance indicators for sports within each category of games.

Rugby falls within the class of invasion games, amongst others like American football, basketball, soccer, and hockey for example. In invasion games the purpose is

to invade the opponents territory while attempting to score points and keep the opposing team's points to a minimum, within a limited time period. Thus common performance indicators for this class of game involve data on tackles, passes, runs with the ball, and loss of ball control, despite differences in the way these physical activities are actually carried out in each sport.

One of the main issues with the use of performance indicators in rugby (and many other invasion games as well), is that most of the contributions by individual players are not easily related to the success of the team, due to the complex interactive nature of the behaviour required in order to score points. This is by contrast to sports like cricket or baseball, where a batters contribution to the team success can be measured by considering the number of runs he has scored and the number of runs he has directly assisted his team mates in scoring. In a rugby game players can perform really well without scoring any points themselves.

Thus according to Bracewell [5]: "...the search for rugby ability must focus on unstructured rugby match participation, with the assumption that over time the nature of an individuals ability will be imposed on collated statistics" (p. 19). This assumption that the on field occurrences which constitute observed performance behaviour relate directly to the abilities and skills of the athletes involved, is the basic premise on which classic performance analysis is based [41].

This paradigm has been adopted from a psychological theory of human personality known as trait theory (see for example Pervin [46]), which focuses on the measurement of underlying personality traits which are assumed to be relatively stable over time and to influence the patterns of behaviour by which they are measured. However, Lames and McGarry [41] argue that game sports are most appropriately conceived of as a dynamic interaction process between two opponents. They suggest that behaviour in game sports emerges as a product of this interaction process and thus unlike many individual sporting disciplines, cannot be considered merely as a reflection of the underlying stable abilities of the athletes involved. If this is the

case, then performance indicators would not be expected to be reliable or stable as a result, and they thus suggest that alternative approaches to performance analysis (e.g. dynamical systems theory) be investigated.

The main form of empirical support for the paradigm suggested by Lames and McGarry is the high variation in performance indicators both within games and between games. There are clearly a few factors which cause this variability in match statistics [5]:

- the variable conditions in which different matches are played (e.g. home & away games, weather conditions etc),
- differing competition - which can affect game strategy and tactics,
- and the volatility of individual or team performance itself.

However some researchers involved in sports science maintain that the analysis of match statistics can be done in such a way that it takes into account the natural variability caused by different match constraints and conditions. The commercially developed Eagle Rating [5, 6, 7] aims to provide a relatively stable overall measure of individual rugby player performance, by considering holistic involvement in a range of different match activities over a series of matches, in order to obtain statistically sound data. This issue is addressed primarily in Bracewell [5] which investigates the impact of non-performance on statistical analyses. Non-performance refers to performance by an individual that is less than expected in a given position and can be caused by superior players being heavily marked or targeted by opponents, or due to variations in the game structure and tactics in particular games which may cause some players to see less of the ball.

However players are trained to adapt their behaviour over time in order to stay involved in the game despite these challenges - an example of this in rugby would be to make use of evasive manoeuvres like side-steps, dive passes and the chip and chase.

Thus by considering performance statistics over multiple games, and thereby allowing athletes to be evaluated in a range of different match conditions it is expected that true ability can be inferred in the long term despite variations in performance due to match conditions. The Eagle Rating aims to produce a reliable indication of an individual rugby players performance by including past performance based on an exponentially weighted moving average. In addition to taking a series of matches into consideration, the effect of match volatility on performance in particular aspects of the game can also be dampened by aggregating player involvement over a range of different activities. Sporting performance is understood to be the physical manifestation of a players ability, and this ability is in turn determined by the combination of desirable skills. In order for a physical task to be successfully completed, it is assumed that relevant physiological and mental skills must be present, and thus it is sensible to assume that by aggregating physical task measures which relate to these skills, we can at least to some extent quantify performance from which ability is inferred.

By applying dimension reduction techniques (factor analysis, self-organising maps and neural networks) to a large amount of physical task data that had been recorded from rugby matches, Bracewell et al. [7] found evidence of a lower intrinsic dimensionality which supports the existence of contextual key performance indicators for different aspects of play. This latent dimensionality of match data is (loosely) interpreted to be a manifestation of the various skills that individual players possess, and these key performance indicators are collapsed into a single overall performance measure for each player.

The Eagle Rating makes use of modified techniques based on the philosophy of statistical process control, in order to monitor match performance so that changes in form and associated strengths and weaknesses can be identified [6]. Since the rating combines a range of key performance indicators, when a significant change is identified the coach can isolate the indicators that caused the change and then re-examine relevant match footage to investigate the potential problem further. In this sense

the measure is useful as a diagnostic tool for exploring strengths and weaknesses of individuals.

In addition to being useful to coaching staff for diagnostic purposes, performance analysis is perhaps more obviously useful when it comes to the selection of players for a team or squad. Team selection in a professional environment is quite clearly of key importance for the overall performance of the team, and Agnew [1] also highlights that it has a significant financial impact on the individuals involved. When selecting a squad (and particularly when selecting a team), the coaching staff and/or selection committee have to take into account a number of different factors, but the primary factor influencing their decision is expected to be how “good” the players are (bearing in mind that this will always be to some extent a subjective measure). Additional factors such as player fitness and tactical considerations are usually invoked subsequently in order to further discern between players.

In this thesis we wish to motivate the practical use of SMAA as a potential decision aid in the selection of players for a rugby team or squad. According to Cameron [8]: “The ranking of individuals is a necessary part of the selection process for professional sports teams.” Accurate observation and evaluation of the players by coaches plays a key roll in this regard, but as highlighted in Section 2.5, it is impossible for the coach to identify and remember all key events in rugby games. For this reason performance analysts seek to provide coaches with valuable support through the use of video analysis, hopefully granting them more detailed insight into the performance of particular players in different aspects of the game. Since the players need to be compared based on a range of different performance attributes and abilities, this ranking problem can be seen to fall within the scope of MCDA.

Trninic et al. [59] suggest that due to the developing nature of modern competitive sports, players are increasingly being required to become more versatile within their preferred position, and to be able to perform in more than one position if required. This may point towards SMAA as a useful tool for player selection since

acceptability indices can provide an indication of a player's 'versatile performance' by considering their ability on a wide range of criteria.

James et al. [28] developed position specific key performance indicators and performance profiles for 10 different rugby positional clusters. They found intra-positional variability (i.e. differences within each positional cluster), and concluded that there is a need for more than one profile per playing position. This provides support for the view that there are many different playing styles within given positions in rugby, each of which can be effective for their team. This is another reason that SMAA could be both useful and suitable in the context of team selection - within the group decision making context, different preferences favouring different playing styles can be incorporated into the analysis.

Chapter 3

Simulation

3.1 Objectives

In this chapter we aim to investigate the use of rank acceptability indices from the SMAA-2 modelling paradigm to inform the selection of a viable short-list of alternatives from a larger set. In addition to the conventional ‘stochastic’ SMAA model we also introduce and compare a few simplified ‘deterministic’ SMAA models which make use of summarized measures of uncertainty instead of a fully fitted statistical distribution. By assessing the efficacy and accuracy of both conventional and simplified SMAA models under a range of different conditions we hope to provide motivation for the use of these simplified decision models in appropriate circumstances.

In the process, we also wish to test the robustness of this approach and gain insights into which factors affect results the most. In addition to varying a few simple characteristics of the hypothetical problem context itself (problem size for example), we consider to a limited extent the effect of the underlying decision maker preferences and the accuracy of available information on the performance of alternatives on different criteria.

3.2 Overview of Simulation Process

The general structure of the simulation procedure for each run is as follows:

1. Form a hypothetical problem context, generating the relevant attribute evaluations.
2. Apply an assumed multi-attribute utility theory model of preference structure in order to derive “true utility” and thus find the “true rank ordering” of all the alternatives.
3. Calculate summarized measures of uncertainty based on the generated data, incorporating an assumed observational error factor.
4. Run different SMAA models based on these summarized uncertainty measures.
5. Calculate various output measures comparing the model results with the true utilities and rank order obtained from step 2.

This process is summarized in the form of a flow diagram in figure 3.1 which outlines a single simulation run.

3.3 Generating Data for Problem Context

Consider a decision problem involving m alternatives evaluated over n criteria, from which a group of decision makers wish to extract a short-list of alternatives. We will consider a range of hypothetical scenarios for this problem based on different parameter combinations dictating:

- the problem size (number of alternatives & criteria)
- the distributional properties of the attribute evaluations on the different criteria

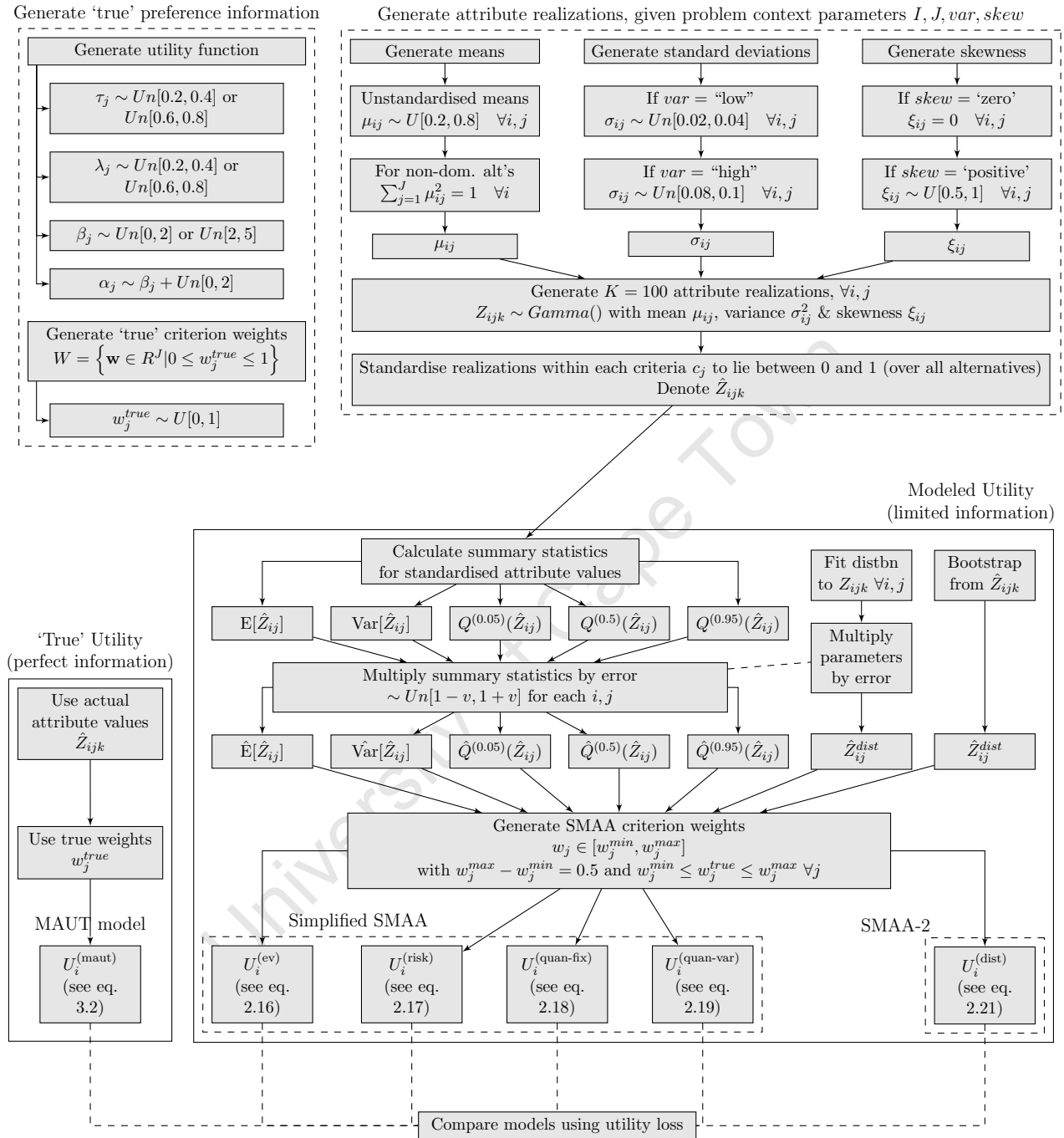


Figure 3.1: Outline of a simulation run

3.3.1 Problem Size

In order to consider decision problems of slightly different sizes, both the number of alternatives ($m = 9, 19$) and the number of criteria ($n = 10, 20$) are varied. Most typical discrete decision problems are unlikely to be larger than this. However, it is expected that any trends observed due to changes in the number of alternatives or the number of criteria for small to medium sized decision problems will apply similarly for larger decision problems.

3.3.2 Attribute Evaluations

Attribute evaluations z_{ij} (for alternative i on criteria j) are assumed to follow a Gamma distribution with mean μ_{ij} , standard deviation σ_{ij} , and skewness ξ_{ij} .

Across all the alternatives and criteria, the mean of each attribute evaluation is drawn randomly from a uniform distribution between 0.2 and 0.8 i.e. $\mu_{ij} \sim U(0.2, 0.8)$ For non-dominated alternatives, the means are then standardized across all the criteria to lie on the unit hypersphere:

$$\sum_{j=1}^n \mu_{ij}^2 = 1 \quad \forall i$$

The standard deviation of the attribute evaluations is allowed to take on either ‘low’ or ‘high’ values generated randomly from a Uniform distribution as follows:

$$\text{low: } \sigma_{ij} \sim U(0.02, 0.04) \quad \text{high: } \sigma_{ij} \sim U(0.08, 0.1) \quad \forall i, j$$

The skewness of the attribute evaluations is considered to be either zero (in which case the Gamma distribution is equivalent to a Gaussian distribution), or a random positive value drawn from a Uniform distribution:

$$\text{symmetric: } \xi_{ij} = 0 \quad \text{positively skew: } \xi_{ij} \sim U(0.5, 1) \quad \forall i, j$$

For each iteration of the simulation run, a set of $nZ = 100$ realizations of these attribute evaluations are generated for each alternative on each criterion using the above parameters (μ_{ij} , σ_{ij} and ξ_{ij}). These attribute evaluations are then standardized across all alternatives and realizations to have a minimum of 0 and a maximum of 1 on each of the n criteria:

$$\min(z_{ijk}) = 0 \text{ and } \max(z_{ijk}) = 1 \quad \forall j$$

3.4 Preference Modelling

We assume that a multi-attribute utility theory (MAUT) model can be used to reflect the underlying preference structure of the group of decision makers with sufficient accuracy. We make use of the basic additive form below, with U_i denoting the overall utility of alternative i . This is adapted and extended for use with the different SMAA models outlined in Section 2.4:

$$U_i = \sum_{j=1}^n w_j u_j(x_{ij}) \quad (3.1)$$

Here x_{ij} is some measure of the performance of alternative i on criterion j , with w_j a weight indicating the importance of criterion j . The form of the partial utility function $u_j(\cdot)$ will be discussed in Section 3.4.1 which follows.

To calculate a ‘true’ measure of utility, the MAUT model is applied and averaged over all the ‘true’ standardized attribute values \hat{Z}_{ijk} :

$$U_i^{maut} = \frac{1}{nZ} \sum_{k=1}^{nZ} \sum_{j=1}^n w_j u_j(\hat{Z}_{ijk}) \quad (3.2)$$

This true measure of utility is based on perfect and complete information, and for the purposes of this simulation study, provides a utility scale on which results from all of the different models can be compared.

Clearly alternatives with higher utility are considered ‘better’ so that the true best alternative is that for which ‘true’ utility is at a maximum i.e. $U_{best}^{true} = \max(U_i^{true})$. Similarly, the true worst alternative is that with the minimum utility, $U_{worst}^{true} = \min(U_i^{true})$, and a complete ‘true’ ranking is specified by ordering all m alternatives in this manner.

The different models outlined in Section 2.4 are all based on incomplete and/or imperfect information, and are expected to result in different utility measures, and thus in different rankings¹ for the alternatives. One common way to analyse the accuracy of a decision support model in the context of utility theory is a simple measure called *utility loss* [3, 14].

