TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

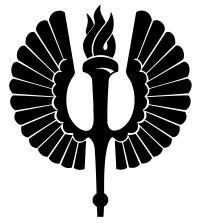
SARJA – SER. AI OSA – TOM. 376

ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA

New directions in Stochastic Multicriteria Acceptability Analysis

by

Tommi Tervonen



TURUN YLIOPISTO
Turku 2007

TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA – SER. AI OSA – TOM. 376

ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA

New directions in Stochastic Multicriteria Acceptability Analysis

by

Tommi Tervonen

This thesis has been completed under an agreement between the following two universities:

University of Turku
Department of Information Technology
20014 Turku, Finland

and

University of Coimbra
Faculty of Economics
3004-512 Coimbra, Portugal

TURUN YLIOPISTO
Turku 2007

Supervisors

Risto Lahdelma
Department of Information Technology
University of Turku
Finland

José Rui Figueira
Department of Engineering and Management
Instituto Superior Técnico, Technical University of Lisbon
Portugal

Pekka Salminen
School of Business and Economics
University of Jyväskylä
Finland

Reviewers

Annika Kangas
Department of Forest Resource Management
University of Helsinki
Finland

Salvatore Greco
Faculty of Economics
University of Catania
Italy

Opponents

Jyrki Kangas
Metsähallitus
Finland

João Paulo Costa
Faculty of Economics
University of Coimbra
Portugal

ISBN 978-951-29-3405-8 (PRINT)
ISBN 978-951-29-3406-5 (PDF)
ISSN 0082-7002
Painosalama Oy - Turku, Finland 2007

Abstract

Decisions taken in modern organizations are often multi-dimensional, involving multiple decision makers and several criteria measured on different scales. Multiple Criteria Decision Making (MCDM) methods are designed to analyze and to give recommendations in this kind of situations. Among the numerous MCDM methods, two large families of methods are the multi-attribute utility theory based methods and the outranking methods. Traditionally both method families require exact values for technical parameters and criteria measurements, as well as for preferences expressed as weights. Often it is hard, if not impossible, to obtain exact values.

Stochastic Multicriteria Acceptability Analysis (SMAA) is a family of methods designed to help in this type of situations where exact values are not available. Different variants of SMAA allow handling all types of MCDM problems. They support defining the model through uncertain, imprecise, or completely missing values. The methods are based on simulation that is applied to obtain descriptive indices characterizing the problem.

In this thesis we present new advances in the SMAA methodology. We present and analyze algorithms for the SMAA-2 method and its extension to handle ordinal preferences. We then present an application of SMAA-2 to an area where MCDM models have not been applied before: planning elevator groups for high-rise buildings. Following this, we introduce two new methods to the family: SMAA-TRI that extends ELECTRE TRI for sorting problems with uncertain parameter values, and SMAA-III that extends ELECTRE III in a similar way. An efficient software implementing these two methods has been developed in conjunction with this work, and is briefly presented in this thesis. The thesis is closed with a comprehensive survey of SMAA methodology including a definition of a unified framework.

Extended abstract in Finnish

Nykypäivänä monet organisaatiot kohtaavat päivittäin tilanteita, joissa niiden tulee tehdä päätöksiä ottaen huomioon useita vaikutuksia. Tämän tyyppiset päätökset vaihtelevat tehtaan sijoituskohteesta pilvenpiirtäjän hissiryhmän tyypin valintaan. Molemmat edellämainituista tilanteista sisältyvät tärkeään ryhmään päätöksenteko-ongelmia, jotka koostuvat äärellisestä määrästä vaihtoehtoja joiden hyvyyttä mitataan usealla kriteerillä. Näiden mittausten perusteella paras vaihtoehto voidaan valita, tai vaihtoehdot voidaan järjestää suosittavuuden mukaan, tai ne voidaan lajitella kategorioihin, jotka ovat etukäteen määriteltyjä ja järjestettyjä paremmuuden mukaan.

Tässä väitöskirjassa käsiteltävät päätöksenteko-ongelmat ovat yllä mainittua tyyppiä. Kriteerit joilla vaihtoehtoja mitataan voivat olla ordinaalisia tai kardinaalisia. Ordinaalisilla kriteereilla ainoastaan vaihtoehtojen suosittavuusjärjestys voidaan määrittää. Kardinaalisilla kriteereillä vaihtoehtoja kyetään mittaamaan numeerisilla arvoilla. Päätöksentekomenetelmät yhdistävät eri kriteerien arvot ottaen huomioon päätöksentekijöiden preferenssit pyrkimyksenä rakentaa preferenssirelaatio joka ratkaisee kyseessä olevan ongelman. Useimmissa käytännön sovelluksissa tätä ratkaisua ei tulisi tulkita kirjaimellisesti, vaan käyttää lähtökohtana syvemmälle analyysille. Jatko-analyysi voi mahdollisesti koostua mallin arvojen tarkentamisesta tai muuttamisesta.

Saavutetut ratkaisut eivät ole riippuvaisia ainoastaan kriteerimittauksista ja päätöksentekijöiden preferensseistä, vaan myös mallin tyypistä ja sen teknisistä parametreista. Tämän vuoksi ongelmaan on suositeltavaa soveltaa useaa eri päätöksentukimenetelmää ja verrata niiden antamia tuloksia, jos vain mahdollista. Vertailu tulee tehdä kuitenkin siten, että päätöksentekijät ja analysoija ymmärtävät sekä edellytykset mallin käytölle, että mallin hyvät ja huonot puolet.

Monikriteerinen päätöksenteko on tieteellisenä kenttänä suhteellisen nuori ja jakautunut eri koulukuntiin. On olemassa useita menetelmiä jotka kukin ottavat huomioon käytännön ongelmissa kohdattavia erityispiirteitä. Kaksi suurta menetelmäperhettä ovat moniattribuuttiseen hyötyteoriaan perustuvat menetelmät ja outranking-menetelmät. Tässä työssä keskitytään out-

ranking-menetelmistä ELECTRE-menetelmäperheeseen. Hyötyteoria antaa pohjan vanhimmille vielä käytössä oleville päätöksentekomenetelmille. Sen menetelmillä on aksiomaattinen pohja toisin kuin ELECTRE-menetelmillä. Nämä sallivat kuitenkin monimuotoisemman preferenssien mallintamisen kynnysfunktioiden avulla. Kynnysfunktiot ja muut ELECTRE:n käsitteet saattavat joillekin päätöksentekijöille olla helpommin ymmärrettävissä kuin hyötyteoria.

Viimeisinä vuosina on tullut selväksi, että päätöksentekomenetelmien tulisi kyetä ottamaan huomioon mallin parametrien arvojen epävarmuudet ja epätarkkuudet. Perinteiset moniattribuuttiseen hyötyteoriaan pohjautuvat menetelmät tai ELECTRE-menetelmät eivät tähän kykene. 1990-luvulla syntyi uusi menetelmäperhe, stokastinen monikriteerinen arvostusanalyysi (SMAA), joka eksplisiittisesti sallii epävarmuuksien mallintamisen. Eri SMAA-menetelmät soveltuvat kaiken tyyppisille päätöksenteko-ongelmille ja mahdollistavat epävarmojen, epätarkkojen, ja puuttuvien parametriarvojen käytön. Ensimmäinen SMAA-menetelmä ja sen SMAA-2--laajennus pohjautuvat moniattribuuttiseen hyötyteoriaan ja käyttävät sitä parhaan vaihtoehdon valintaan (SMAA) tai vaihtoehtojen järjestämiseen (SMAA-2). SMAA-menetelmiä voidaan käyttää myös sellaisissa päätöksenteko-ongelmissa, joissa ei ole ollenkaan preferenssitietoa saatavilla. Myös epätarkkaa preferenssitietoa voidaa käyttää SMAA:ssa. Samoin kriteerimittaukset ja muut mallin parametrit voivat olla epätarkkoja.

SMAA-menetelmät käyttävät simulaatiota laskeakseen ongelmaa kuvaavia indeksejä. Perheen eri menetelmät tuottavat erilaisia indeksejä. Näistä useasti tärkeimpiä ovat preferenssien, kriteerimittausten ja muiden parametrien osuus, joka sijoittaa vaihtoehdon tietylle lajittelusijalle tai tiettyyn kategoriaan. Nämä indeksit lasketaan teoriassa moniulotteisina integraaleina, mutta käytännössä niitä arvioidaan Monte Carlo-simulaatiotekniikalla. Kuten tässä väitöskirjassa myöhemmin näytetään, SMAA:n algoritmit ovat nopeita ja tarpeeksi tarkkoja käytettäväksi kaikissa käytännön kokoa olevissa päätöksenteko-ongelmissa.

Tämä väitöskirja koostuu artikkeleista jotka käsittelevät useita SMAA-menetelmäperheen osa-alueita. Tutkimuksen ensimmäisessä osassa analysoitiin SMAA-2 ja SMAA-O menetelmien algoritmit sekä teoreettisesti että käytännön testeillä. Teoreettinen osa johti laskennallisen kompleksisuuden määrittämiseen sekä algoritmien tarkkuuden laskemiseen. Tarkkuuslaskelmien pohjalta voidaan määrittää tarvittavien Monte Carlo simulaatioiden määrä, jotta saavutetaan riittävän pienet luottamusvälit ko. indekseille. Tässä osassa kuvaamme myös SMAA-2 ja SMAA-O algoritmit pseudokoodina.

Toinen osa sisältää realistisen sovelluksen, jossa SMAA:ta on sovellettu hissisuunnitteluun. Tässä sovelluksessa tutkittiin SMAA:n soveltuvuutta hissiryhmän valintaan korkeiden rakennusten suunnittelussa. Tutkimus

tehtiin yhdessä KONE-yhtiön työntekijän kanssa. Käytimme KONE:en rakennussimulaattoria tuottamaan mittauksia suoritustehokkuuskriteereille. Muodostimme näistä monimuuttujisen normaalijakauman, jossa epävarmuudet suoritustehokkuuskriteereiden kesken olivat riippuvaisia. Tämän lisäksi käytimme ordinaalista kriteeriä hinnan määrittämiseen (tarkkoja arvoja ei ollut saatavilla), sekä kardinaalista epätarkkaa kriteeriä hissiryhmän vaatiman lattiapinta-alan mittaamiseen. Mallin avulla kykenimme erottamaan alkuperäisistä 10:stä vaihtoehdosta neljä mahdollisesti implementoitavaa vaihtoehtoa, joista yksi osoittautui selvästi parhaaksi kompromissivaihtoehdoksi.

Väitöskirjan kolmas osa koostuu ELECTRE-menetelmien ja SMAA:n yhdistämisestä. ELECTRE-perheen kaksi menetelmää, ELECTRE III ja ELECTRE TRI laajennettiin käyttämään epätarkkoja arvoja. Nämä laajennukset kantavat nimiä SMAA-III ja SMAA-TRI. Tässä yhteydessä laajennettiin myös SMAA-menetelmäperheen käsitettä: sen sijaan että keskitytään eri SMAA-menetelmiin, on tärkeämpää nähdä niiden idea simulaation käyttämisestä eri indeksien laskemisessa. Tällöin SMAA:ta voidaan käyttää "ulkoisten" menetelmien soveltamiseen epätarkkojen mittausten kanssa. SMAA:ta voidaan käyttää tällä tavalla myös herkkyysanalyysiin ja parametrien "herkkyyden" kvantifioimiseen.

Väitöskirjan viimeinen osa sisältää kaiken menneen tutkimuksen yhdistämisen yksittäiseksi kehykseksi, jonka perusteella käytettävä SMAA-menetelmä voidaan helposti valita ongelman erityispiirteiden perusteella. Tämä kehys auttaa myös SMAA-menetelmien puutteiden kartoituksessa ja siten tulevan tutkimuksen suunnittelussa.

Jotta päätöksenteko-menetelmä saavuttaisi suosiota myös kehittäjäpiirinsä ulkopuolella, tarvitaan sille käyttäjäystävällinen ohjelmisto. Teoreettisen tutkimuksen lisäksi tämän väitöskirjatutkimuksen aikana allekirjoittanut kehitti ohjelmiston joka implementoi SMAA-III- ja SMAA-TRI-menetelmät. Ohjelmisto on kirjoitettu C++-kielellä ja se käyttää graafista käyttöliittymää varten gtkmm-kirjastoa. Tämän vuoksi ohjelmisto on helposti siirrettävissä uusille alustoille. Tällä hetkellä se onkin jo käännetty Windows XP:lle, Max OS X:lle ja Linuxille. Tulevaisuuden tutkimustyö tulee keskittymään menetelmäperheen laajentamiseen sekä näiden laajennusten implementoimiseen ohjelmistossa.

Extended abstract in Portuguese

As organizações actuais deparam-se diariamente com situações de tomada de decisões com base em múltiplos critérios. Estas decisões podem ir desde a selecção de um sítio para localizar uma fábrica até à escolha de um conjunto de elevadores para um arranha-céu. Ambos os exemplos mencionados têm como característica comum o facto de disporem de um conjunto finito de alternativas avaliadas a partir de um conjunto ou família coerente de critérios. Dependendo da problemática em questão, podemo-nos preocupar com a escolha da melhor ou melhores alternativas, com a ordenação das alternativas de melhor para a pior ou ainda com a classificação das alternativas em categorias predefinidas.

Os problemas de decisão tratados nesta tese são dos tipos mencionados anteriormente. Os critérios definidos para avaliar as diferentes alternativas podem ser de natureza ordinal ou cardinal. Em relação aos critérios ordinais apenas pode ser construída uma ordenação das alternativas de acordo com as preferências do decisor. Relativamente aos critérios cardinais, estes têm a vantagem de poder ser traduzidos por valores numéricos. Nesta tese partimos do princípio que a família de critérios possa comportar ambos os tipos de critérios, que são usados para modelar ou construir as preferências através de uma relação de prevalência que será explorada para "resolver" os problemas atrás mencionados. Nos problemas reais os resultados provenientes da aplicação directa de um determinado método não deverão ser interpretados literalmente como tais, ou seja, tal como nos aparecem após a aplicação do método. Uma análise mais profunda será necessária, as imprecisões, inexactidões, insuficiências, incertezas ou indeterminações nas avaliações das alternativas segundos os diferentes critérios bem como aquelas que estão associadas aos parâmetros (preferências ou técnicos) dos modelos fazem intervir uma parte de arbitrário que necessita de ser estudada de forma mais aprofundada.

O apoio multicritério à decisão é um campo da ciência relativamente novo e dispersa-se por várias escolas de pensamento. Existem vários métodos propostos na literatura, cada um privilegiando algumas características

particulares encontradas nos problemas reais. Há, no entanto, duas grandes famílias de métodos que por várias razões, incluindo algumas de natureza histórica, se impuseram: métodos baseados na teoria da utilidade ou valor multicritério e métodos baseados nas relações de prevalência. Este trabalho concentra-se na família dos métodos ELECTRE, que são métodos baseados na construção de uma ou várias relações de prevalência seguida de uma exploração dessa ou dessas relações.

A teoria de utilidade dá a base para os métodos de apoio à decisão mais antigos, mas ainda em uso. Estes métodos têm fortes bases axiomáticas, o que não é o caso dos métodos ELECTRE. De qualquer maneira, os métodos ELECTRE têm uma vantagem em relação aos métodos baseados na teoria da utilidade, dado que partem do princípio que não existe uma função utilidade que por algum processo se podem determinar. Os métodos ELECTRE têm ainda a característica de trabalharem com um modelo com lineares, baseiam-se assim no chamado modelo do pseudo-critério.

Recentemente, tem-se atendido com alguma profundidade e preocupação para o facto de que os parâmetros dos modelos são incertos ou inexactos bem como os desempenhos das alternativas nos critérios. Os métodos tradicionais baseados na teoria de utilidade multicritério e os métodos de ELECTRE não o fizeram de forma sistemática. No início dos anos 90 nasceu uma nova família dos métodos, Stochastic Multricriteria Acceptability Analysis (SMAA), que explicitamente permite modelar esta imprecisão. Os vários métodos da família SMAA foram concebidos para problemas de apoio à decisão de todos os tipos, e possibilitam usar parâmetros e desempenhos imprecisos, inexactos e/ou insuficientes. O primeiro método da família SMAA e a sua extensão SMAA-2 são baseados na teoria da utilidade multicritério e forma concebidos para as problemáticas da escolha da melhor alternativa (o caso do SMAA) ou para a ordenação das alternativas (SMAA-2). Os métodos SMAA podem igualmente ser usados em problemas de decisão onde não se disponha de nenhuma informação preferencial.