The formula for *utility loss* resulting from the selection of alternative i (as the ‘model best’) is:

$$\text{utility loss}_i = \frac{U_{best}^{true} - U_i^{true}}{U_{best}^{true} - U_{worst}^{true}} \quad (3.3)$$

Utility loss will be used as the primary measure for model accuracy in the results of this simulation study, but in conjunction with a few other model output measures which are described in Section 3.5.

3.4.1 Utility Function

It is assumed that each marginal utility function is concave above some reference level τ_j , convex below it, and that the slope and curvature is greater below the reference level than above it. This is in line with general prospect theory and peoples’ tendency for loss aversion as highlighted by Kahneman and Tversky [30].

¹Note again that there are many different ways to compare decision models depending on the specific aims and priorities of the decision context. Here the focus is on the ranking of alternatives for the purpose of selecting the best alternative or a short-list of alternatives containing the best alternative i.e. the ranking problematique

The following standardised exponential form is used:

$$u_j(x) = \left\{ \begin{array}{ll} \frac{\lambda_j(e^{\alpha_j x} - 1)}{e^{\alpha_j \tau_j} - 1} & \text{for } 0 < x \leq \tau_j \\ \lambda_j + \frac{(1 - \lambda_j)(1 - e^{-\beta_j(x - \tau_j)})}{1 - e^{-\beta_j(1 - \tau_j)}} & \text{for } \tau_j < x \leq 1 \end{array} \right\} \quad (3.4)$$

The same approach is used by Stewart [52] and Durbach and Stewart [14], with

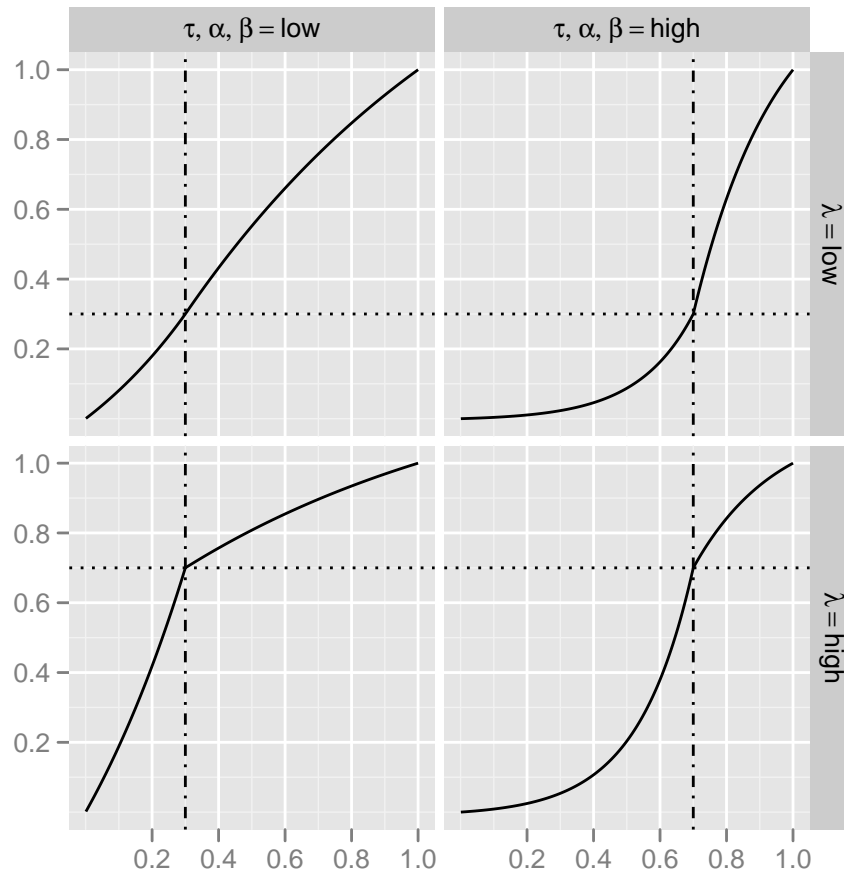


Figure 3.2: Graphs of utility function (3.4) showing the effect of different values for τ , λ , α & β

$\alpha_j > \beta_j > 0$ so that the properties assumed above are satisfied.

We will include τ , λ , and β as simulation parameters with either ‘low’ or ‘high’ values drawn from a uniform distribution as outlined below:

τ_j	reference level for u_j	low: $U[0.2, 0.4]$	high: $U[0.6, 0.8]$
λ_j	value of u_j at reference level	low: $U[0.2, 0.4]$	high: $U[0.6, 0.8]$
β_j	curvature of u_j above τ_j	low: $U[0, 2]$	high: $U[2, 5]$

The value of α is then fixed relative to the value of β : $\alpha_j = \beta_j + U[0, 2]$

The parameters dictating problem size (m and n), distribution of the performance attributes (σ_{ij} and ξ_{ij}) and assessment error (v) are all as outlined in the previous sections. Table 3.1 provides a summary of all the simulation parameters and the values they can take on.

3.4.2 Inter-Criterion Weights

As highlighted in Section 2.3, the feasible weights are usually considered to be non-negative and normalized to sum to 1. However, Lahdelma and Salminen [37] highlight that weight vector normalization can also be done in other ways or omitted entirely. The generation of normalized and uniformly distributed weights in an unrestricted weight space (or even with certain simple restrictions) is relatively straightforward (e.g. as outlined in 2.4.4), but it turns out that in the more general case, generating normalized, uniformly distributed weight vectors in a restricted weight space is far more complex and can be highly demanding from a computational perspective. A common approach in the case of single SMAA (or MCDA) problems is to use rejection based sampling in order to generate uniform weights within the restricted space (which can then be arbitrarily specified). This approach would be very time consuming in a large scale simulation such as ours.

Parameter	Description	Values
Problem Context:		
m	number of alternatives	9; 19
n	number of criteria	10; 20
σ_{ij}	std. dev. of attribute evaluations	$U(0.02, 0.04)$; $U(0.08, 0.1)$
ξ_{ij}	skewness of attribute evaluations	0; $U(0.5, 1)$
Errors in assessments of attributes and attribute uncertainty:		
v	interval width for assessment error	0; 0.1; 0.2
Marginal utility functions:		
τ_j	reference level for u_j	$U[0.2, 0.4]$; $U[0.6, 0.8]$
λ_j	value of u_j at reference level	$U[0.2, 0.4]$; $U[0.6, 0.8]$
β_j	curvature of u_j above reference level	$U[0, 2]$; $U[2, 5]$
α_j	curvature of u_j below reference level	$\beta_j + U[0, 2]$
SMAA modelling:		
k	ranks used in acceptability index	1; 3; 5
w^{range}	interval for weight generation	0.5 (see Sec. 4.1)

Table 3.1: Summary of all simulation parameters used

The issue of efficient weight generation in restricted space in higher dimensions has been explored (and addressed to some extent) by means of an implementation of the hit-and-run algorithm which can be applied in order to sample uniformly from a subset of the n -simplex defined by linear constraints. Tervonen et al. [58] have demonstrated the transformation to $n - 1$ dimensions which allows for more efficient Monte Carlo Markov Chain (MCMC) sampling and have illustrated that with sufficient thinning an acceptable deviation from a uniform distribution over the constrained weight space can be obtained. However, this approach is still considerably slower than conventional methods for sampling preference weights (e.g. those suggested by Tervonen and Lahdelma [57]) and given that this adds significantly to the computational complexity of what is already a lengthy and demanding simulation study, it could not be implemented here.

We have instead opted to omit the normalization of weight vectors, utilizing uniformly distributed weights in n -dimensional space in place of the conventional uniformly distributed weights on the $n - 1$ simplex. It is acknowledged that this approach is less than ideal due to the divergence from the conventional scaling and distribution of preference weights, but since the ensuing analysis will focus primarily on utility loss (which is an entirely relative measure) in order to compare results, the interpretability of the utility measurements in an absolute sense is not important.

Thus we consider the weights:

$$W = \{w \in R^n | 0 \leq w_j \leq 1\} \quad (3.5)$$

The ‘true’ criterion weights w_j^{true} are generated randomly for each of the criteria from a uniform distribution on the interval $[0, 1]$. To emulate the inclusion of partial preference information in the SMAA-2 model for this simulation, the weight interval

for each criterion weight is restricted as follows:

$$\{w_j \in [w_j^{min}, w_j^{max}] \text{ with } w_j^{max} - w_j^{min} = w^{range} \forall j\} \quad (3.6)$$

Clearly the restricted weight interval must contain the true criterion weight i.e. $0 \leq w_j^{min} \leq w_j^{true} \leq w_j^{max} \leq 1$. Furthermore, the intervals are constructed so as to be as symmetric around the true criterion weights as is possible given the global minimum (0) and maximum (1) constraints. For example, given a weight range restriction of $w^{range} = 0.7$ and a true criterion weight of $w_j^{true} = 0.3$, the restricted interval would then be set as $w_j \in [0, 0.7]$ so that the true weight is as close as possible to the middle of the restricted weight interval.

In simulation studies such as this it is usually necessary to include only some of the variables of interest as simulation parameters and to hold the remaining variables constant in order to limit the size of the study. This decreases total computation time and keeps the overall focus of the study within reasonable limits.

Since the SMAA paradigm is generally intended for use in situations where a limited amount of preference information is available, the inter-criterion weight range restriction (as outlined in Section 3.4.2) will be fixed to allow for comparison of the different SMAA models under a (fixed) partial preference information assumption. A suitable restriction of this weight range is determined by means of an investigation into the effect of inter-criterion weight range restrictions on the different SMAA models which is described in Section 4.1. Intra-criterion preference information is still varied by means of the simulation parameters τ_j , λ_j , α_j and β_j which determine the shape of utility function for each of the criteria. For the purposes of this simulation study the shape parameters of these marginal utility functions are assumed to be known.

Model Name	Details	Error (v)		
		none	small	significant
True model				
MAUT	Full MAUT model	not applicable		
Conventional stochastic SMAA model				
SMAA ^{dist}	Fitted distribution model	n/a	0	0.1
Simplified uncertainty models				
SMAA ^{ev}	Expected Value model	0	0.1	0.2
SMAA ^{risk}	Risk model (EV & Variance)	0	0.1	0.2
SMAA ^{quan-fix}	Fixed Quantile model	0	0.1	0.2
SMAA ^{quan-var}	Variables Quantile model	0	0.1	0.2

Table 3.2: Models used in simulation

3.5 Output Measures

It is assumed that the primary form of decision support required is some ranking of the alternatives in order to inform a short-list of ‘good’ options. To this end, the rank acceptability index b_i^r will form the core tool of SMAA to be utilized in the calculation of various output measures which can then be used to compare the performance of the different SMAA models.

The rank acceptability indices described in 2.3.2 can be used to inform estimates on the ‘true’ ranking of the alternatives. To obtain rank estimates we simply rank the alternatives based on the k -best ranks acceptability indices a_i^k resulting from each SMAA model. We will then refer to this estimated ranking of the alternatives as the “ k -best model rank”.

In general it is expected that for higher values of k the k -best model rank will provide more robust measures of overall suitability for the alternatives, while a more accurate indication of the true best alternatives will be provided with smaller values of k . Currently we are concerned primarily with choosing or shortlisting the ‘best’ alternatives so will consider the k -best model rank for $k=1,3$ & 5 . This means that for each simulation run of each model type, three different rankings of the alternatives are calculated based on the different rank acceptability indices.

As highlighted in Section 3.4, utility loss (UL) is a well known and widely used metric for ranking and choosing problems in the context of MAUT and will be used in the results as the primary indication of model accuracy. Utility loss can range between 0 (when a model results in the selection of the true best alternative) and 1 (when a model results in the selection of the worst possible alternative). In theory, a randomly selected alternative will result in utility loss of 0.5 on average.

Chapter 4

Simulation Results

In the results sections which follow we will first look at a preliminary investigation into the effect of weight range restrictions on model accuracy, before delving into a broader exploration of the results returned over a variety of different simulation parameters. The primary focus of these simulation results is on the evaluation of model accuracy resulting from the different simplified models of uncertainty by comparison to the conventional ‘distributional’ SMAA approach. In addition we will analyse the effects of changes in assessment error, problem size, attribute distribution, and shape of utility function. Lastly we look briefly at the impact of different choices of k for the k -best ranks acceptability indices.

4.1 Weight Range Restrictions

We begin by exploring the effect of weight restrictions on the average accuracy of the SMAA models. All of the different SMAA models were run while varying the range of allowed intra-criterion importance weights in order to first try and establish the kind of weight restrictions that are necessary to give reasonable performance, on average, for each of the different models. In addition, we also wish to investigate the possibility of an interaction effect between model type and weight range (i.e. to check whether certain model types are much better with more accurate preference weights).

The weight range interval was varied between 0 (perfect weights) and 1 (unrestricted weights) in intervals of 0.1, with 120 runs of each model type on each weight range interval. The number of starting realizations generated for the hypothetical problem context was fixed at $nZ = 10000$ (so in this case the empirical distribution of criterion attributes is expected to match the theoretical Gamma distribution more closely than in the subsequent results sections where $nZ = 100$). All models were implemented with no assessment errors. Within each weight range, two distinct cases were also considered by utilizing ‘all low’ and ‘all high’ values on the remaining simulation parameters. The ‘low’ and ‘high’ values for the simulation parameters are as defined in Table 3.1 for the rest of the simulation study.

A factorial-ANOVA model was fitted, to test for possible interaction effects between the three factors: model type, weight range interval, and case (other simulation parameters).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	5	6.30	1.26	44.75	0.0000
Weight Range	10	217.03	21.70	771.23	0.0000
Case	1	16.29	16.29	578.84	0.0000
Model: Weight Range	50	0.62	0.01	0.44	0.9998
Model: Case	5	2.32	0.46	16.48	0.0000
Weight Range: Case	10	3.64	0.36	12.94	0.0000
Residuals	19718	554.87	0.03		

Table 4.1: Factorial-ANOVA results showing first and second order effects of model type, weight range, and case on utility loss

As is to be expected, the individual ‘main’ effects of weight range interval, model type, and case are all highly significant, and have thus been correctly identified as factors of interest. Average utility loss was observed to range roughly between 0 and 0.4, with lower utility losses associated with tighter weight restrictions as expected. Figure 4.1 illustrates this relationship, showing the smoothed average utility loss by

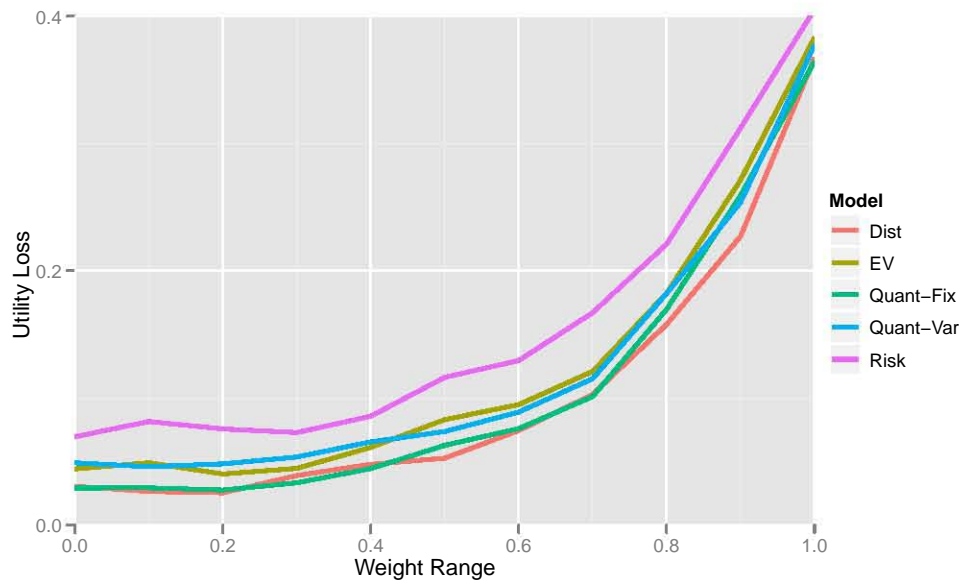


Figure 4.1: Mean utility loss by model for different weight restrictions

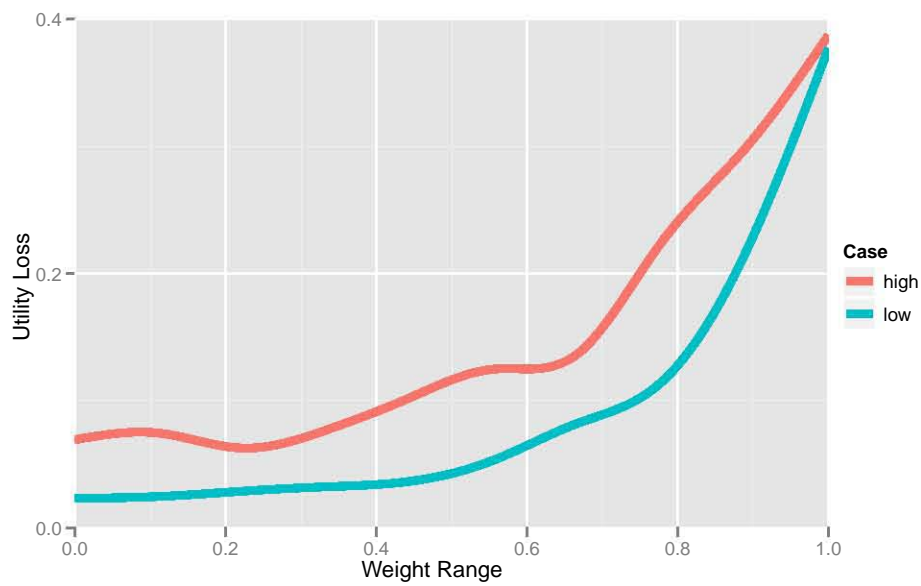


Figure 4.2: Mean utility loss by case for different weight restrictions

model over the different weight range restrictions. With little or no weight restrictions, the resulting average utility loss for most of the models is in the region of 0.35 to 0.4. Use of the acceptability indices with no weight restrictions is thus essentially only a little bit better than a random guess, and it would be difficult to compare the models with each other when all of them provide very poor results. When the weights are restricted to 80% of the full range however, average utility loss quickly drops below 20% for most of the SMAA models, and at 50% of the full range average utility loss is under 0.1 for most of the models.

The key result shown in Table 4.1 is that there appears to be no significant interaction (p-value = 0.9998) between the type of model used and the weight range interval. If one accepts this result, that means it is then acceptable to fix the weight range restriction for all models in order to ensure a suitable level of accuracy, while continuing to explore in more detail the effect that other parameters (here combined into a simple ‘two-case’ factor) can have on the different model types.

Since there doesn’t seem to be an interaction effect between the weight range restrictions and the different model types, the weight range restriction will be fixed at $w^{range} = 0.5$ for all the models run in the simulation study. Based on figure 4.1, this weight range should be narrow enough to provide sufficient accuracy for the simulation results (as primarily indicated by utility loss) while requiring only a limited amount of preference information.

It is expected that the results in the following sections will be indicative of the various SMAA models behaviour over a wider range of weight restrictions (i.e. varying degrees of preference information), but extrapolation of the results is only loosely supported without a more detailed analysis on the effect that different weight restrictions can have when varying the ‘case’ parameters more extensively.

The ANOVA results in Table 4.1 suggest that there is an interaction effect between weight range and case (F-statistic: 12.94), but this effect is less significant than the others and may be due to over-fitting. Figure 4.2 shows the effect of changes in the

weight range interval on the two parameter case factor used here. Since any interaction effect present between these two factors could have a bearing on the validity of results in the following sections, this must be noted as a possible limitation of the study, and a potential area for future research. The above results also suggest that there is a significant interaction effect present between the type of model and the case (which represents factors such as variance, shape of utility function etc) - these important factors will be explored in more detail in Sections 4.4 and 4.5. Table 4.4 has also been included for reference since it highlights statistical significance for a number of first and second order interaction effects across a more detailed selection of different simulation parameters based on factorial-ANOVA.

4.2 Overview of Simulation Results

With the weight range interval fixed at $w^{range} = 0.5$, we now consider results from SMAA simulations run over the entire combination of parameters outlined in Table 3.1. With 16 parameter combinations for problem context, 3 for assessment error, and 8 for shape of utility function there are a total of 384 simulation parameter combinations. Running 100 simulations for each of the 5 model types on each of these 384 parameter combinations results in 192000 runs in total.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	4	104.04	26.01	992.48	0.0000
Error	2	188.53	94.26	3596.83	0.0000
Alternatives	1	0.00	0.00	0.18	0.6706
Criteria	1	7.14	7.14	272.37	0.0000
Variance	1	52.02	52.02	1985.07	0.0000
Skewness	1	0.41	0.41	15.50	0.0001
τ	1	107.34	107.34	4095.67	0.0000
λ	1	9.68	9.68	369.52	0.0000
β	1	2.22	2.22	84.65	0.0000
Residuals	191986	5031.44	0.03		

Table 4.2: Summary of ANOVA results showing the main effects of all simulation parameters

Table 4.2 shows that almost all the simulation parameters under consideration have a significant effect on the utility loss resulting from the SMAA models. Based on these results, of all the factors considered only the effect of the number of alternatives on utility loss is not significant. However it must be noted that there is very high power due to the large sample size, so the usual measures of statistical significance are somewhat over-inflated. In order to investigate the extent to which these different factors influence model accuracy we will thus focus largely on effect size rather than looking at measures of significance.

Analysis on the effects of different decision parameters in this simulation study is broken down into three areas in the sections which follow: model type and error, problem size and attribute distribution, and utility function. Table 4.3 highlights the factors under consideration in each section and reflects the primary effect of an

¹Here an ‘increase’ in the model parameter is associated with a change from the conventional stochastic model to the worst of the simplified models of uncertainty.

²Since the error parameter can take on three different values, the increase here reflects a change from no error to medium error.

Section	Parameter \uparrow	Utility Loss
4.3 Model Type & Error	Model ¹	\uparrow
	Error ²	\uparrow
4.4 Size & Distribution	Criteria	\uparrow
	Variance	\uparrow
	Skewness	\downarrow
4.5 Utility function	τ	\uparrow
	λ	\downarrow
	β	\uparrow

Table 4.3: Table showing the direction of the average change in utility loss when each of the simulation parameters is increased

‘increase’ in each of the simulation parameters, showing the direction of the change in utility loss, averaged over all levels of the remaining parameters.

4.3 Comparison of Different Uncertainty Models under Error

As stated at the start of this chapter, the predominant focus here is on evaluating the performance of the different model types under consideration in the present study, by comparing them under a range of different conditions. Since the choice of model type is usually the only factor under the control of either the decision maker or analyst, any interaction effects between model type and the other decision parameters are also of interest. These interaction effects with model type will be considered in Sections 4.4 and 4.5.

We begin here with a comparison of all the models at different levels of assessment error, with performance aggregated over all the remaining simulation parameters (i.e. main effects). Table 4.5 provides some summary statistics on utility loss which are

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	4	104.04	26.01	1011.90	0.0000
Error	2	188.53	94.26	3667.22	0.0000
Criteria	1	7.14	7.14	277.70	0.0000
Variance	1	52.02	52.02	2023.92	0.0000
Skewness	1	0.41	0.41	15.80	0.0001
τ	1	107.34	107.34	4175.83	0.0000
λ	1	9.68	9.68	376.75	0.0000
β	1	2.22	2.22	86.30	0.0000
Model:Error	8	2.18	0.27	10.61	0.0000
Model:Criteria	4	1.63	0.41	15.87	0.0000
Model:Variance	4	8.72	2.18	84.86	0.0000
Model:Skewness	4	0.92	0.23	8.92	0.0000
Model: τ	4	7.90	1.97	76.82	0.0000
Model: λ	4	2.22	0.55	21.54	0.0000
Model: β	4	0.74	0.18	7.19	0.0000
Error:Criteria	2	0.30	0.15	5.81	0.0030
Error:Variance	2	7.11	3.55	138.28	0.0000
Error:Skewness	2	0.65	0.33	12.71	0.0000
Error: τ	2	20.90	10.45	406.46	0.0000
Error: λ	2	5.41	2.71	105.25	0.0000
Error: β	2	0.06	0.03	1.26	0.2827
Criteria:Variance	1	0.06	0.06	2.39	0.1219
Criteria:Skewness	1	0.04	0.04	1.70	0.1917
Criteria: τ	1	0.02	0.02	0.64	0.4239
Criteria: λ	1	0.20	0.20	7.62	0.0058
Criteria: β	1	0.11	0.11	4.38	0.0364
Variance:Skewness	1	0.65	0.65	25.38	0.0000
Variance: τ	1	0.05	0.05	1.94	0.1641
Variance: λ	1	0.03	0.03	1.19	0.2761
Variance: β	1	0.84	0.84	32.50	0.0000
Skewness: τ	1	0.98	0.98	37.94	0.0000
Skewness: λ	1	0.01	0.01	0.51	0.4746
Skewness: β	1	0.66	0.66	25.59	0.0000
τ : λ	1	30.34	30.34	1180.20	0.0000
τ : β	1	5.30	5.30	206.34	0.0000
λ : β	1	0.05	0.05	1.97	0.1610
Residuals	191928	4933.37	0.03		

Table 4.4: Factorial-ANOVA results showing first and second order interaction effects for the different simulation parameters

also reflected in graphical form in figures 4.3 and 4.4 on page 77.

	Error	Model	Mean	St. Dev.	Median	Upp. Quartile
1		Dist	0.06	0.11	0.00	0.07
2		Quan-Fix	0.06	0.11	0.00	0.08
3	none	Quan-Var	0.07	0.12	0.00	0.10
4		EV	0.08	0.14	0.00	0.12
5		Risk	0.12	0.17	0.02	0.19
6		Dist	0.09	0.14	0.00	0.14
7		Quan-Fix	0.09	0.14	0.00	0.13
8	small	Quan-Var	0.10	0.15	0.00	0.16
9		EV	0.12	0.18	0.03	0.20
10		Risk	0.15	0.19	0.07	0.25
11		Dist	0.13	0.18	0.04	0.21
12		Quan-Fix	0.13	0.18	0.04	0.21
13	med	Quan-Var	0.14	0.19	0.05	0.23
14		EV	0.17	0.21	0.10	0.29
15		Risk	0.19	0.22	0.13	0.32

Table 4.5: Summary statistics of utility loss by model type for different assessment errors

As one would expect, the conventional stochastic model (Dist) performs best overall and has the lowest mean utility loss since it makes use of full distributional representations of uncertainty on the criteria attributes. On average the quantile models also perform very well with mean utility loss only marginally higher than for the SMAA-Dist model. The fixed-weight quantile model outperforms the variable-weight quantile model by a small margin - probably due to the fact that it provides a more stable and consistent measure of the attributes' central location. The expected value (which ignores uncertainty altogether) ranks fourth in terms of accuracy and the risk model is clearly significantly worse than all the other models. These results are consistent with earlier assessments of model performance outside of the SMAA context such as those of Durbach and Stewart [15] who investigated the performance

of complete and simplified models in the context of ‘full-information’ MAUT.

Naturally as the size of the assessment error increases the performance of all the decision models is negatively effected. The same performance ranking of the models is preserved across differing levels of assessment error, with the only exception being that the fixed-quantile model has a slightly lower mean utility than the dist-model under ‘small’ error. However, it must be noted that the level of error is not directly comparable across the different model types due to the way these errors have had to be incorporated into the simulation structure. For example one cannot directly compare a 10% assessment error on the quantiles with a 10% error on the expected values. Rather than comparing models and errors in absolute terms it is hoped that the error factor can be used to assess the relative robustness of different models.

It is notable in the results that for many of the SMAA models, when assessment error is sufficiently small the median utility loss is zero. This means that under these conditions, there is a 50% probability on average that the model will result in zero utility loss (or equivalently that the model will indicate the true best alternative). Note that in these cases with both the lower quartile and median equal to zero, only the upper quartile is distinctly visible on the box plots in figure 4.4.

4.4 Changes in Problem Size and Attribute Distribution

As highlighted earlier, the number of alternatives appears to have no effect on model accuracy. Since the allocation of performance over the alternatives is deliberately random (uniform), on average one wouldn’t expect this distribution to change simply by increasing the number of alternatives (i.e. the probability of picking a good or bad alternative at random shouldn’t be affected by changing the number of alternatives).

The number of criteria on the other hand, clearly influences the accuracy of these

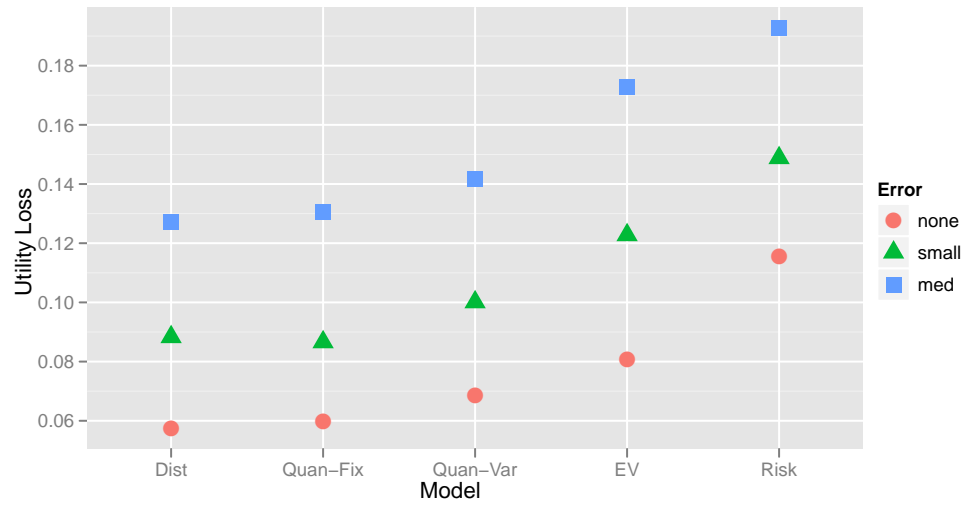


Figure 4.3: Plot of mean utility loss for different models and assessment errors

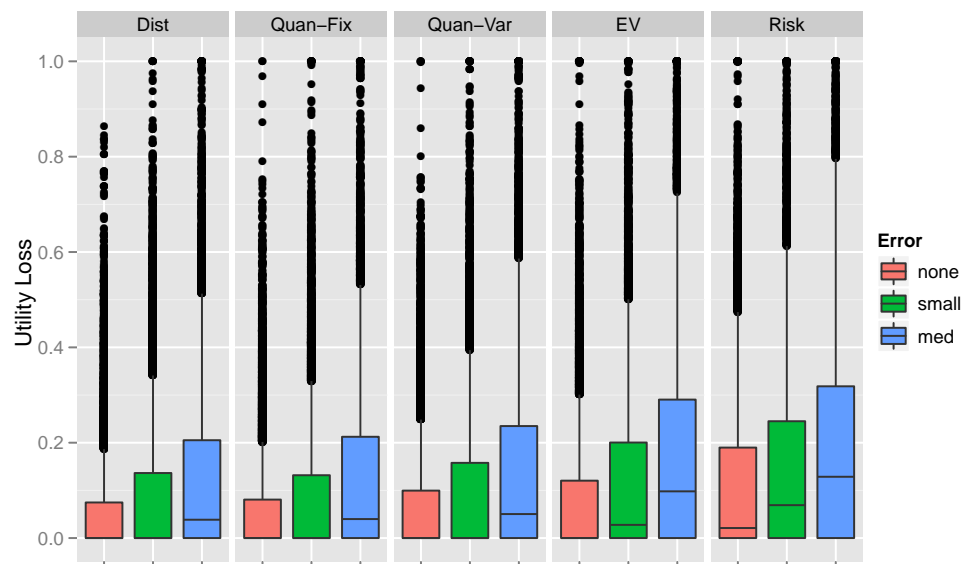


Figure 4.4: Box-plots of utility loss for the different models and assessment errors

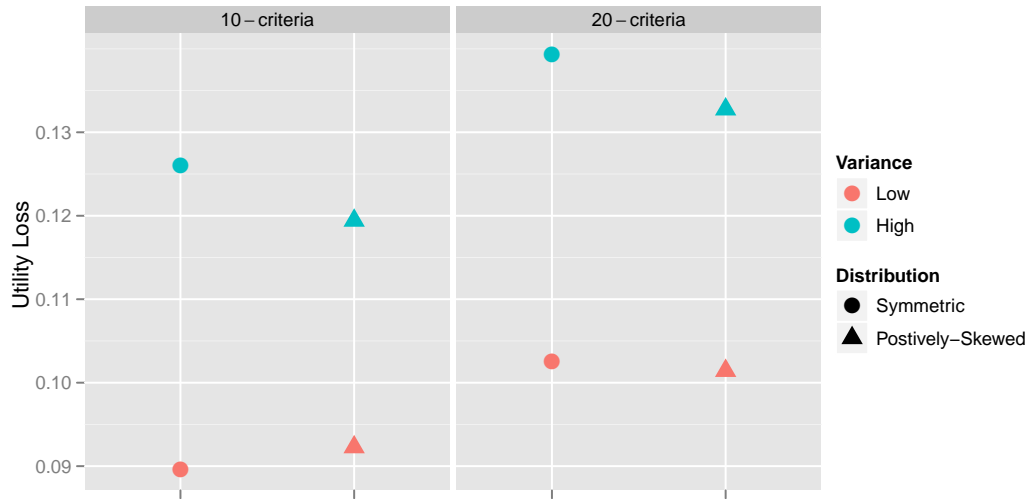


Figure 4.5: Mean utility loss broken down according to the variance and skewness of the performance attributes, with 10 and 20 criteria

models. This is illustrated in figure 4.5, by the higher mean utility losses on the right hand side (20 criteria). An increase in the number of decision criteria is thus associated with an increase in utility loss. This is in line with what intuition would suggest: that it is harder to make the right decision when you have more criteria to take into account.