Os métodos da família SMAA usam as técnicas de simulação para as medidas descritivas ou índices que servem para dar informação estatística sobre o problema. Os diferentes métodos propõem diferentes índices, dentre os mais usuais destaca-se aquele que nos permite ter informação sobre as preferências que colocam uma certa alternativa numa dada posição da ordenação. Estes índices são calculados de forma exacta a partir da teoria dos integrais múltiplos. Mas, na prática são estimadas através da simulação de Monte Carlo. Nesta tese pode-se constatar que os algoritmos dos métodos SMAA são rápidos e suficientemente exactos para serem usados em todos os problemas de decisão de dimensões aceitáveis.

Esta tese é composta de vários artigos que tratam também várias subáreas de aplicação dos métodos SMAA. Na primeira parte, fez-se uma análise teórica e experimental dos algoritmos SMAA-2 e SMAA-O. A parte

teórica resultou no estudo da complexidade e no cálculo de precisão dos algoritmos. Baseado nos cálculos sobre a precisão foi possível determinar a quantidade de simulações de Monte Carlo para obter intervalos de confiança suficientemente pequenos, mas significativos para justificar os índices em questão. Nesta parte apresenta-se igualmente os algoritmos de SMAA-2 e SMAA-O em pseudo-código.

A segunda parte contém uma aplicação real do SMAA na área do planeamento da instalação de um conjunto de elevadores. Nesta aplicação foi investigada a aplicabilidade do SMAA na escolha de um conjunto de elevadores para arranha-céus. A investigação e aplicação foram efectuadas junto da empresa KONE. Usou-se o simulador de prédios da KONE para construir a matriz de desempenhos. A imprecisão relativamente a estes desempenhos foi modelada através de uma distribuição de Gauss multivariada. Para além disso usou-se um critério ordinal para modelar o preço, dado que os valores exactos não eram conhecidos, e um critério cardinal para representar a área necessária. A partir de uma análise preliminar 4 das 10 alternativas inicias puderam ser seleccionadas como potenciais opções a implementar. Seguidamente pode observar-se que uma destas 4 alternativas é claramente a melhor alternativa de compromisso.

A terceira parte da tese é consagrada à uma combinação entre os métodos ELECTRE e SMAA. Dois métodos da família ELECTRE, ELECTRE III e ELECTRE TRI, foram estendidos para usar valores inexactos. Estas extensões chamam-se SMAA-III e SMAA-TRI. O conceito de família dos métodos de SMAA foi também estendido neste contexto: em vez de nos concentrarmos em métodos diferentes, o que é importante é a ideia de usar simulação para calcular os índices. Assim, o SMAA pode-se usar para aplicar métodos "externos" com parâmetros inexactos para analisar a sua robustez.

A última parte desta tese conte apresenta uma estrutura geral onde se enquadram os métodos SMAA e que pode facilitar a escolha de um determinado método em função das características específicas do problema. Esta estrutura também é útil para construir um mapa das lacunas dos métodos SMAA e ajuda, assim, no planeamento da investigação futura.

Para que um método de apoio à decisão alcance alguma popularidade fora do círculo de seus autores e comunidade científica da área, torna-se necessário dispor de um software amigável para os utilizadores. Para além da investigação teórica, dos testes experimentais e das aplicações reais efectuados no quadro desta tese, também foi consagrado algum tempo à implementação informática e construção de uma interface amigável dos métodos SMAA-III e SMAA-TRI. Os métodos foram implementados em linguagem C++ e usa a biblioteca gtkmm necessária para a construção da interface gráfica. Tal torna possível a portabilidade para novas plataformas. Neste momento já existem versões para Windows XP, Mac OS X, e

Linux. Parte da investigação futura será consagrada a novas extensões dos métodos SMAA e sua implementação.

Acknowledgements

The work leading into this thesis has been completed in various universities. The first year of the thesis I was working as a part-time assistant at the Department of Information Technology of the University of Turku (Finland). During this time the research was supported through my salary. In September 2004 I moved to the Faculty of Economics of the University of Coimbra, Portugal, where I worked in the INESC-Coimbra research center. This lasted until May 2006, when I moved to my current location: the Centre for Management Studies of Instituto Superior Técnico at the Technical University of Lisbon.

Although this thesis has been completed in a somewhat unorthodox fashion, by working in three different universities under supervision of three professors, I consider this being an advantage. I have enjoyed the possibility of experiencing working cultures of different universities, which is something that would never have happened by completing the whole thesis in Finland. I am grateful for all three supervisors of my thesis: Risto Lahdelma, José Rui Figueira, and Pekka Salminen. I thank them all both for their professional guidance as well as for making me feel welcome in all places I was working in.

My work abroad has been supported by numerous grants for research as well as for attending conferences. The major supporters have been the Finnish Cultural Foundation and the University Foundation of Turku (Turun Yliopistosäätiö). I thank them as well as the minor supporters that are not listed – without their support this work would have never been completed.

I also thank all my friends in Turku, Coimbra as well as in Lisbon, for providing content in my life outside science. And last but not least, I thank my parents, Jari and Leena Tervonen, for growing me up to question everything, as well as for the support they have provided me during this work.

List of original publications

[I] Tervonen T., Lahdelma R., 2007. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research 178(2), 500–513.

[II] Tervonen T., Hakonen H., Lahdelma R., 2008. Elevator planning with stochastic multicriteria acceptability analysis. Omega 36(3), 352–362.

[III] Tervonen T., Figueira J. R., Lahdelma R., Almeida Dias J., Salminen, P., 2007. A stochastic method for robustness analysis in sorting problems. European Journal of Operational Research (in press).

[IV] Tervonen T., Figueira J. R., Lahdelma R., Salminen P., 2007. Towards robust ELECTRE III with simulation: theory and software of SMAA-III. Working paper 1/2007 of CEG-IST. Submitted to European Journal of Operational Research.

[V] Tervonen T., Figueira J. R., 2007. A survey on stochastic multicriteria acceptability analysis methods. Journal of Multi-Criteria Decision Analysis (in press).

Contents

Chapter 1

Introduction

Organizations of all sizes often face situations in which decisions have to be taken based on multi-dimensional data. This type of decisions range from siting a factory to choosing a type of elevator system for a high-rise building. Both of these belong to an important class of decision making problems characterized by being composed of a finite set of alternatives evaluated on the basis of several criteria. Based on this evaluation the best alternative can be chosen, or the alternatives can be ranked, or they can be sorted into pre-defined and ordered categories. These types of problems are termed multiple criteria choosing, ranking, and sorting problems, respectively (Figueira et al., 2005). This type of decision making is called Multiple Criteria Decision Making (MCDM) or Multiple Criteria Decision Analysis (MCDA). Sometimes MCDA is used to abbreviate Multiple Criteria Decision Aiding as well (Belton and Stewart, 2002). In this thesis we use the term MCDM.

1.1 Multiple Criteria Decision Making

The problems considered in this thesis consist of a finite set of alternatives evaluated on basis of a set of criteria. The criteria can be ordinal, in the case that only the ranking of alternatives with respect to the criteria is available, or cardinal, if they can be measured on numerical scales. MCDM methods then aim to aggregate these values in a way that takes into account the preferences of the Decision Makers (DMs), in order to construct preference relations that solve the problem. In most real-life cases the solution should not be taken *per se*, but only as a starting point for a more through analysis. This includes possibly refining the model and re-performing the analysis in an iterative way.

The solutions are dependent not only on the criteria measurements and the preferences of the DMs, but also on the type of model and its technical

parameters. Therefore, if several methods can be applied to the problem setting in question, it is advisable to compare their results. For all the methods applied, the analyst as well as the DMs should acknowledge the prerequisites for its use, as well as the advantages and disadvantages the method has.

MCDM as a scientific field is relatively young and quite dispersed in different schools of thought. There exists a large amount of methods each designed to tackle certain specificities of real-life MCDM problems. Two large families of methods are the multiattribute utility theory (MAUT) based methods (see e.g. Keeney and Raiffa, 1976) and the outranking methods. Some well-known outranking methods are the ELECTRE methods (Roy, 1996), PROMETHEE methods (Brans and Mareschal, 2005), and the SIR method (Xu, 2001). In this thesis we concentrate in ELECTRE methods from the outranking approach. The utility theory gives a basis for the oldest MCDM methods still in use today, and is mathematically more firmly based than the ELECTRE methods. Nevertheless, the ELECTRE methods allow more versatile modelling of preferences in terms of thresholds. These might be more easy for some DMs to understand than the concepts of MAUT.