The distribution of performance on each of the criteria is naturally also seen to be of importance in determining the resulting accuracy of this kind of utility model. High variability in the distribution has a strong negative effect on model accuracy. Utility loss is observed to increase with variance consistently across all the model types, and at all examined levels of error. This is not surprising, given that high variance is an indication of inconsistency (volatility) in the relevant criterion performance levels. High levels of fluctuation on the criteria make it more difficult to accurately assess the overall ('aggregate') performance of an alternative.

The general effect of introducing skewness into the distribution of criterion performance is a slight reduction in utility loss. However there is also a noteworthy interaction effect between the skewness of the distribution and the amount of variance. As evident in figure 4.5, the reduction in (mean) utility loss is clearly larger in the case of high variance. In the case of low variance, the introduction of skewness results only in a small decrease in utility loss (20 criteria cases) or even in a slight increase (10 criteria cases).

Figure 4.6 breaks down the effect of attribute variance and skewness even further by showing the mean utilities for each model type and at each error level. Probably the most striking interaction effect is that between variance and model type: at low variance, use of the fixed quantile model (Quan-Fix) results in lower utility loss than for the conventional full stochastic model (Dist), while at high variance, the stochastic model remains the best.

This effect is also more pronounced when there is skewness in the distribution and/or when error is introduced. With low variance and medium error both the fixed and variable quantile models outperform the stochastic model. When variance is low and there is positive skewness, the fixed and variable quantile models outperform the stochastic model at both small and medium error levels.

Another seemingly evident interaction effect is present between skewness and model type: the quantile models both improve in accuracy relative to the stochastic model when skewness is introduced.

4.5 Changes in Utility Function

As outlined in the previous chapter, the marginal utility functions follow an exponential form, with four parameters controlling their shape. τ is a reference level, λ is the value of the utility function at this reference level, and α and β determine the curvature of the utility function below and above the reference level respectively. Since we

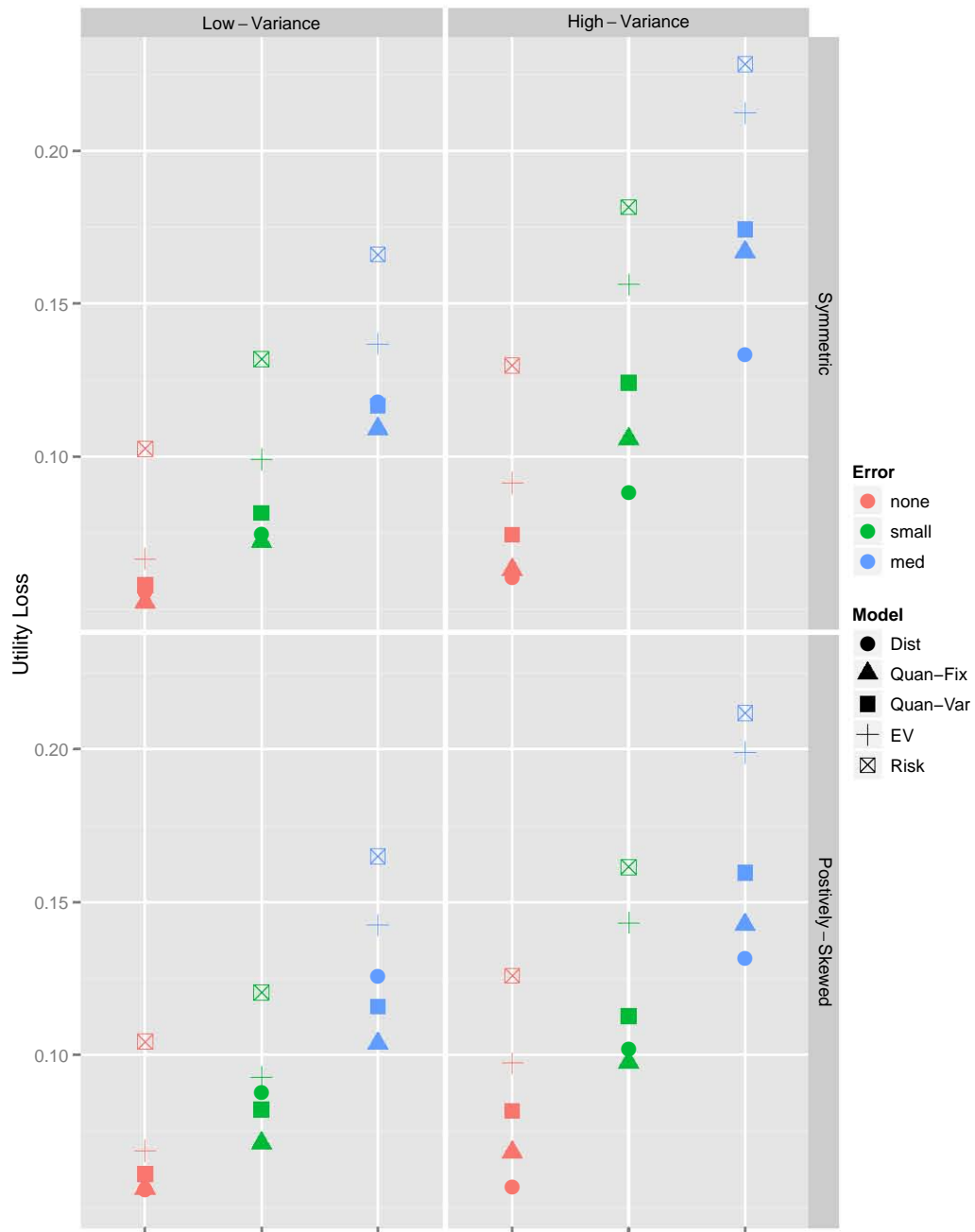


Figure 4.6: Mean utility loss broken down according to attribute variance and skewness for the different models and assessment errors

have fixed the value of α relative to the value of β , here we need only consider β as a simulation parameter, and can essentially ignore α . The utility function parameters work very much in tandem with each other, and thus need to be considered together in order for their effect to be understood (to see this refer back to figure 3.2).

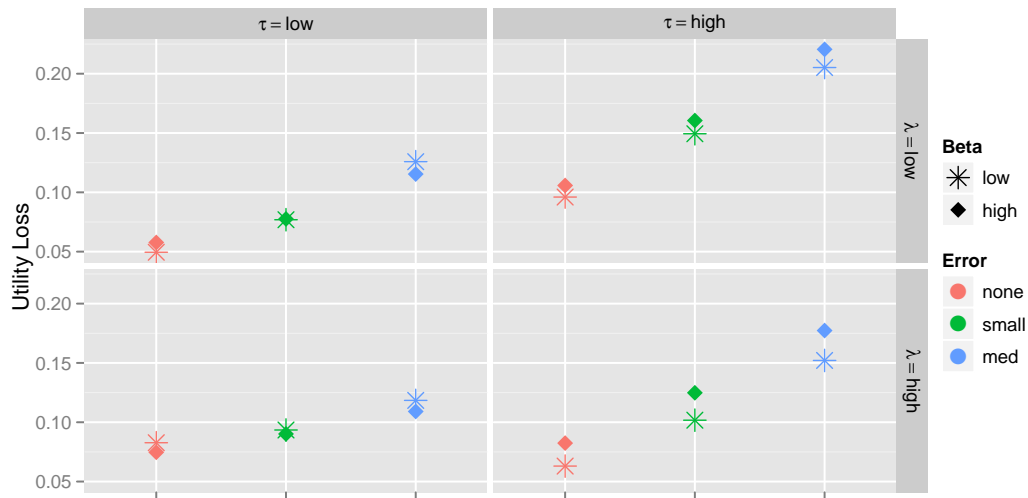


Figure 4.7: Mean utility loss for different levels of τ , λ and β , at the different error levels

The main effect associated with the reference level (τ) is that a higher reference level results in higher average utility loss. A possible explanation for this is that in determining the best alternative, the distinction between utility levels at high reference levels are most influential (i.e ordering the top performing alternatives for each criterion is likely more influential in determining the overall best performing alternative than distinguishing the exact ordering of the poorer performing alternatives on each criterion). The negative effect of erroneous assessment is also more pronounced when τ is higher (as evident in figure 4.7 by the steeper ‘gradient’ across the different error levels on the right hand side).

When considering the effect of the shape of utility, generally it is observed that

deviations from linear utility are what hampers the accuracy of all the models most. So when τ and λ are both low we observe lower mean utility loss than when τ is low and λ is high, since the latter is a deviation from linearity. Similarly when τ and λ are both high we observe lower mean utility loss than when τ is high and λ is low.

The strongest interaction effect present is between τ and β : for low levels of τ , high levels of β generally result in smaller average utility loss than low levels of β , but for high levels of τ an increase in β has the opposite effect. This effect is generally also consistent across the different error levels - the only exception is for the no error case when both τ and λ are low.

When considering the different models performance under varying utility, it is evident that when the reference level (τ) is low, and value at that reference level (λ) is high, the quantile models outperform the conventional stochastic model slightly, both with or without error. Otherwise the previously observed hierarchy of model performance is generally preserved across the other utility parameter combinations.

4.6 k -Best Ranks Acceptability Indices

In all the previous sections, utility loss has been calculated based on use of the first rank acceptability index ($b_i^1 = a_i$) to determine the ‘best’ alternative according to the model. We now briefly examine the effect of incorporating information from some of the rank acceptability indices by varying the value of k .

It can be seen in figure 4.9 that increasing the value of k can in some cases result in a slight improvement in utility loss. For the more accurate models (e.g. Dist & Quan-Var) and particularly at low levels of error, the mean utility loss is generally higher when $k = 3$ or 5 than when $k = 1$. However, in the cases where there is higher utility loss (≥ 0.1), the mean utility loss is slightly lower for $k = 3$ than for $k = 1$. This effect is evident for all the model types, and for all levels of error, provided that

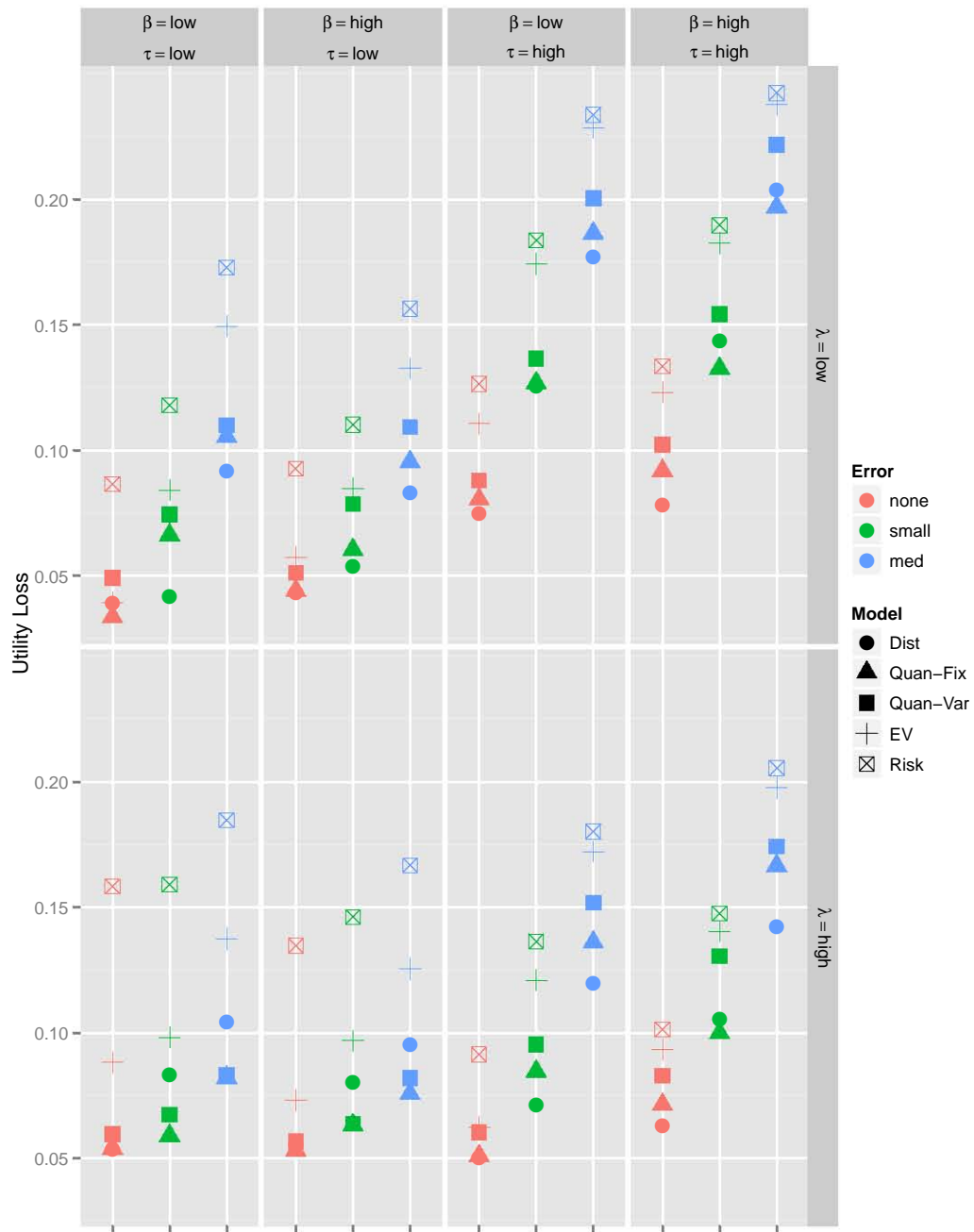


Figure 4.8: Mean utility loss broken down according to the three utility function parameters, for the different models and assessment errors

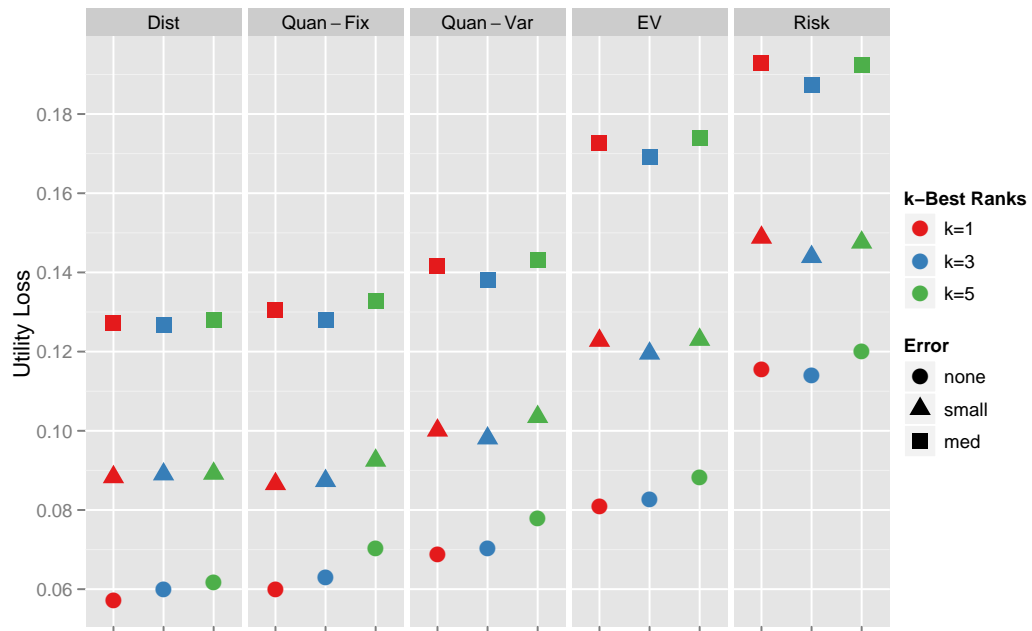


Figure 4.9: Mean utility loss at different error levels when using different values of k for the k -best ranks acceptability indices

mean utility loss is high enough. In some extreme cases (e.g. risk model with small or medium error), mean utility loss is slightly lower for both $k = 3$ and $k = 5$.