In the recent years it has become more than apparent that MCDM methods should be able to take into account uncertainty and imprecision in the parameters. The classical methods applying both MAUT and outranking model do not accomplish this. In the 1990s emerged another family of MCDM methods, the Stochastic Multicriteria Acceptability Analysis (SMAA) that explicitly allows to model uncertainty.

1.2 Stochastic Multicriteria Acceptability Analysis

SMAA is a family of methods that allows to handle various types of MCDM problems having uncertain, imprecise, or missing values for the model. The original SMAA method (Lahdelma et al., 1998) and its extension SMAA-2 (Lahdelma and Salminen, 2001) applied MAUT in order to choose the best alternative (SMAA) or to rank the alternatives (SMAA-2). SMAA methods can handle decision making situations with completely missing preference information. If some preference information is available, it can be incorporated to the model. Also criteria measurements as well as other parameters of the model can be imprecise.

SMAA methods apply simulation in order to provide the DMs with indices describing the problem. Different methods of the family produce different indices, but the most important ones are usually the share of weights, criteria measurements, and other parameters that assign an alternative to

a certain rank or category. These indices are calculated in theory through multidimensional integrals, but in practice Monte Carlo simulation is used to computed approximations for the values. As shall be shown later on, the SMAA algorithms are fast and accurate enough to use in all decision making problems of sizes encountered in practice.

1.3 Contributions

This thesis is composed of contributions in various parts of the SMAA methodology. We have analyzed the classical SMAA and SMAA-2 algorithms to give bounds on computational complexity and the amount of Monte Carlo iterations needed to obtain sufficient accuracy for the analysis. In this work we also described the algorithms for SMAA-2 and SMAA-O in pseudo-code.

A realistic case study of applying SMAA to elevator planning was made by us. In this study we applied the KONE building simulator in order to generate measurements from which a multivariate Gaussian distribution could be defined. This had to be done, because the measurements were highly correlated. Otherwise the results would have contained biases.

Another part of the work comprises of combining ELECTRE methods with the SMAA methodology. This work resulted in two new methods, SMAA-III and SMAA-TRI, that allow ELECTRE III and ELECTRE TRI, respectively, to be used with imprecise values. An important direction explored in these works was the usage of SMAA as an "external" method for performing robustness analysis with third party MCDM methods.

We have combined all the important past research into an integrated SMAA framework. This allows to get a complete picture of the current state of SMAA as well as its shortcomings. This framework allows to easily choose the SMAA variant to use based on the particularities of the decision making problem in question.

1.4 Outline of the thesis

We begin this thesis by giving a brief introduction to two classical MCDM methodologies considered in the contributions: MAUT and the outranking model as applied in ELECTRE methods. We then continue, in Chapter 3, by presenting the basic SMAA methodology. We introduce the basic SMAA method, SMAA-2, and an extension to handle ordinal criteria. We review some computational results, and present an application to elevator planning. We continue on the theory of SMAA by presenting outranking-based SMAA methods in Chapter 4. We define a unified SMAA framework in Chapter 5. In Chapter 6 we introduce the software produced in con-

junction to this research. The software implements the two new methods, SMAA-III and SMAA-TRI. We summarize the publications in Chapter 7 before proceeding to give concluding remarks in Chapter 8.

Chapter 2

Multiple Criteria Decision Making

In this chapter we briefly present two major MCDM methodologies: MAUT and the outranking model. We will concentrate in MAUT from the point of view that it will be used for ranking the alternatives, and in outranking model as applied in the ELECTRE methods. The small introduction to these two approaches is given because they are extended in SMAA approaches considered in the contributions. For more information and references on both approaches, see Belton and Stewart (2002). A detailed description of MAUT can be found in Keeney and Raiffa (1976), and one of ELECTRE methods in Roy (1996).

2.1 Multiattribute utility theory

Unidimensional utility theory is based on the concept that each alternative, when evaluated with respect to uncertain conditions, is assigned an expected utility value. These values describe the "goodness" of alternatives taking into account the preferences of the DM. The alternative with the highest expected utility is the most preferred one, or "best" in the considered problem setting.

The expected utility values are formed based on lotteries. These are defined as follows: consider a set of consequences $c_1, \ldots, c_n$, which are ordered so that c_n is the most preferred one and c_1 the least preferred one. Then consider two alternatives, x_1 and x_2, that each have for it assigned a probability p_i^1 or p_i^2 that when the alternative x_* is implemented, it results in consequence c_i with probability p_i^*. Now suppose that the DM asserts that for each i, he is indifferent between the two options:

1. Certainty. Receive c_i.

2. Risky. Receive c_n (the best consequence) with probability π_i and c_1 with the probability $1 - \pi_i$.

Then the expected values of the π's can also be used to numerically scale probability distributions over the c's (Keeney and Raiffa, 1976). That is, risks associated with consequences happening are used to calculate expected utilities of the alternatives. The DM's attitude towards risk defines which alternatives obtain the highest utilities. From these we can form utility functions that map the values into utility scores in an arbitrary, non-linear way that take into account the DM's attitude towards risk.

MAUT extends the unidimensional utility theory so that alternatives are considered with respect to several attributes, that is, criteria. In MAUT the utility functions of individual criteria are combined with scaling factors. These describe trade-offs the DM accepts to be consistent with his/her preferences. The weights of SMAA and SMAA-2 models that apply utility theory are these very scaling factors.

Another way of presenting the utility theory is to consider the utility functions to represent a complete preorder. This is defined with strongly complete and transitive binary relation based on trade-offs between criteria. By presenting utility theory in this way the difference between it and the outranking model considered in the next section comes clearer: the outranking model is based on the outranking relation that is weak, simply reflexive and neither strongly complete nor transitive.

2.2 Outranking model

Unlike MAUT, outranking model does not have an axiomatic basis, but instead relies on the intuition of how "goodness" of the alternatives is judged. The basic idea is that small differences between alternatives are indifferent, and differences over some certain magnitude do not bring any additional value. For example, when buying a car, it does not make a difference for most of the DMs whether the car costs 10000 euros or 20 more. In analogy, if one car costs 10000 and two others 2000000 and 3000000, probably there is no difference between preferability of the first over the second one to the first over the third one. Both of the latter ones are considered "bad" with respect to the price of the first one.

One of the largest families of outranking methods are the ELECTRE methods. It includes ELECTRE I (Roy, 1968), II (Roy, 1971; Roy and Bertier, 1973), III (Roy, 1978), IV (Roy and Hugonnard, 1982), TRI (Yu, 1992b), and 1S (Roy and Skalka, 1984). The two above mentioned characteristics of outranking models are modelled in ELECTRE methods as thresholds. That is, as an indifference threshold defining the difference until which the values are considered indifferent, and a preference threshold

for over which the differences do not bear additional value no matter how big they are. The thresholds can be defined as constant ones or, for example, as a percentage of the value. When a criterion is defined with two such thresholds, it is called a *pseudo-criterion*. All the ELECTRE methods extended in this thesis use pseudo-criteria.

Outranking methods are called such, because instead of aggregating their criteria values to a single attribute describing goodness of the alternative, they form an outranking relation between alternatives. An alternative is said to outrank another if it is considered *as good as or better*. The outranking methods then exploit these outranking relations, for example, to form a ranking of the alternatives (as in ELECTRE III, see Roy, 1978), or to assign the alternatives into categories (as in ELECTRE TRI, see Yu, 1992b). The weights in ELECTRE methods are not scaling factors as in MAUT-based models, but interpreted as votes for the criteria (Vincke, 1992).

2.3 Imprecision

The basic methods of both of the above-mentioned approaches require exact values to be defined for the model. In MAUT this means that although the attitude towards risk should take uncertainty into account, the attitude must be strictly defined with numerical values. The scaling factors (weights) between pairs of criteria must be exact. In ELECTRE methods the situation is similar: deterministic weights are needed, as well as exact values for the thresholds. In both methodologies the basic methods require exact values for the cardinal criteria measurements.