Thus the results suggest that for the poorer performing models, or when there is some error, utilizing the 3-best ranks acceptability indices will produce a slight improvement on overall utility loss. Given that increasing the value of k further to 5 generally results again in an increase in mean utility loss, it seems that the choice of value for k is essentially a trade-off between accuracy and robustness. Increasing the value of k improves robustness to error and is thus beneficial only when there is significant inbuilt error, either in the model itself or in the inputs to the model. Since only 3 values of k have been considered here, the conclusions that can be drawn are limited. Future research should thus investigate this relationship further by considering additional values of k (e.g. $k = 2$) and exploring whether there may be particular factors influencing what the optimal choice of k should be in different situations.

4.7 Discussion

This simulation experiment illustrates a number of simplified SMAA models for dealing with problem contexts in which decision makers are unable or unwilling to assess trade-off information precisely. The primary focus is to assess how closely a rank order of alternatives based on partial information and SMAA can approximate results obtained using full-information MAUT, both for the conventional (distributional) SMAA model and for the simplified SMAA models. In addition we explore how some characteristics of the decision problem can influence the accuracy of this approximation.

The results highlight that the conventional SMAA model with fairly limited preference information can return good approximations to the full MAUT model provided that uncertainty is assessed with little or no errors. If no weight information is provided, then the average accuracy of any SMAA model (conventional or simplified)

tends to be only slightly better than a random guess (where the average utility loss is 0.5). However, the simulation results suggest that by restricting the specified weight range for the SMAA models to cover less than 70% of the full range of possible outcomes the accuracy can be improved significantly. With weights restricted this way, application of the conventional SMAA model resulted in average utility loss just less than 0.1.

Another key finding is that it is not essential to make use of a full probability distribution to represent attribute uncertainty when implementing the SMAA model. Good results can be obtained by using simplified uncertainty formats, and the quantile model in particular appears to be the most accurate of these simplified SMAA models. The risk model is clearly the worst of the simplified SMAA models and should thus not be used, but the expected value model doesn't perform too badly given that it is the simplest of all models considered here and is thus the easiest and quickest in terms of implementation. Given that it can take considerably more time and effort to implement a full MAUT model, the use of a simplified SMAA model appears to be justifiable for decision problems where preference information is not easily obtainable or time is constrained.

With no assessment error, almost all of the models (excluding the risk model) resulted in a median utility loss very close to zero. Average utility loss was found to increase significantly when relatively small assessment errors were introduced into any of the SMAA models. This means that even the worst of the SMAA models when applied without assessment error provides better results than the best of the SMAA models when these have been applied with small errors in the assessment of criteria attributes. For this reason, it would appear that avoiding assessment errors is more important than the choice of model for a particular problem.

Though an increase in the number of alternatives can affect the manageability of a decision problem from a cognitive point of view, this appears to have little or no effect on model accuracy within the context of the simulation study. However, an

increase in the number of criteria under consideration is detrimental to model accuracy, both for the conventional and simplified SMAA models.

The distribution of performance on the criteria attributes is a key area of influence when it comes to the selection of a suitable SMAA model. High variance results in significantly higher utility loss across all the model types and at all examined levels of error, and is generally seen to have a much stronger negative effect on all of the simplified uncertainty models. Positive skewness in the distribution decreases utility loss across most of the models, and appears to counteract the negative affect of high variance in particular. It is interesting to note that while the conventional SMAA model generally gives the best accuracy, it is outperformed by the quantile SMAA model (particularly the fixed-weight quantile model) in a number of the scenarios with low variance criteria distributions, and especially when there is also skewness in the distribution and/or assessment error is introduced.

The three parameters determining the shape of utility function in the study were all found to be influential in terms of model accuracy, and the general finding is that for all models utility loss deteriorates with any significant deviations from linear utility. This means that for a given reference level (τ), if the value at this reference level (λ) is the same then utility loss is observed to be much lower than when the value at this reference level is much higher or lower. Since the best performing alternative overall is more likely to be strongly influenced by the shape of utility towards the top end of the performance spectrum, deviations from linearity at higher reference levels (e.g.. higher curvature β) are observed to have a more severe negative impact on utility loss. An alternative view on acceptability was also considered by utilizing the k -best ranks acceptability indices for $k = 3$ and $k = 5$ instead of just looking at the acceptability index (i.e. $k = 1$). In cases when the resulting utility loss is higher - either due to inbuilt model error (the simplified models) or assessment error - it was found that incorporating information from the 2nd and 3rd rank acceptability indices in order to chose the ‘best’ performing alternative actually improved accuracy slightly on average. Intuition suggests that this might be a trade-off between between accuracy and robustness. We suppose that the additional information obtained by

looking at additional rank acceptability indices results in a more ‘robust’ view of the ‘preferred alternatives’, which is most useful in situations where model accuracy is compromised.

Chapter 5

Decision Support For The Evaluation of Rugby Players

The previous chapter dealt with the use of SMAA in an artificial, simulated context. In this context ‘environmental’ parameters can be adjusted to facilitate controlled direct comparisons which are not possible in real world experiments. These simulation experiments are however obviously limited in their ability to address and explain the practical real-world implementation of decision support methods. This chapter will report on the application of stochastic multi-criteria acceptability analysis (SMAA) with different uncertainty formats in the analysis of rugby player performance. We aim to demonstrate how SMAA can be utilized to incorporate statistical match data into a versatile analysis of player performance which can then be used to aid coaches with player analysis and squad selection.

This chapter will begin with a discussion of the data and then detail the structuring of the problem in accordance with the theory outlined in Section 2.1.2 before describing the results of the analysis.

5.1 Discussion of Data

The analysis will draw on data taken from the 2008 and 2009 Super Rugby tournaments. Super Rugby is the pre-eminent professional Rugby Union competition in the Southern Hemisphere, with matches broadcast in many countries all over the world. Originally known as the Super 12, Super Rugby officially started in 1996 with 12 teams from Australia, New Zealand and South Africa, but has since grown to include 15 teams - 5 teams from each of the three countries. The format has changed over time with the addition of new teams, but incorporates both home and away games against local and international opposition as the teams travel between the three countries during the course of the tournament.

During the 2008 and 2009 seasons of Super Rugby, there were 14 teams participating in the tournament - four Australian teams, five New Zealand teams, and five South African teams. The data was captured by *Fair Play Pty Ltd - Sports Analysis Systems - Australia*. Using both live and recorded match footage, a detailed log of events from each game are coded into a database. Data can then be viewed and extracted from the (proprietary) database using *The Rugby Analyst* software [53]. This data does have some qualitative characteristics but is primarily quantitative and includes a large variety of ‘on-the-ball’ activities for players involved in rucks, mauls, kicks, tackles, passes etc.

In it’s ‘raw format’, data for each match consists of a list of transactions which contains 6000 – 8000 timestamped events, each coded by type and by player, along with some additional details as well. Over 400 event types are specified, ranging from very rare events such as red cards to very frequent events like tackles. However, at a player level probably less than half of these are meaningful since many of these ‘events’ are simply markers to provide context for the transactions list and others are duplicates of each other since some events are recorded once from the perspective of each of the teams. The full dataset for the 2008 and 2009 Super 14 included 7 & 11 pre-season warm-up games respectively, as well as match data for the 94 tournament

games played in each year.

5.2 Problem Structuring

5.2.1 Research Objectives

The following research objectives are identified:

- to illustrate the suitability of the SMAA methodology with different incorporated representations of uncertainty as a decision support tool for performance analysis and player selection
- to identify difficulties with the practical implementation of the SMAA methodology with different representations of uncertainty in this context
- to assess whether decision makers find the decision support provided by the SMAA approach to be useful in general and in particular which incorporated uncertainty formats might be preferred

5.2.2 Stakeholder Involvement

With the aim of informing player selection through the analysis of player performance, the primary stakeholders are the relevant rugby coaching staff and any others involved in the selection process. The project was originally commissioned in 2008 by an analyst working for the Stormers, a South African team participating in the then Super 14, on behalf of their coaching staff. While the real stakeholders for the problem are the coaching staff and management of the Stormers, our only contact was with the analyst, who provided insights into the data as well as possible performance criteria. The analyst left the Stormers in 2010, before the project could be completed. Thus while we have obtained some feedback on our results; none has been obtained from the Stormers. However in the final stages of the research the individual was presented with a summary of the modelling process and some of the

results and his feedback was positive regarding the insights this approach offers and its potential for further use.

5.2.3 Grouping of Alternatives

In choosing a team of rugby players, it is understood that from a squad of players, a specific player is chosen for each position, usually taking into account a range of factors which go beyond a player's ability and specialities. These factors can include a player's history of combinations with other players in nearby positions on the team or how a player might aid in specific strategies and game plans for particular matches. This is beyond the scope of our application here. We will focus instead on aiding in the selection of a rugby squad.

Squad selection generally does not involve the strategic, tactical or game theory type dynamics of team selection, and is usually focused more simply on assembling the 'best group' of players based on their ability. The process of building a rugby squad involves short-listing a few players in turn from each of a range of positional profiles in order to build a talented pool of players from which to select teams for each game during a season. In order to do this, one needs to have some means of evaluating and comparing different players.

When evaluating the performance of a rugby player, it is only sensible to compare that player to other players who play in a similar position. Attempting to directly compare the performance of a big strong forward with a small and speedy back is futile since these players each contribute different skills and abilities to different aspects of the game. Comparing an inside and outside centre (12 & 13) or two flanks (6 & 7) is far more reasonable though, since these players usually have more similar job roles within the team despite playing in slightly different positions.

Positional 'Profile'/Group	Numbers
Props	1 & 3
Hookers	2
Locks	4 & 5
Loose Forwards	6,7 & 8
Half-Backs	9 & 10
Centres	12 & 13
Outside-Backs	11,14 & 15

Table 5.1: Positional clusters for players

The fifteen positions which make up a team of rugby players all have unique characteristics (some more than others), but like those mentioned above, many of them also have a number of characteristics or 'job functions' in common and can thus be grouped together based on these similarities. To avoid having to consider fifteen separate groups of players for all the positions, we have opted to make use of such positional groupings. Table 5.1 shows which positions (denoted by the number worn by the player who plays in the position) have been grouped to form the seven positional profiles adopted for use here. Using each of these groups, we will focus on the selection of a small subset of 'best performing' players from a larger group of players who fit that positional profile.

5.2.4 Constructing Attributes

In keeping with the approach of grouping players by position, a variety of attributes were constructed on which to measure the performance of players in different aspects of the game based on the position they play in. The choice and construction of all of these attributes was dictated to a large extent by the available data.

Given that many basic rugby skills and disciplines apply to players in all positions,

Attribute	Description	Positions
Running		
Adv. Line	Count of ‘makes advantage line’	All
Adv. Line%	Percent success for ‘makes advantage line’	All
Carries	Count of ball carries - incl. open & closed runs, pick & drive etc	All
Tackles Evaded	Count of tackles evaded	6-15
Breaks	Count of all half-breaks & line-breaks	9-15
Kicking		
Kick%	Percent of successful ‘territory’ kicks (e.g. find touch, ball regained, kick to space etc)	9-15
Conversion%	Percent of successful goal kicks - including conversions, penalties & drop goal attempts	9-15
Discipline		
Cards	Count of yellow and/or red cards issued	All
Frees Against	Count of all free kicks & penalties against	All
Frees For	Count of all free kicks & penalties for	All
Handling Errors	Count of all knock-ons & handling errors	All

Table 5.2: Performance attributes for player positions (part 1)

a number of ‘core’ attributes relating to some of these behaviours were specified for use in all positional groups. Attributes pertaining to other more specialized performance areas were utilized only for certain positional groups. This assignment of different attributes to positions is based both on prior understanding of the game of rugby and the different ‘job roles’ of different positions as well as on the frequency of occurrence of different types of involvement by players in these different positions.

In Tables 5.2 and 5.3, each of the attributes is listed along with a basic description as well as which positions the attribute applies to. The attributes are arranged into the categories running, kicking, discipline, general play, breakdown involvement and special/other. Two general types of metric are used in all of these areas:

1. a success rate or efficiency measure of some kind, expressed as a percentage and thus denoted by a % symbol at the end of the attribute name
2. a simple frequency measure counting the number of occurrences of an event during a match (all other attributes which are not of the first type)

In many cases, both of these metrics are utilized on the same variables, since they convey different types of information about a players ability in an area. In the general case, the success or efficiency measure tends to indicate the quality of execution of a task whereas the frequency measure provides an indication of the work-rate of a player in executing that task. Both of these are important since a low score in either is indicative of poor performance. For example a player who attempts many tackles all over the field but hardly ever succeeds in stopping his opponents is not considered effective. On the other hand, being able to stop every opponent he tries to tackle is not worth much if he hardly ever attempts any tackles. In some cases, the frequency measure is a count of the number of successful executions of a task (e.g. the advantage line attribute), in which case it can be considered as a combined measure of success/effectiveness and work-rate.

Attribute	Description	Positions
General Play		
Passes	Count of all passes and offloads made	All
Pass%	Percent of successful passes (excl. offloads)	All
Tackles	Count of all tackle involvements (initiating or assisting with a tackle)	All
Tackle%	Percent of successful tackles	All
EffGenPlay	Count of all effective contributions in general play (defence & attack)	All
Breakdown Involvement		
Mauls	Weighted sum of all maul arrivals	1-8
Maul%	Percent of efficient maul involvements	1-8
Rucks	Weighted sum of all ruck arrivals	1-8
Ruck%	Percent of efficient ruck involvements	1-8
Turnovers For	Count of successful turnovers & pilfers	All
Turnovers Against	Count of turnovers & pilfers against	All
Special/Other		
Line-out%	Percent of own line-outs won	2,4-8
Try Create	Count of all try creating movements (kick, pass, break etc) plus tries scored	All

Table 5.3: Performance attributes for player positions (part 2)

5.2.5 Data Cleaning and Transformation

In order to ensure that players are evaluated fairly the analysis should make use of sufficient data to provide a reasonably balanced assessment of each player's ability. Since many players spend a lot of time on the bench and have limited game time, the available data was trimmed down to ignore match data for players who; started on the bench, spent less than 10 minutes of the game on the field, and/or played in less than 8 games in total across the 2008 & 2009 super rugby seasons.

The trimmed data set contained game data for 265 different players in total, who played between 8 and 28 games each (avg. 16.3); 39 props, 17 hookers, 37 locks, 49 loose-forwards, 39 half-backs, 36 centres, and 49 outside backs. Since their involvement in these games ranged from a minimum of 10 minutes to the full 80+ minutes and is therefore not directly comparable, all frequency data must be re-scaled on a per match basis for each player based on the length of time they spent on the field.

We make use of the following transformation as suggested by [28]:

$$F^{new} = F\left(\frac{80}{n}\right)\left[\log_{10}\frac{80}{n} + 1\right] \quad (5.1)$$

Here n is the number of minutes a player spends on the field and F denotes the frequency of a particular action performed by the player. As shown in figure 5.1, by contrast to a simple average rate transformation (e.g. multiply event frequencies by 8 for a player with 10 min of game time), this log transformation scales frequencies slightly higher for players with a small amount of game time, in an effort to compensate for potential imbalance in game-play during the time they were on the field.

For example, an attacking wing who spends 10 minutes on the field while his team is one player short (due to a yellow card) and forced to defend against constant attacks from the opposition might not see much of the ball and therefore likely won't have much opportunity to prove himself with ball carries and breaks. By contrast, he may come on and shine in a defensive role, but the transformation is chosen to

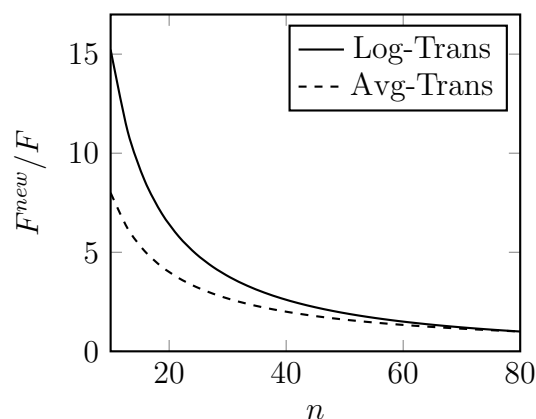


Figure 5.1: Log transformation used to scale frequency of events for players with less than 80 min of match time

depict these players in the best light, allowing players with limited game time an opportunity to measure up competitively against those who play for much longer periods.

5.3 Performance evaluation models

5.3.1 Treatment of uncertain attribute evaluations

As mentioned above, to measure a player's performance on a particular attribute is to measure an inherently uncertain quantity – performance varies from game to game. A common response is to evaluate players based on average performance over the course of a fixed time period such as a season (e.g. median performance is used by James et al. [28]). Such an assessment strategy, however, effectively ignores uncertainty in the evaluations. For example, after averaging performance it is impossible to distinguish a player who averages 10 tackles per game from one who makes no tackles in half his games and 20 tackles per game in the other half. We conjecture that in at least some cases the variability of performance constitutes an attribute of interest.

Because information is collected for each game, we can assess players both in terms of their average performance over the course of a season and in terms of the variability of this performance. In order to assess the contribution made by uncertainty, we will construct three different evaluations for each player:

1. average performance model
2. distributional performance model
3. quantile performance model

In the first model, average performances are calculated over all the games played in the season (in our case we're combining data from both 2008 and 2009 and treating this as one season). At each iteration of the SMAA simulation the same performance evaluations are drawn, so that it is only the weights associated with the different performance criteria that differ at each iteration. The resulting preference model is defined as:

$$U_i^{(\text{ev})} = \sum_{j=1}^J w_j u_j(\bar{z}_{ij}) \quad (5.2)$$

where \bar{z}_{ij} denotes the mean performance for the season.

The second model considers the full distribution of performances over the course of the season. It treats these performances as an empirical (discrete) probability distribution, and at each SMAA iteration samples directly from these values i.e. both the performance values and criterion weights differ. The resulting model is given by

$$U_i^{(\text{dist})} = \sum_{j=1}^J w_j u_j(z_{ij}^*) \quad (5.3)$$

where z_{ij}^* is an attribute value drawn from a simple random sample of the z_{ijk} .

The third model makes use of quantiles as a simplified representation of the distribution of each players performance over the season. Three quantiles (10%, 50%

and 90%) are used, and the weights x_r are constrained to sum to 1.

$$U_i^{(\text{quan})} = \sum_{j=1}^n w_j \sum_{r=1}^3 x_r u_j(\hat{q}_r[\hat{z}_{ij}]) \tag{5.4}$$

where $\hat{q}_1 = 10\%$ quantile, $\hat{q}_2 = 50\%$ quantile, $\hat{q}_3 = 90\%$ quantile

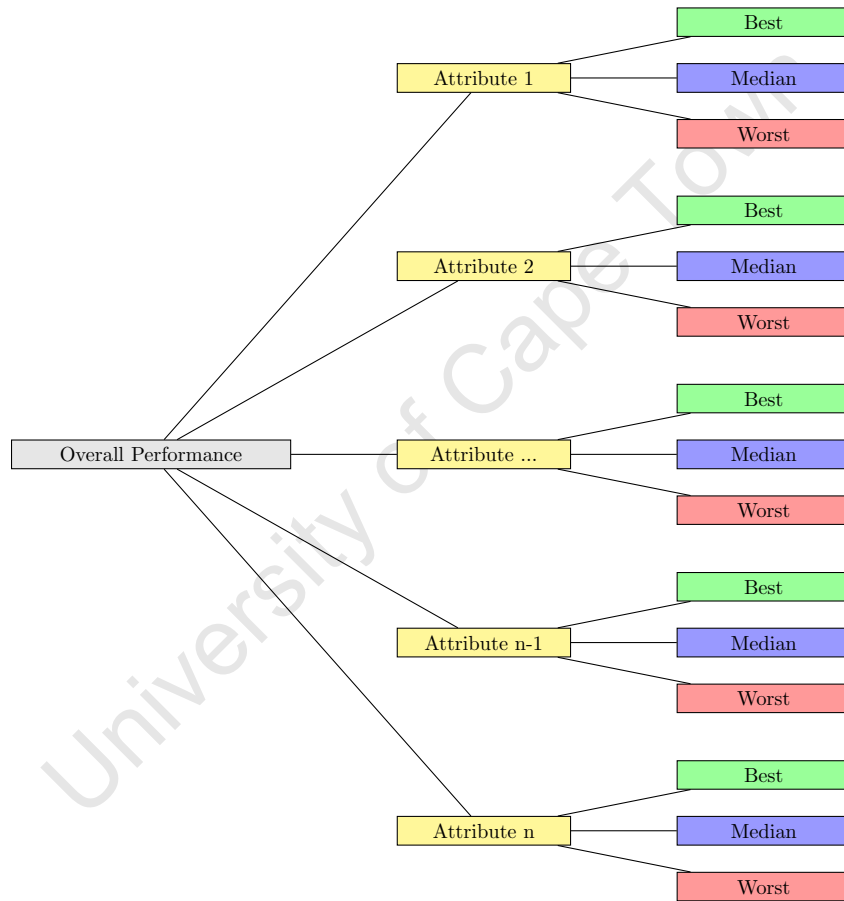


Figure 5.2: Value Tree for Quantile Model of Player Ability

5.3.2 Implementation details

For each position, our input data is game-specific performance evaluations captured over the criteria defined in Table 5.3. The performance evaluations on each criterion are first standardized so that their minimum over all players is zero and their maximum one. This is done to ensure commensurability of the simulated attribute importance weights.

We then apply our three SMAA models. The “average performance” model uses pre-calculated mean performances for each player on each attribute, the “distributional” model uses the full range of game-specific evaluations directly, and the quantile model uses three pre-calculated quantiles for each player on each attribute. At each iteration of the SMAA model, three quantities are randomly simulated:

1. Performance evaluations. In the case of the average and quantile performance models, the same evaluations will be used in all iterations¹.
2. Attribute weights. Weights are generated to be uniformly distributed and to sum to one over all attributes, following the approach used in [64, 12, 13].
3. Utility functions. Performance evaluations are transformed into utilities using a simple exponential utility function which implies either constant or diminishing marginal returns to some degree. A single parameter β controls the amount of curvature in the utility function, and is randomly generated as part of the simulation to lie between 0 (linear) and 4 (fairly rapidly diminishing returns). Some possible utility curves are sketched in Figure 5.3.

We employ 20 000 simulations per position, twice the number suggested by Tervonen and Lahdelma [57]. Once the full set of simulations is completed, we compute the acceptability indices and central weight vectors using numerical approximations of equations 2.9 and 2.10 respectively. These results were plotted in R and presented to stakeholders in an interactive workshop and as part of a final report.

¹Since an additive (positive) utility model is being used, all criteria which indicate mistakes (e.g. handling errors & penalties against) were first transformed to negatives before being rescaled on the positive axis i.e. inverted so that a high score on a negative criterion results in low utility

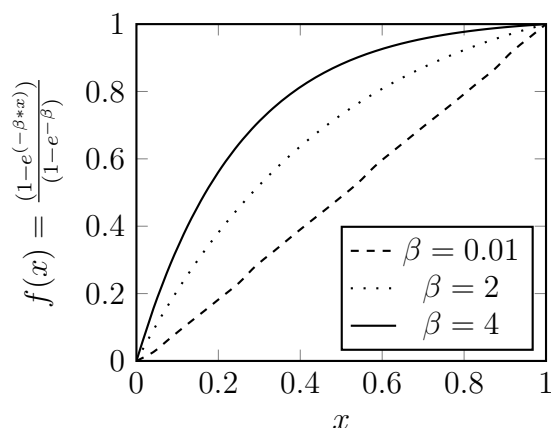


Figure 5.3: Possible simulated utility functions

5.4 Results

As highlighted earlier, SMAA results in two primary descriptive measures: rank acceptability indices which can be used to highlight a particular player's versatility within their 'positional profile', and central weight vectors which indicate the areas in which a player is inferior or superior to other players. Looking at results from all three performance SMAA models we will now analyze these measures in order to compare some players within each of two position groups; centres and loose forwards. Since many of the younger players present in the analysis are still playing currently it is worth re-emphasizing that these results are specifically based on 2008 and 2009 data and most if-not-all of the players who are still active will have evolved and improved since then.

5.4.1 Centers

Acceptability results

Using the rank acceptability indices b_i^r resulting from (2.9), we calculate for each player the total acceptability for the top 3 ranks (i.e. $\sum_{r=1}^3 b_i^r$). This indicates what proportion of all preferences would result in a particular player being ranked in the

top 3 for that position and is thus expected to be high only for very good players with solid all round performance (i.e. notable versatility).

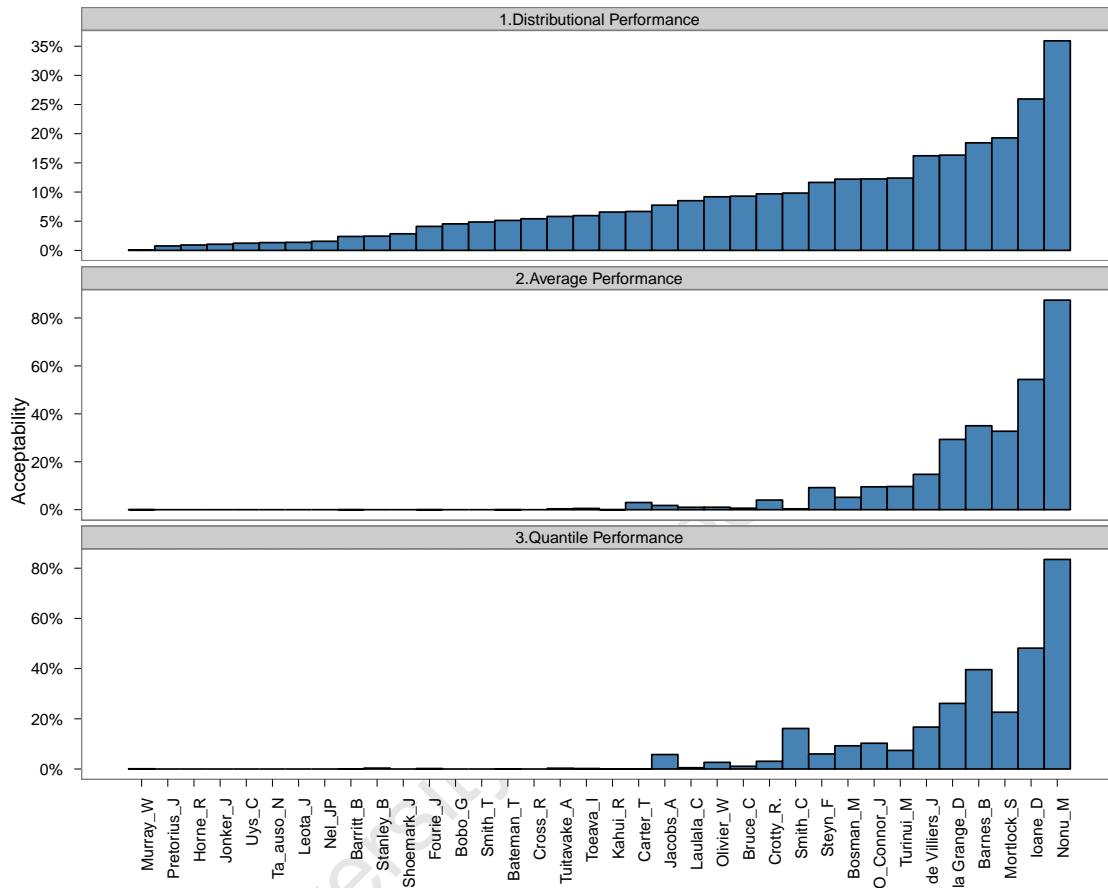


Figure 5.4: Rank acceptability indices for the centres based on the empirical and average performance models

Figure 5.4 shows the 36 centres included in the analysis ordered along the horizontal axis by the sum of their top 3 rank acceptability indices. Loosely speaking, these acceptability results provide a rough ‘objective ranking’ of the players based on their performance on each of the criteria under consideration.

It is worth noting that while the results of the three performance models are gen-

erally agreeable in terms of the overall ‘ranking’ of players, there is a much quicker drop-off in acceptabilities for the average performance model. Since it does not take performance uncertainty into account, in the SMAA context it tends to exaggerate the ability of players with higher overall average performance and understates the ability of players with lower overall average performance. Another way of considering this difference is that the average performance model is overconfident in its assessment of differences between the players’ ability since it ignores fluctuation in performance from game to game. Although to a slightly lesser extent, the quantile performance model also exhibits a quicker drop-off in acceptability than the distributional performance model since it summarizes uncertainty.

Overall the results suggest Nonu, Ioane, Mortlock and Barnes are examples of very well rounded players with above average performance and ability in most aspects of the game. Or to put it slightly differently, they are more likely to be preferred over other centres by coaches. By contrast, Pretorius and Murray are unlikely to be preferred by many coaches (if all possible preference inputs are equally likely for any given coach).

Strictly speaking though, players like Pretorius and Murray are not necessarily the worst since they have non-zero rank 1 acceptabilities. The fact that they score low acceptability in the top ranks means that there are only a small variety of preference inputs that support each of them over and above the other players. But it is of course possible that small variety of preference inputs could represent precisely the balance of traits that a particular coach wants in a player, in which case they could be considered the best.

Since no particular constraints have been placed on the preference inputs to the model, this is a very general overview and would differ based on any additional information which could be utilized to restrict the preference space in some way. For example, if we could say definitively that all coaches think running criteria are at least twice as important as breakdown criteria for a center, that would restrict

the allowed combination of weights and influence both the acceptability and central weight vector output of the model.

Interpreting the central weight vectors: player profiles

While the acceptability index provides an indication of how likely it is that a particular player will be considered ‘best’, the central weight vector highlights the average preferences under which a player is assigned rank one (i.e. it identifies the strong and weak criteria areas for a player). Note that although we have considered acceptability for the top 3 ranks, the central weight vector pertains strictly to the rank 1 acceptability index. Players with a high level of acceptability tend to have very flat central weight vectors which indicates that they score on or above average on almost all criteria under which they are being considered. Since it is very rare for a player to be better than other players on all criteria, a flat central weight vector generally indicates that a player is versatile (rather than particularly good in a few areas of the game and average or weaker elsewhere).

An above average weight on a particular criterion indicates above average performance on that criterion and similarly below average weights reflect weaker performance on the relevant criteria. 19 criteria were utilized for the center position, so the average criterion weight for a flat central weight vector is just over 5%. Since it is not feasible to display central weight vectors for all 36 centers for all models, figure 5.5 shows central weight vectors for just 5 players in the center position for the purpose of illustration.

Nonu, Ioane, Mortlock and Barnes are the four players with the highest acceptability while O’Connor is slightly further down. As with the acceptabilities shown in the previous section, the results of the different performance models seem to agree in many respects, with similar sets and strengths and weaknesses emphasized for each player. Again though, it is clear that the average performance model ‘exaggerates’ the strong and weak criteria of each player by comparison to the empirical performance model (note that in the figures the scales are different for the models).