Some MAUT extensions and the ELECTRE methods allow to use also poorer, ordinal scales. In these only the ranking of alternatives is required. But through the years it had become apparent that more free modelling of imprecision is needed. This applies to all parameters of the models: preferences, criteria measurements, as well as to the technical parameters. A new approach that allows explicitly to account for uncertainties and imprecision as well as missing values in all parameters is the Stochastic Multicriteria Acceptability Analysis.

Chapter 3

Stochastic Multicriteria Acceptability Analysis

One way to overcome the weaknesses of the utility theory based approach is through an inverse method: instead of asking parameter values and giving an answer to the problem in question, the values resulting in different outcomes are described. The Stochastic Multicriteria Acceptability Analysis (SMAA) (Lahdelma et al., 1998; Lahdelma and Salminen, 2001) methods include computing multidimensional integrals over feasible parameter spaces in order to support DMs with descriptive measures. The methods solve various problems encountered in the traditional approach by allowing to use parameters with ignorance on the values. For example, usually different weight elicitation techniques produce different values, and therefore deterministic weights are harder to justify than, for example, weight intervals.

There have been similar approaches before SMAA. The first one was the comparative hypervolume criterion by Charnetski (1973) and Charnetski and Soland (1978). Rietveld (1980) and Rietveld and Ouwersloot (1992) presented similar methods for problems with ordinal criteria and ordinal preference information. Bana e Costa (1986, 1988) presented the overall compromise criterion. We note that the probability distributions used in SMAA are not the only possibility for modelling uncertain parameter values. Other possible approaches include entropy methods (Abbas, 2006; Jessop, 1999), rough sets (Greco et al., 1999, 2000, 2001, 2002; Pawlak and Słowiński, 1994), fuzzy sets (Roubens, 1997), interval methods (Mustajoki et al., 2006, 2005), and Dempster-Shafer theory (Beynon, 2002; Beynon et al., 2001a,b, 2000).

We will describe here the SMAA-2 (Lahdelma and Salminen, 2001) and SMAA-O (Lahdelma et al., 1998) methods, as well as our new ELECTRE-based SMAA methods, SMAA-TRI ([III]; Tervonen et al. (2007)) and SMAA-

III ([IV]). Other methods/extensions of the SMAA family not presented here are a technique for handling dependent criteria (Lahdelma et al., 2006a,b), cross confidence factors (Lahdelma and Salminen, 2006a), SMAA-D (Lahdelma and Salminen, 2006b) for data envelopment analysis, SMAA-P (Lahdelma and Salminen, 2003) applying prospect-theory, and Ref-SMAA (Lahdelma et al., 2005) for using reference points in SMAA. A method similar to Ref-SMAA has been presented by Durbach (2006). Different SMAA methods have been applied in various real-life cases: harbour citing (Hokkanen et al., 1999), waste treatment facility citing (Lahdelma et al., 2002), determining the implementation order of a general plan (Hokkanen et al., 1998), choosing a clearer for polluted soil (Hokkanen et al., 2000), forest planning (Kangas et al., 2006, 2003a; Kangas and Kangas, 2003; Kangas et al., 2005), elevator planning ([II]), and designing a framework for an oil spill response effectiveness (Linkov et al., 2007). For a complete survey on SMAA, see [V].

3.1 SMAA-2

The discrete decision-making problem considered in SMAA-2 (Lahdelma and Salminen, 2001) refers to a set of m alternatives $X = \{x_1, \ldots, x_i, \ldots, x_m\}$, that are evaluated on the basis of n criteria $\{g_1, \ldots, g_j, \ldots, g_n\}$. The evaluation of alternative x_i on criterion g_j is denoted $g_j(x_i)$. Without loss of generality we assume that all the criteria are to be maximized. The model considers multiple DMs, each having a preference structure representable through an individual weight vector w and a real-valued utility function $u(x_i, w)$ that has a commonly accepted shape. The most commonly used utility function is the linear one:

$$u(x_i, w) = \sum_{j=1}^{n} w_j g_j(x_i). \tag{3.1}$$

The weights will be assumed non-negative and normalized. Therefore the feasible weight space will be:

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

The feasible weight space of a 3-criteria problem with no preference information is illustrated in Figure 3.1.

The SMAA methods are developed for situations where criteria values and/or weights or other model parameters are not precisely known. Uncertain or imprecise criteria values are represented by stochastic variables

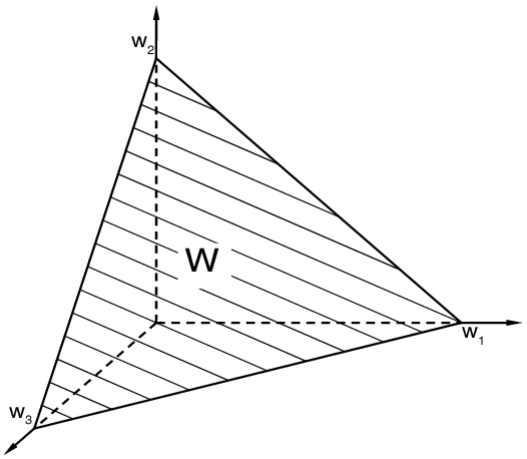


Figure 3.1: The feasible weight space of a 3-criteria problem.

ξ_{ij} corresponding to the deterministic evaluations $g_j(x_i)$ with density function $f_\chi(\xi)$ in the space $\chi \subseteq R^{m \times n}$. In principle, arbitrary distributions can be used, but in practice a uniform distribution in a certain interval or a Gaussian distribution is often used.

Similarly, the DMs unknown or partially known preferences are represented by a weight distribution with a joint density function $f_W(w)$ in the feasible weight space W. Total lack of preference information on weights is represented by the uniform weight distribution in W:

$$f_W(w) = 1/\mathrm{vol}(W).$$

As for the utility function based approaches, one should note that the weights are defined as scaling factors: the weights rescale the values of partial utility functions in such a way that the full swing in the scaled function indicates the importance of the criterion (see Belton and Stewart, 2002, Sect. 5.4).

The fundamental idea of SMAA is to provide decision support through descriptive measures calculated as multidimensional integrals over stochastic parameter spaces. Approximations for these measures are computed through Monte Carlo simulation. This means that they might contain errors, but the error margins are so small that usually they do not have to be taken into account (when the number of Monte Carlo iterations is large enough, see Section 3.4). SMAA-2 (Lahdelma and Salminen, 2001) defines three main types of descriptive indices for decision support: rank acceptability indices, central weight vectors, and confidence factors. These measures do not give definite answers, but rather provide DMs with more insight into

the decision making problem. In order to introduce these indices we first need to define a ranking function as follows:

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

where $\rho(true) = 1$ and $\rho(false) = 0$. Note that $rank(i, \xi, w) \in \{1, \ldots, m\}$. Let us also define the sets of favourable rank weights $W_i^r(\xi)$ as follows,

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

3.1.1 Rank acceptability index

The rank acceptability index b_i^r describes the share of parameter values granting alternative x_i rank r. It is computed as a multidimensional integral over the criteria distributions and the favourable rank weights as follows,

$$b_i^r = \int_{\xi \in \chi} f_\chi(\xi) \int_{w \in W_i^r(\xi)} f_W(w)\, dw\, d\xi.$$

The most acceptable (best) alternatives are those with high acceptabilities for the best (smallest) ranks. Evidently, the rank acceptability indices are within the range [0,1], where 0 indicates that the alternative will never obtain a given rank and 1 indicates that it will obtain the given rank always with any choice of weights.

Rank acceptability indices can be used to classify alternatives into stochastically efficient ($b_i^1 >> 0$) or inefficient ones (b_i^1 near zero, for example, < 0.05). A zero first rank acceptability index means that an alternative is never considered the best with the assumed preference model. For stochastically efficient alternatives, the index measures the strength of the efficiency considering simultaneously the uncertainties on the criteria measurements and the DMs' preferences.

Scaling of the criteria affects the rank acceptability indices. Therefore scaling must not be done abritrarily when trying to classify the alternatives on the basis of rank acceptability indices (Lahdelma and Salminen, 2001). For example, if the minimum and maximum criterion values are chosen as the corresponding scaling points, the possible introduction of a new alternative might change these values and, therefore, also the rank acceptability indices to a large extent (Bana e Costa, 1988).