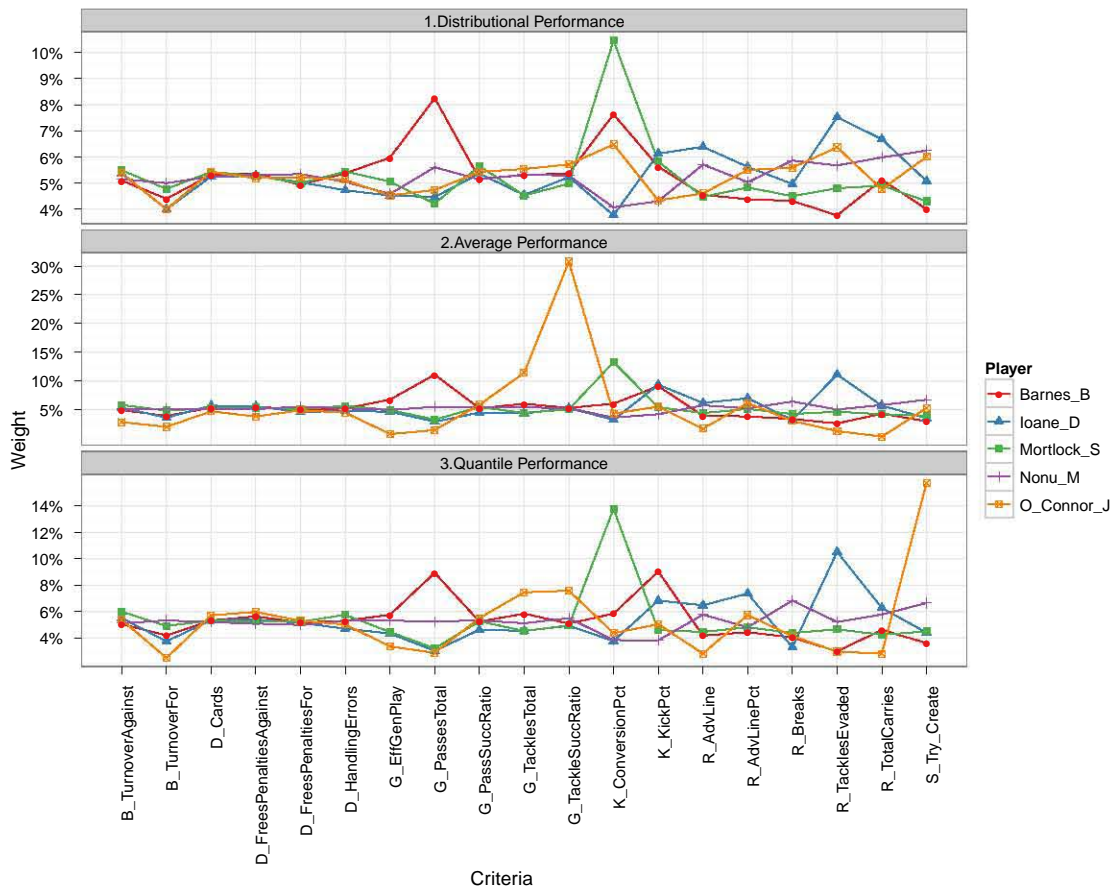


Figure 5.5: Central weight vectors for some of the centres based on the empirical and average performance models

There are clear differences in the extent to which the models support the strength or weakness of a player’s performance on some criteria. This is because the average performance model ignores uncertainty so will reward above average performance on a particular criteria even if this performance is highly inconsistent. By contrast, the distributional and quantile models will take into account the consistency of performance (to different extents).

Looking at the central weights highlights that Nonu seems to be the most versa-

tile and solid center all round. He has the flattest central weight vector, with the results suggesting the only kink in his armour might be his kicking ability. His central weight vector is skewed slightly towards running criteria like carries, breaks, advantage line and try creation which re-asserts us of the fact that he is a powerful running center.

Ioane is also fairly versatile and shows a somewhat similar profile to Nonu, but with slightly more focus on the running variables. The central weight assigned to the breaks criteria is relatively average, but his ability to evade tackles is clearly emphasized across all three models, with the advantage line and total carries criteria also consistently above the average weights. In addition, his kicking in general play does appear to be an asset. By contrast to Ioane and Nonu, Mortlock and Barnes are centers with the ability to kick consistently (both for conversions and in general play). Thus they obtain higher weights on the kicking variables (especially conversion), seemingly at the cost of lower weights on most of the running variables.

The high weight assigned to the conversion kicks criteria for Mortlock in all models points to the fact that he is a prolific kicker, not only in terms of his overall average performance but also when considering the full spread of his good and bad kicking results. The central weights for Barnes also highlight what looks to be an interesting trade-off: weaker performance in the area of running criteria and particularly tackles evaded in favour of much higher pass counts i.e. either he doesn't often evade tackles because he chooses to pass the ball instead of hanging onto it, or he chooses to pass the ball more often because he isn't usually able to evade tackles.

The central weights for O'Connor resulting from the three models are quite different, with a much bigger spread of weights under the average and quantile performance models. Under the empirical performance model the results suggest that his best assets are his ability to kick conversions and to create tries (supported further by strong ability on the tackle evasion and breaks criteria). His tackle count and passing success are also strong. However under the average performance model his tackling

ability (tackle count and success) is emphasized to a much greater extent along with passing success, with everything else weaker on an overall basis. The quantile model also supports these three criteria pretty strongly, but emphasizes his try creation substantially more, due to his high median performance on this criterion. Since O'Connor was still very young and inexperienced in 2008 and 2009, a possible explanation for the difference in results under the three models might be that as a younger player he was very committed and passionate (as suggested by his tackling) but not yet very consistent in terms of his match performance.

5.4.2 Loose forwards

Acceptability results

The same measure of acceptability is utilized here as was described in section 5.4.1. The 49 loose forwards are again ordered along the horizontal axis by their (top 3 ranks) acceptability based on the empirical performance model. McCaw, Tuiali, and Palu clearly have the highest acceptability, while Higginbotham, Soakai and Grobelaar have the lowest acceptability. As was noted with the centers, the acceptability results of the distributional and average performance models are generally agreeable in terms of the overall 'ranking' of the loose forwards but with a few discrepancies (e.g. Waldrom and Setepthano obtain higher acceptability than Messam and Hoiles under the average performance model). Again there is also a much quicker drop-off in acceptabilities for the average performance model which assigns very high acceptabilities to the top few players. This again supports the assertion that in the context of SMAA a model which does not take performance uncertainty into account will exaggerate the ability of players with higher overall average performance and understate the ability of players with lower overall average performance.

Interpreting the central weight vectors: player profiles

With 21 criteria utilized for evaluating the group of loose forwards, an average criterion weight for a flat central weight vector is just under 5%. As the player with

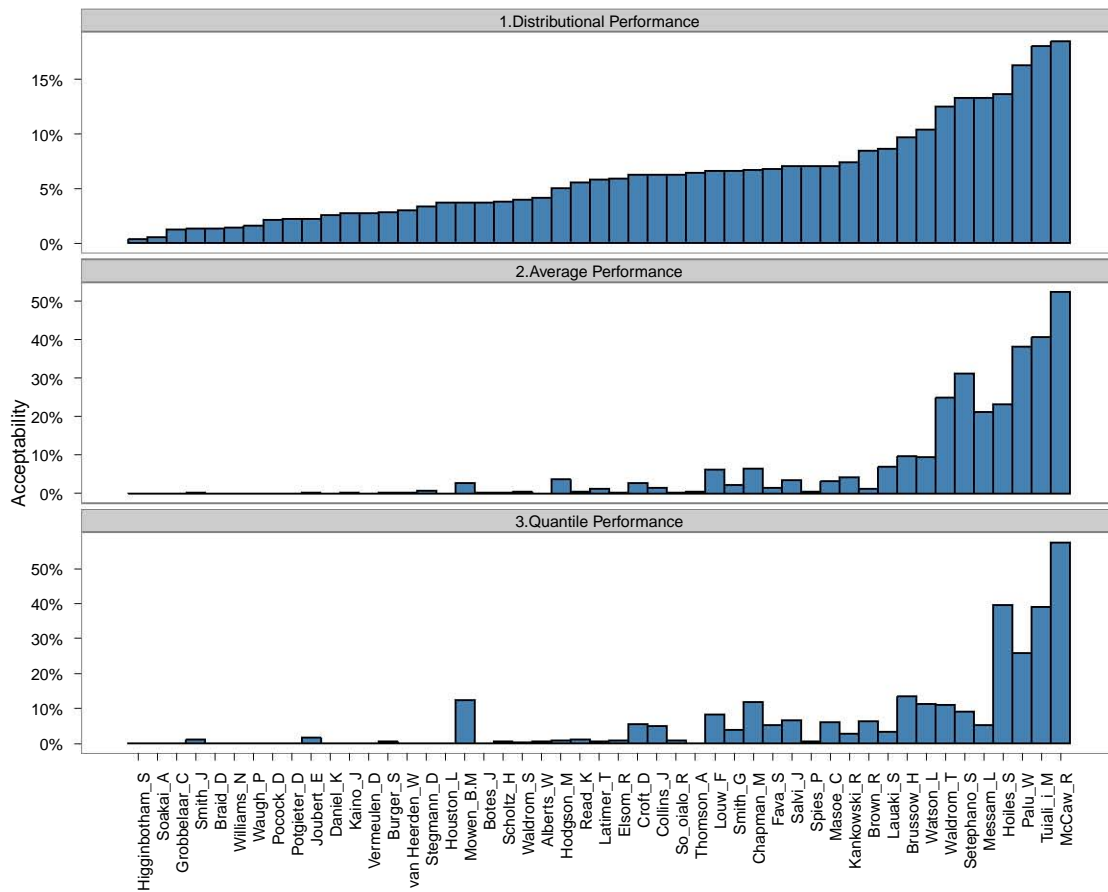


Figure 5.6: Rank acceptability indices for the loose forwards based on the empirical and average performance models

the highest acceptability, McCaw has a fairly flat central weight vector but it does emphasize both his work rate and effectiveness in the rucks and his ability to force turnovers and penalties. His lowest weights are for total passes - likely because he prefers to carry the ball into contact; and try creation - which he may not often be directly involved in due to his commitment to the breakdown area.

Like McCaw, Tuiali also has high acceptability and a relatively flat central weight vector which suggests his primary strength area is at the maul with high involvement

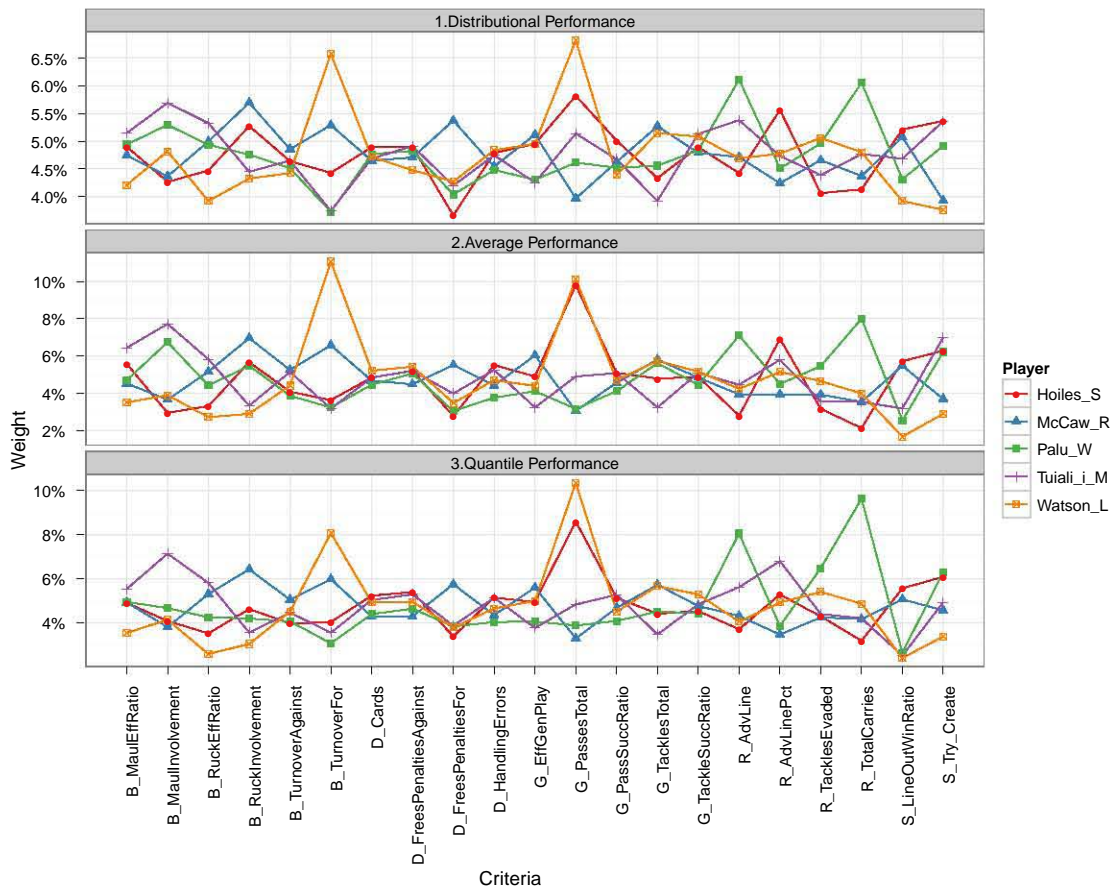


Figure 5.7: Central weight vectors for some of the loose forwards based on the empirical and average performance models

and efficiency scores. His pass counts and try creation are above average for a loose forward and he is also able to make the advantage line consistently but has slightly lower tackle counts and effectiveness in general play.

Palu and Hoiles seem to have contrasting profiles. Palu definitely fulfills the role of ball carrier, since his central weights highlight good performance in terms of total carries, reaching the advantage line and evading tackles. Hoiles seems to be more focused on passing the ball, and outperforms Palu slightly in a few other areas. The

two also differ slightly at the ruck and maul: Palu has higher maul involvement but lower efficiency whereas at the ruck Hoiles tends to be more involved but less efficient.

Watson has lower acceptability than the other four players, so it expected that his central weight vector won't be very smooth (i.e. that he may be slightly more specialized). With clear evidence from all three models, his standout strengths are an ability to win turnovers and to distribute the ball consistently with a high number of passes per game. By comparison to other loose forwards his lower scoring areas are in terms of general involvement and efficiency at the ruck, number of frees and penalties won, ability to win lineout ball and try creation.

5.4.3 Feedback from rugby coaches

As mentioned previously, there has been little opportunity for feedback from the analyst who commissioned the project since he left the Stormers in 2010, however he was walked through both the methodology and some of the results described above for the center and loose forward positions. He expressed that the approach shows plenty of potential and does seem to reveal some interesting insights on player profiles and performance.

Again it must be stressed that the strength of the model and the analysis is highly dependent on the accuracy and integrity of the underlying data, but the above results do provide evidence that the technique is effectively able to pick up the nature of a players game. Furthermore, the method allows for presentation of the results in a manner that is easy to understand and interpret. Based on this study, another conclusion was that this technique would be of most value in situations where there are a lot of players to choose from and the coaching staff are not already entirely familiar with all of the players. A prime example of this would be in a trials situation preceding squad selection.

Chapter 6

Conclusions

When faced with a decision between a number of potential alternatives, the strength of these alternatives is often shrouded by uncertainty. Loosely speaking, there are generally two sources of uncertainty: uncertainty around the exact consequences of choosing a particular alternative (criteria measurements), and uncertainty around the extent to which these consequences might influence the desirability of that alternative (preference information). MAUT provides an axiomatic basis for choice, but requires extensive involvement of the decision makers themselves and can be very time consuming since model inputs generally need to be highly specified. In many cases, the uncertainty referred to above is not always dealt with explicitly but rather addressed by means of sensitivity analysis following the application of a full MAUT model.

SMAA offers a generalized framework which attempts to handle both of the above-mentioned aspects of uncertainty by allowing both criteria measurements and preference information to be expressed as arbitrarily distributed stochastic variables. Since the approach can be implemented with little or no preference information, a major advantage is that it can offer a potential solution for limited decision support in low involvement contexts where direct interaction with the decision makers themselves is limited due to availability and/or time constraints. Despite the flexibility of this approach however, in practice it can still be difficult and time consuming to fully

assess the distribution of criteria measurements for all of the different alternatives present in a given decision.