3.1.2 Central weight vector

The central weight vector w_i^c is defined as the expected center of gravity of the favourable weight space. It is computed as a multidimensional integral

over the criteria and weight distributions as

$$w_i^c = \frac{1}{b_i^1} \int_{\xi \in \chi} f_\chi(\xi) \int_{w \in W_i(\xi)} f_W(w) w \, dw \, d\xi.$$

The central weight vector describes the preferences of a typical DM supporting this alternative with the assumed preference model. By presenting the central weight vectors to the DMs, an inverse approach for decision support can be applied: instead of eliciting preferences and building a solution to the problem, the DMs can learn what kind of preferences lead into which actions, without providing any preference information.

3.1.3 Confidence factor

The confidence factor p_i^c is defined as the probability for an alternative to be the preferred one with the preferences expressed by its central weight vector. It is computed as a multidimensional integral over the criteria distributions as follows,

$$p_i^c = \int_{\substack{\xi \in \chi : u(\xi_i, w_i^c) \geq u(\xi_k, w_i^c) \\ \forall k = 1, \ldots, m}} f_\chi(\xi) \, d\xi.$$

The confidence factors measure whether the criteria measurements are accurate enough to discern the efficient alternatives. If the problem formulation is to choose an alternative to realize, the ones with low confidence factors should not be chosen. If they are deemed as attractive ones, more accurate criteria data should be collected in order to make a reliable decision.

3.2 Preference information

In most decision-making problems it is possible to elicit some preference information from the DMs. This information can possibly be imprecise and uncertain. Although SMAA methods allow preference information to be represented with an arbitrary density function, it is usually easier to elicit the preferences as constraints for the weight space. Then, the density function is defined with a uniform distribution in the restricted weight space W' as

$$f_{W'}(w) = \begin{cases} 1/\mathrm{vol}(W'), & \text{if } w \in W', \\ 0, & \text{if } w \in W \setminus W'. \end{cases}$$

In particular, we can have the following types of constraints (Lahdelma and Salminen, 2001):

1. Intervals for weights ($w_j \in [w_j^{min}, w_j^{max}]$).

2. Intervals for weight ratios (trade-offs) ($w_j / w_k \in [w_{jk}^{min}, w_{jk}^{max}]$).

3. Linear inequality constraints for weights ($Aw \leq c$).

4. Nonlinear inequality constraints for weights ($f(w) \leq 0$).

5. Partial or complete ranking of the weights ($w_j > w_k$).

Figure 3.2 illustrates the feasible weight space of a 3-criteria problem with interval constraints for weight w_1. Figure 3.3 illustrates the feasible weight space of a 3-criteria problem with complete ranking of the weights.

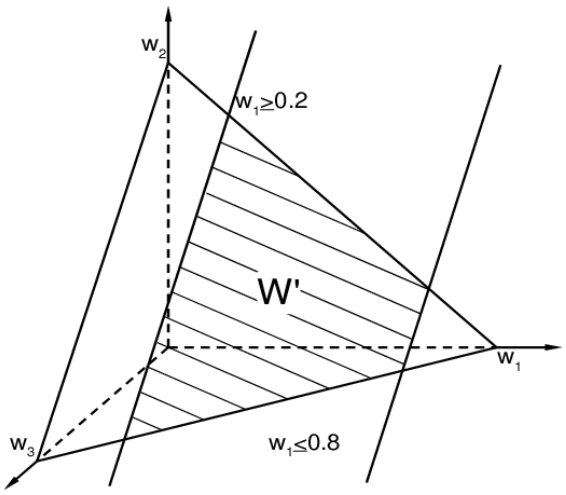


Figure 3.2: The feasible weight space of a 3-criteria problem with constraints on w_1.

When there are multiple DMs, the constraints have to be aggregated before applying. Possible non-interactive aggregation techniques include forming union or intersection, or averaging weight space density functions of different DMs. There exists also a technique based on belief functions for eliciting and aggregating the preference information, see Tervonen et al. (2004a,b).

3.3 Ordinal criteria (SMAA-O)

SMAA-O (Lahdelma et al., 2003) extends SMAA to consider ordinal criteria measurements, meaning that the DMs have ranked the alternatives according to each (ordinal) criterion. In SMAA-O, the ordinal information is mapped to cardinal without forcing any specific mapping. This means that nothing is assumed about the weights of criteria ranks in the piecewise linear mapping.

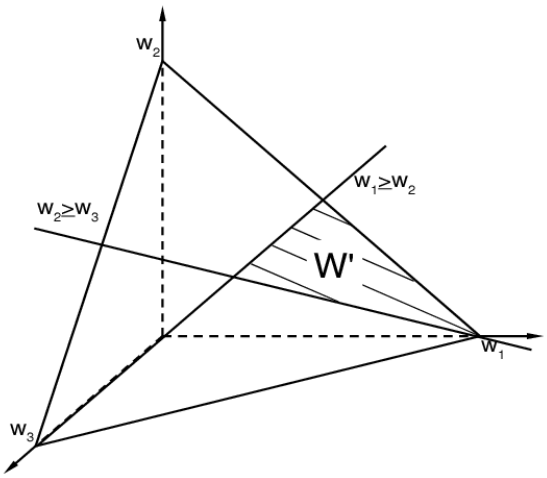


Figure 3.3: The feasible weight space of a 3-criteria problem with complete ranking of the weights.

The possibility of using ordinal measurements has its advantages. Usually the experts defining criteria measurements can rank alternatives with respect to each criterion faster than if they use cardinal measurements. Therefore, if ordinal measurements provide sufficient accuracy for the decision-making problem in question, savings can be obtained.

Ordinal criteria are measured by assigning for each alternative a rank level number $r_j = 1, \ldots, j^{max}$, where 1 is the best and j^{max} the worst rank level. Alternatives considered equally good are placed on the same rank level and rank levels are numbered consecutively. On an ordinal scale, the scale intervals do not contain any information, and should be therefore treated as such without imposing any extra assumptions. However, some mapping can be assumed to underlie the ordinal information. In SMAA-O, all mappings that are consistent with the ordinal information are simulated numerically during Monte Carlo iterations. This means generating random cardinal values for the corresponding ordinal criteria measurements in a way that preserves the ordinal rank information. Figure 3.4 illustrates a sample mapping generated in this way.

The MAUT-based SMAA methods can be used with any kind of utility function jointly accepted by the DMs, but if we have an additive utility function, the shape of the function will be considered unknown. In this case, the DMs partial utility functions are simulated in the same way as the ordinal to cardinal mappings. However, simulation is not necessary for ordinal criteria, because the simulated cardinal values can be interpreted directly as partial values on a linear scale. Therefore, if the DMs accept

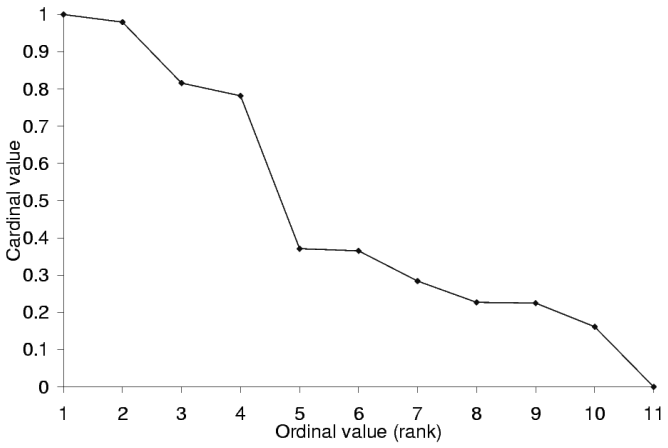


Figure 3.4: A sample ordinal to cardinal mapping of SMAA-O. (Lahdelma et al., 2003)

an additive utility function, it is not necessary for the DMs to agree on a common shape of the partial utility functions for the ordinal criteria.

SMAA-O has been combined with the so-called SWOT methodology in the work of Kangas et al. (2003b). For an alternative technique for applying ordinal criteria in simulation-based approaches, see Leskinen et al. (2004).

3.4 Simulation

The various distributions applied in the integrals of SMAA vary according to the application and can be arbitrarily complex. Usually the integrals have high dimensionality as well. Numerical integration techniques based on discretizing the distributions with respect to each dimension are infeasible, because the required effort depends exponentially on the number of dimensions. Therefore, instead of trying to obtain exact values for the integrals, Monte Carlo simulation is applied to obtain sufficiently accurate approximations. In this section we address the simulation technique, accuracy of the computations, and the complexity issues. For a full description of the algorithms, see [I].