It is for this reason that we consider the application and suitability of some SMAA models which use simplified formats for representing uncertainty in the attribute evaluations. The key advantages are ease of use and greater transparency, which come at the cost of some accuracy. The simulation experiment setup in Chapter 3 aims to evaluate this loss of accuracy by assessing the extent to which these simplified models can approximate results obtained using MAUT. We do not intend to use these results to conclude a detailed apparatus prescribing rules for using particular SMAA models in particular decision contexts. Rather the results suggest a general course of action for practitioners who for some reason are unable to fully elicit weight information. As with any other simulation experiment, we stress that all our findings are limited by the range of simulated cases. The complexity of the simulation apparatus is largely to ensure that a suitable range of problems have been covered, although doubtless there are counterexamples to our findings which could be constructed.

Our main conclusions regarding the implementation of SMAA to support low-involvement decision making are as follows:

1. Acceptability indices obtained from SMAA models can be expected to return good approximations to the full MAUT model, provided that some preference information is provided and that uncertainty is assessed with little or no error. Our results suggest that specifying ranges of possible weights covering around 70% of the full range of possible outcomes would be sufficient to begin to obtain accurate approximations.
2. If no weight information is provided at all, the average accuracy of all SMAA models is very poor – only somewhat better than a random guess. Ranking alternatives based on SMAA acceptability indices without some restriction of the weight space cannot be recommended except possibly to begin short-listing

by removing the poorest performing alternatives e.g any alternatives with zero acceptability which are stochastically dominated by other alternatives.

3. It is not essential that the SMAA model uses probability distributions to represent attribute uncertainty. Good results can also be obtained using simplified uncertainty formats, particularly quantiles. Since assessing quantiles is considerably simpler than fitting a full distribution to each alternative's performance on each attribute, we believe the results provide reasonable motivation for the use of a quantile-based SMAA model in low-involvement contexts.
4. Accuracy is highest when only the best rank is used to construct the SMAA acceptability index, under the condition that no assessment errors are made. If errors are made, then using the top three ranks can be expected to return slightly better results.

The simulation results also provide a number of insights with regard to the effect of different problem contexts on the accuracy of the different models considered. These are largely in agreement with what has been reported in [13] for general utility/value function models under conditions of uncertainty (i.e. outside the context of SMAA):

5. Though the number of alternatives in a decision problem will affect its manageability from a cognitive point of view, our results suggest that the accuracy of the decision support tools under consideration is not affected by a large number of alternatives.
6. An increase in the number of criteria under consideration is significantly detrimental to accuracy for both the conventional and simplified SMAA models.
7. Results pertaining to model accuracy across different shapes of utility suggest that increasing deviations from linearity decrease the accuracy of all the models.
8. Quantile models perform best when attribute evaluation distributions are relatively "simple" – symmetric and with moderate variability. When distributions are skew and highly variable, it becomes more beneficial to capture the variation using a probability distribution.

Taken together, our results suggest some role for the SMAA methodology in low-involvement decision making. It seems that with limited preference information and summarised representations of uncertain attribute evaluations, the SMAA acceptability index can be used to rank order alternatives and select one which, on average, performs relatively well. In a simulation study such as this, however, we are unable to assess a number of practical issues. For example: whether the reduced accuracy is “acceptable” for decision makers; whether the trade-off between accuracy and the time saved on weight elicitation is felt to be worthwhile; or whether the limited weight information that must be assessed for the SMAA models is still felt to be too onerous a task.

To try and gain insight into some aspects of these issues, we sought to apply SMAA using simplified uncertainty formats in a real-world case study, in order to provide decision support around the selection of rugby players. We have demonstrated how SMAA can be successfully applied to the problem of player selection in rugby and have highlighted its potential use as a tool for identifying the strengths and weaknesses of players within positional profiles. Even though our implementation here utilized the full range (unrestricted) of inter-criterion weights, results from the expected value and quantile SMAA models for both the acceptability indices and the central weight vectors were shown to be largely in agreement with output from the conventional ‘full’ SMAA model for the highly ranked players. Though there was limited opportunity for interaction with the coach who was initially involved with the study, after looking at results across the 7 position groups the feedback was that the approach shows plenty of potential and does seem to reveal some interesting insights on player profiles and performance. Analysis of the results for the two position groups considered here produced sensible insights and was in line with our previous knowledge of these players. In conjunction with the findings of the simulation study, this provides further support for the assertion that simplified SMAA models can offer both acceptable accuracy and useful insight in practical decision problems.

By providing a framework in which player performance can be measured on a scale

relative to the spread of performances by other players in a similar position, we believe SMAA can be used to inform coaches as they make squad - and to some extent also team - selections. As highlighted at the outset, decision modelling cannot seek to replace the decision making process but only to inform it. When starting out with a low involvement context, SMAA is best considered as a descriptive aid, which can be used in an interactive manner with the decision maker(s) in order to assist them in better understanding of the alternatives involved in a decision and the preferences which favor them. In using SMAA to assess and compare the ability of individual rugby players, it is assumed that there are multiple specific performance criteria on the basis of which these players can be realistically compared. If the performance criteria utilized aren't reliable indicators of a player's ability, then no model or methodology (SMAA included) will be able to provide information on which players might be better and how they differ from each other. Given that rugby is a game which involves a wide variety of team dynamics, obvious 'successful' output (or criteria) measures are often ascribed to the combined efforts of the team as a whole rather than to particular individuals, and the problem of choosing the right evaluation criteria for each position is thus a particularly challenging one. This is certainly a key area where future research can be focused.

Bibliography

- [1] Agnew, M. S. B. (2006). *Game Analysis in Rugby Union : A Review of the Literature*. Ph. D. thesis, Auckland University of Technology.
- [2] Alvarez-Guerra, M., L. Canis, N. Voulvoulis, J. R. Viguri, and I. Linkov (2010, September). Prioritization of sediment management alternatives using stochastic multicriteria acceptability analysis. *The Science of the total environment* 408(20), 4354–67.
- [3] Barron, F. (1987). Influence of missing attributes on selecting a best multiattributed alternative. *Decision Sciences* 18, 194–205.
- [4] Belton, V. and T. Stewart (2001). *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers.
- [5] Bracewell, P. (2001). Perception of individual rugby player performance and the impact of non-performance on statistical analyses. *Research Letters in the Information and Mathematical Sciences* 2, 19–22.
- [6] Bracewell, P. (2002). Implementing statistics in a diagnostic coaching structure for rugby. *Research Letters in the Information and Mathematical Sciences* 3, 79–84.
- [7] Bracewell, P., D. Meyer, and S. Ganesh (2003). Creating and Monitoring Meaningful Individual Rugby Ratings. *Research Letters in the Information and Mathematical Sciences* 4, 19–22.

- [8] Cameron, M. (2004). Using performance and results to measure the quality of sports competitors. In *Paper presented at the Seventh Mathematics and Computers in Sport Conference, Palmerston North, New Zealand*.
- [9] Cox, R. (2002). *Sport Psychology: Concepts and Applications* (Fifth Edit ed.). McGraw Hill.
- [10] Deffenbacher, K., B. Bornstein, S. Penrod, and E. McGorty (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior* 28(6), 687–706.
- [11] Durbach, I. (2006, May). A simulation-based test of stochastic multicriteria acceptability analysis using achievement functions. *European Journal of Operational Research* 170(3), 923–934.
- [12] Durbach, I. (2009, August). The use of the SMAA acceptability index in descriptive decision analysis. *European Journal of Operational Research* 196(3), 1229–1237.
- [13] Durbach, I. and S. Davis (2012). Decision support for selecting a shortlist of electricity-saving options: A modified SMAA approach. *ORiON* 28(2), 99–116.
- [14] Durbach, I. and T. Stewart (2009, April). Using expected values to simplify decision making under uncertainty. *Omega* 37(2), 312–330.
- [15] Durbach, I. and T. J. Stewart (2012). A comparison of simplified value function approaches for treating uncertainty in multi-criteria decision analysis. *Omega* 40(4), 456–464.
- [16] Dyer, J. S. and R. K. Sarin (1979). Measurable Multiattribute Value Functions. *Operations Research* 27(4), 810–822.
- [17] Evert, A. (2006). *The significance of dominant ball carrying collisions as an indicator of success in rugby union and the biomechanical analysis thereof*. Ph. D. thesis, University of Pretoria.

- [18] Figueira, J. R., S. Greco, and M. Ehrgott (2005). *Multiple criteria decision analysis: state of the art surveys*. Springer Verlag.
- [19] Franks, I. M. and G. Miller (1986). Eyewitness testimony in sport. *Journal of Sport Behaviour* 9(1), 38–45.
- [20] Gilboa, I. (2011). *Making Better Decisions*. John Wiley & Sons Ltd.
- [21] Gilovich, T., D. Griffin, and D. Kahneman (2002). *Heuristics and biases: the psychology of intuitive judgement*. Cambridge University Press.
- [22] Hakes, J. and R. Sauer (2006). An economic evaluation of the Moneyball hypothesis. *The Journal of Economic Perspectives* 20(3), 173–186.
- [23] Hammersley, J. M. and D. Handscomb (1964). *Monte Carlo Methods*. Methuen Publishing Ltd.
- [24] Hokkanen, J., R. Lahdelma, and P. Salminen (1999, March). A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor. *Socio-Economic Planning Sciences* 33(1), 1–23.
- [25] Hughes, M. (1996). *Science and soccer* (First ed.). SPON Press, Taylor and Francis Group.
- [26] Hughes, M. D. and R. M. Bartlett (2002, October). The use of performance indicators in performance analysis. *Journal of Sports Sciences* 20(10), 739–754.
- [27] International-Rugby-Board (2008). *LAWS OF THE GAME*. International Rugby Board.
- [28] James, N., S. D. Mellalieu, and N. M. P. Jones (2005). The development of position-specific performance indicators in professional rugby union. *Journal of Sports Sciences* 23(1), 63–72.
- [29] Jeukendrup, A. E. and M. Gleeson (2004). *Sport Nutrition: An introduction to energy production and performance*. Human Kinetics.

- [30] Kahneman, D. and A. Tversky (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291.
- [31] Kangas, J. and A. Kangas (2003, January). Multicriteria approval and SMAA-O in natural resources decision analysis with both ordinal and cardinal criteria. *Journal of Multi-Criteria Decision Analysis* 12(1), 3–15.
- [32] Keefer, D. L. and S. E. Bodily (1983, May). Three-Point Approximations for Continuous Random Variables. *Management Science* 29(5), 595–609.
- [33] Keeney, R. L. and H. Raiffa (1993). *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge: Cambridge University Press.
- [34] Lahdelma, R., J. Hokkanen, and P. Salminen (1998). SMAA-stochastic multiobjective acceptability analysis. *European Journal of Operational Research* 106(1), 137–143.
- [35] Lahdelma, R., K. Miettinen, and P. Salminen (2003, May). Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA). *European Journal of Operational Research* 147(1), 117–127.
- [36] Lahdelma, R., K. Miettinen, and P. Salminen (2005, August). Reference point approach for multiple decision makers. *European Journal of Operational Research* 164(3), 785–791.
- [37] Lahdelma, R. and P. Salminen (2001). SMAA-2 : Stochastic Multicriteria Analysis for Group Decision Making. *Operations Research* 49(3), 444–454.
- [38] Lahdelma, R. and P. Salminen (2002, September). Pseudo-criteria versus linear utility function in stochastic multi-criteria acceptability analysis. *European Journal of Operational Research* 141(2), 454–469.
- [39] Lahdelma, R. and P. Salminen (2006). Stochastic multicriteria acceptability analysis using the data envelopment model. *European Journal of Operational Research* 170(1), 241–252.

- [40] Lahdelma, R., P. Salminen, and J. Hokkanen (2002, October). Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. *European Journal of Operational Research* 142(2), 345–356.
- [41] Lames, M. and T. McGarry (2007). On the search for reliable performance indicators in game sports. *International Journal of Performance Analysis in Sport* 7(1), 62–79.
- [42] Lewis, M. (2003). *Moneyball - The Art of Winning an Unfair Game*. W.W. Norton & Company Inc.
- [43] Mellalieu, S. (2008). Editorial - Science and rugby union. *Journal of Sports Sciences* 26(8), 791–794.
- [44] Nakagawa, A. and K. Hirose (2007). Factors to Heighten Success of Ball Acquisition in Kick-off and 50m Restart Kick Play in Rugby Football. *International Journal of Sport and Health Science* 5, 135–146.
- [45] Neisser, U. (1982). *Memory Observed: remembering in natural contexts*. W.H. Freeman and Company.
- [46] Pervin, L. (1994). A critical analysis of current trait theory. *Psychological Inquiry* 5(2), 103–113.
- [47] Prim, S., M. van Rooyen, and M. Lambert (2006). A comparison of performance indicators between the four South African teams and the winners of the 2005 Super 12 Rugby competition. What separates top from bottom? *International Journal of Performance Analysis in Sport* 6(2), 126–133.
- [48] Reilly, T., N. Seecher, P. Snell, and C. Williams (Eds.) (1990). *Physiology of Sports*. Taylor and Francis Ltd.
- [49] Simon, H. (1955). A Behavioral Model of Rational Choice Author (s): Herbert A . Simon Published by : Oxford University Press. *The Quarterly Journal of Economics* 69(1), 99–118.

- [50] Simon, H. (1956, March). Rational choice and the structure of the environment. *Psychological review* 63(2), 129–138.
- [51] Smets, P. (1991, September). Varieties of ignorance and the need for well-founded theories. *Information Sciences* 57-58, 135–144.
- [52] Stewart, T. J. (1996, December). Robustness of Additive Value Function Methods in MCDM. *Journal of Multi-Criteria Decision Analysis* 5(4), 301–309.
- [53] Systems, S. A. (2009). The Rugby Analyst. Fair Play Pty Ltd, Queensland, Australia.
- [54] Tervonen, T., J. Almeida-Dias, J. R. Figueira, R. Lahdelma, and P. Salminen (2005). SMAA-TRI: A parameter stability analysis method for ELECTRE TRI. *Research Report* 6(6).
- [55] Tervonen, T., J. R. Figueira, R. Lahdelma, J. Almeida-Dias, and P. Salminen (2009, January). A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research* 192(1), 236–242.
- [56] Tervonen, T., H. Hakonen, and R. Lahdelma (2008, June). Elevator planning with stochastic multicriteria acceptability analysis. *Omega* 36(3), 352–362.
- [57] Tervonen, T. and R. Lahdelma (2007). Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research* 178(2), 500–513.
- [58] Tervonen, T., G. van Valkenhoef, N. Basturk, and D. Postmus (2013). Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research* 224(3), 552–559.
- [59] Trninic, S., V. Papic, V. Trninic, and D. Vukicevic (2008). Player selection procedures in team sports games. *Acta Kinesiologica* 2, 24–28.
- [60] van den Berg, P. and D. Malan (2010). Match analysis of the 2006 Super 14 Rugby Union tournament. *African Journal for Physical, Health Education, Recreation and Dance* 16(4), 580–593.

- [61] van Rooyen, M., M. Lambert, and T. Noakes (2006). A Retrospective analysis of the IRB statistics and video analysis of match play to explain the performance of four teams in the 2003 Rugby World Cup. *International Journal of Performance Analysis in Sport* 6(1), 57–72.
- [62] van Rooyen, M. and T. Noakes (2006). An Analysis of the movements, both duration and field location, of 4 teams in the 2003 Rugby World Cup. *International Journal of Performance Analysis in Sport* 6(1), 40–56.
- [63] Vincke, P. (1992). *Multicriteria Decision-Aid* (English ed.). John Wiley & Sons Ltd.
- [64] Wang, J. and S. Zionts (2006). Random Weight Generation in Multiple Criteria Decision Models. In *Conference of the International Society on Multiple Criteria Decision Making*, Chania, Greece, pp. 1–10.
- [65] Yu, W. (1992). *Aide multicritère à la décision dans le cadre de la problématique du tri: Concepts, méthodes et applications*. Ph. D. thesis, Université Paris-Dauphine.
- [66] Zatsiorsky, V. M. (Ed.) (2000). *Biomechanics in Sport: Performance Enhancement and Injury Prevention*. Blackwell Science Ltd.