3.4.1 Simulation technique

Monte Carlo simulation is applied in computation of the integrals. For all the acceptability index-type measures, a similar technique is applied: in each iteration, measurements for the parameters (criteria measurements, weights, ...) are drawn from their corresponding joint distributions, and a

ranking or a classification is built based on these values. After this, counters for the corresponding ranks or classes with respect to the alternatives are increased. After a number of iterations, the indices are obtained by dividing the counters with the number of iterations. The central weights are computed in a similar fashion, so that in each iteration, when an alternative obtains first rank, the weight vector is added to its "summed weight vector". This vector is divided component-wise in the end by the number of iterations to obtain the central weight vector.

Weight generation is an important part of the simulation technique. If there is no preference information available, the n uniform distributed weights are generated as follows: first $n-1$ independent random numbers are generated from the uniform distribution within the range $[0,1]$, and sorted into ascending order $(q_1,\ldots,q_{n-1})$. After that, 0 and 1 are inserted as the first (q_0) and last (q_n) numbers, respectively. The weights are then obtained as intervals between consecutive numbers $(w_j = q_j - q_{j-1})$.

If there exists preference information, the weight generation technique must be altered. In the case of complete ordinal preference information, the weights can simply be sorted according to the ranking. Lower bounds for weights can be handled by using a simple transformation technique, because the lower-bounded feasible weight space is homomorphic with the original one. The lower bounded weights are defined by generating the random numbers from interval $[0, 1-s]$, where s is the sum of all lower bounds, and adding to them the corresponding lower bounds.

Upper bounds for weights cannot be handled with a similar technique, but instead a simple rejection technique is applied, in which the weight vectors not satisfying the upper bounds are rejected. The tip of the simplex cut off by the upper bounds has relatively small area compared to the one of lower bounds. Therefore the increase in computational complexity due to upper bounds is relatively low. In addition, lower bounds might even render some of the upper bounds redundant. Consider for example a 3-criteria problem with lower bounds of 0.2 for all weights. The maximum value that any weight can obtain is $1-0.2-0.2=0.6$, and therefore all upper bounds higher than 0.6 are redundant. The amount of weights rejected due to upper bounds can be estimated in the following way: if we consider all weights to have a common upper bound w^{max}, the probability for the largest of the generated weights to exceed the upper bound is

$$\begin{aligned} P[max\{w_j\} > w^{max})] =& n(1-w^{max})^{n-1} - \binom{n}{2}(1-2w^{max})^{n-1} \\ &+ \cdots (-1)^{k-1}\binom{n}{k}(1-kw^{max})^{n-1}\cdots , \end{aligned}$$

where the series continues as long as $1-kw^{max} > 0$ (David, 1970).

3.4.2 Accuracy of computations

Accuracy of computations can be calculated by considering the Monte Carlo simulations as point estimators for the descriptive measures. To achieve accuracy of A with 95% confidence for the rank acceptability indices, we need the following number of Monte Carlo iterations K (Milton and Arnold, 1995):

$$K = \frac{1.96^2}{4A^2}.$$

For example, to achieve 95% confidence on error limits of ± 0.01 for the rank acceptability indices, we need to execute 9604 Monte Carlo iterations. The accuracy of confidence factors depends on the accuracy of central weight vectors in a complicated manner, but if we disregard this source of error, the same equation for accuracy applies. The accuracy of the central weight vectors depends on the acceptability indices, and the required amount of iterations is calculated as follows:

$$K = \frac{1.96^2}{b_i^1 4A^2}.$$

It should be noted that the accuracy of the computations does not depend on the dimensionality of the problem, but only on the number of iterations.

3.4.3 Complexity issues

The required number of Monte Carlo iterations in typical SMAA applications is fairly high, and therefore for having practical applicability the complexity of SMAA computations should not be too high with respect to the number of criteria and alternatives. The complexity of SMAA-2 and SMAA-O has been analyzed in [I]. The complexity of computing the acceptability indices and central weight vectors with independent criteria measurements and cardinal criteria is $O(K \cdot (n \log(n) + m \cdot n + m \log(m)))$. The complexity of computing the confidence factors is $O(K \cdot m^2 \cdot n)$. In these formulas K is the number of Monte Carlo iterations, m the number of alternatives, and n the number of criteria.

The use of ordinal criteria adds to the complexity with a factor of $\log(m)$. In practice this has very little effect. What has a larger impact to the running times is the handling of preference information. The formulas above assume that there are no constraints on the weights, which in practice is usually not the case. As described in Section 3.4.1, lower bounds for weights do not affect the complexity of the weight generation, but upper bounds might have a great impact on it.

3.5 Application: elevator planning

In modern high-rise buildings workers and inhabitants are transported between floors mainly by means of multiple elevators. Elevators are usually operated by elevator group control systems in order to provide efficient transportation. When a high-rise building is designed, a suitable configuration for the elevator group has to be designed. The DMs should consider performance as well as price and other non-performance criteria of alternative elevator group configurations. Because analytical methods are limited to the up-peak traffic situation and cannot evaluate the effect of a group control algorithm, the performance has to be measured using computer simulation, which produces stochastic measurements for the performance criteria of alternative configurations. The performance of an elevator group can be measured using several criteria, such as the average waiting time or the average ride time of the passengers. The price and other non-performance criteria can usually be assessed with sufficient accuracy or by ranking the alternatives. We present here an application of SMAA in elevator planning. For full details of the application, see [II]. For more on the history of elevator planning, see e.g. Basset (1923); Browne and Kelly (1968); Morley (1962); Parlow (1966); Phillips (1966); Pinfold (1966); Strakosch (1967); Tregenza (1971).

The goal in elevator planning is to find a suitable elevator group to serve the traffic of a high-rise building. Because the buildings do not exist at the planning stage, the traffic must be estimated by using the building specifications: the number of floors, their heights, the floor area and the building type. The travel height can be calculated from the number of floors and their heights, and the total population can be estimated according to the type of building and the floor area. Building types have characteristic *traffic profiles*. For example, office buildings typically have up-peak traffic in the morning when employees enter the building, intense two-way or inter-floor traffic during the lunch time, and down-peak traffic when employees exit the building (Siikonen and Leppälä, 1991).

The performance of a group of elevators is mainly determined by the number and size of the cars and their speed. Also acceleration, door types and the group control algorithm affect performance. Usual performance criteria are the handling capacity and the interval calculated in the up-peak situation. The *up-peak handling capacity* is the percentage of population per five minutes that can be transported from the lobby to the upper floors. It is assumed that elevators are filled to 80% of rated load (although it is possible to fill elevator up to rated load that does not happen in practice). The (up-peak) *interval* is an interval between two starts from the lobby. The interval is also related to the waiting time. The up-peak is used since it is the most demanding situation considering elevator handling capacity

at least in office buildings, and because there are analytical formulas for calculating the up-peak handling capacity and interval (Barney and dos Santos, 1985). The usual recommendations state that the up-peak handling capacity for an office building should be 11-17% and interval 20-30s (Barney et al., 1998).

Non-performance criteria, such as cost and occupied floor area should also be considered. The *cost* of an elevator system consists of build and maintenance costs. The *floor area* occupied by the elevator group consists of the shaft space and the waiting area for passengers. In high-rise buildings the population is large and distances are long, thus the portion of shafts is large compared to the total floor area. This means more costs, since the rentable area is reduced. In some cases the building design constraints the occupied area, sometimes there is more freedom to use space. The elevator planning is not independent of building design; the architect should take advice from the elevator planner.

Instead of considering only up-peak traffic, we take into account the entire daily traffic and consider all criteria simultaneously. In this study the following 6 criteria are considered. The cost and area criteria take into account the building owners point of view. Passengers point of view is taken into account by waiting time, journey time, the percentage of waiting times exceeding 60s, and the percentage of journey times exceeding 120s. The waiting time is measured from the moment a passenger enters the waiting area to the moment he/she enters the elevator. The journey time is the total time from entering the waiting area to exiting the elevator. The last two criteria measure unsatisfactory service, which may happen especially in intense traffic peaks.

To obtain stochastic criteria measurements for the performance criteria, we executed simulations with the KONE Building Traffic Simulator (Hakonen, 2003; Leinonen, 1999). The simulation model consists of the elevator model and traffic generation. For more details of the model, see [II]. The simulated building has a lobby floor and 19 populated floors. The estimated number of people is 60 per floor.

Figure 3.5 shows the intensities of incoming, outgoing and inter-floor passengers during the day from 7 a.m. to 7.15 p.m. The traffic profile is measured from an office building. The profile shows typical morning, lunch time and afternoon traffic peaks. When passengers are generated according to the traffic profile, the expected number of passengers is 11502. Since total population of the building is uncertain, the traffic is varied between 80% and 120% of forecasted traffic. With these parameters, we generated 21 traffic situations according to the traffic profile. The same passengers were used for all 10 alternatives in order to reduce the covariance between the measurements of different alternatives.

The number of elevators in the alternatives varied between 6 and 8,

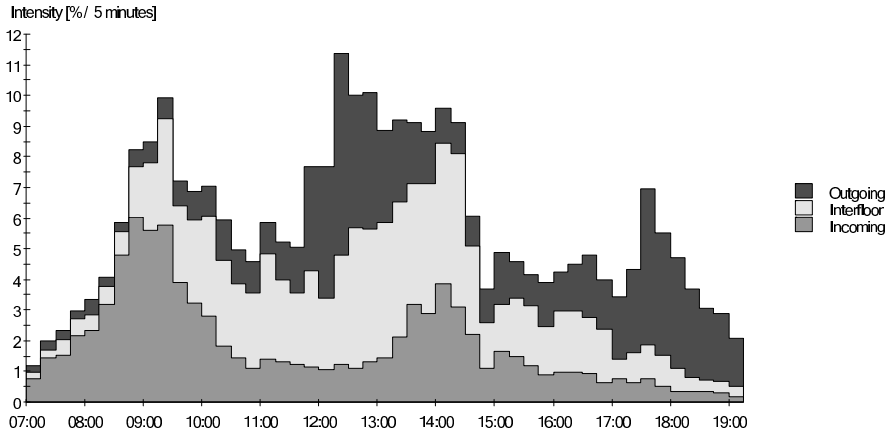


Figure 3.5: Traffic profile of the simulated building. Siikonen and Leppälä (1991)

rated load from 13 to 24 and speed from 3.5 m/s to 5 m/s. Area is the shaft space plus waiting area space. The exact costs were unknown, but alternatives could be ranked with respect to the cost. All alternatives were feasible with respect to up-peak handling capacity and interval.

The uncertainties of the performance criteria were assessed based on the simulations for each of the 10 configurations. Based on the simulation results we estimated the parameters for a multivariate Gaussian distribution, i.e. the expected value of each criteria measurement and the covariance matrix for the uncertainty dependencies. The uncertainties of the performance criteria were quite dependent, with multivariate correlations in the interval [0.8,1]. The cost was modelled as an ordinal criterion (see Section 3.3), because exact price information was not available. The required floor area was measured on a cardinal scale with 5m^2 uncertainty for all alternatives.

Preference information was added to the model in form of weight bounds to the model; weights for cost and shaft space were constrained to be in the interval [0.1,1]. The preference information was added to the model because of the strong dependencies between performance criteria, which shows that they all ultimately measure a single criterion, performance from the passengers point of view. Because of the additivity of weights, the performance would obtain too high significance in the analysis without balancing accomplished by using weight constraints.

We analyzed the model using 100 000 Monte-Carlo iterations, which gives error limits ≤ 0.01 ([I]). For results of the SMAA computations, see

[II]. Rank acceptability indices are illustrated graphically in Figure 3.6, and central weights as stacked columns in Figure 3.7. The analysis of this application allowed directly to eliminate half of the alternatives based on their confidence factors. The rank acceptability indicated four good choices for the alternative to implement. The trade-offs between the four alternatives could be stated based on the central weight vectors. From these four alternatives, one was recommended as a good compromise solution.

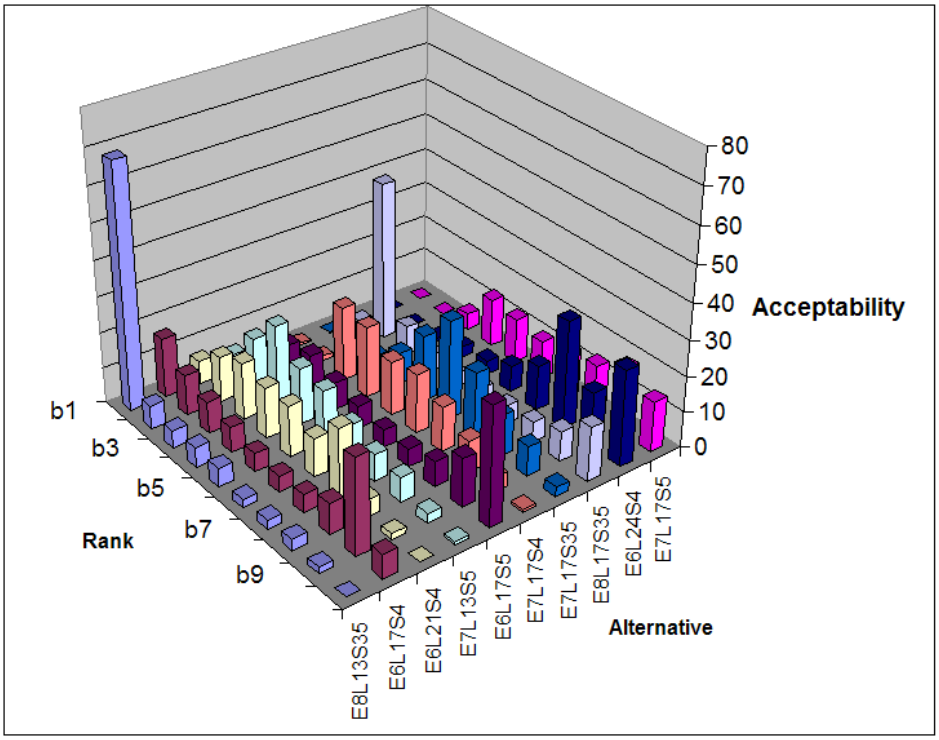


Figure 3.6: Rank acceptability indices of the study.

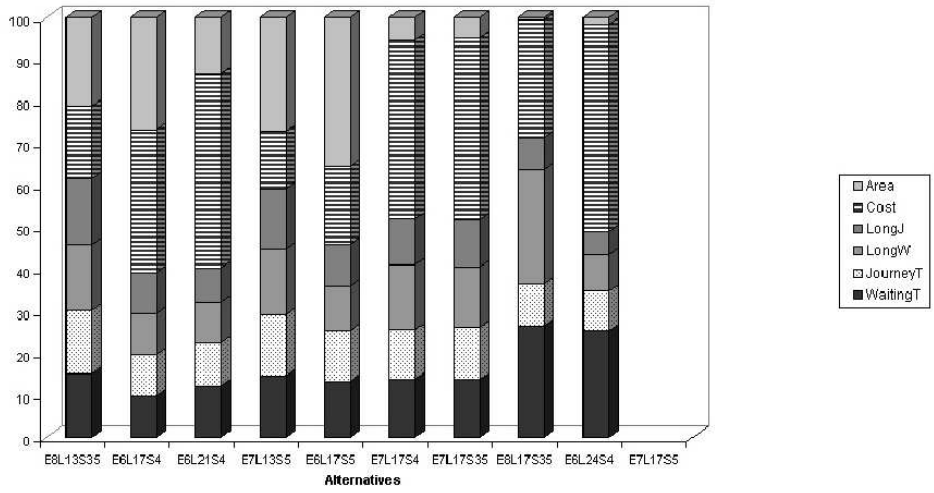


Figure 3.7: Central weight vectors of the study.

Chapter 4

Outranking based SMAA approaches

SMAA has been extended for using instead of utility function (3.1) an outranking-based aggregation procedure to rank alternatives. This and other approaches described in this chapter are based on using ELECTRE type pseudo-criteria. The pseudo-criteria are defined by using thresholds that are denoted as follows:

- $q_j(g_j(\cdot))$ is the *indifference threshold* for criterion g_j,
- $p_j(g_j(\cdot))$ is the *preference threshold* for criterion g_j, and, finally,
- $v_j(g_j(\cdot))$ is the *veto threshold* for criterion g_j.

By using these thresholds a *concordance index* is defined. It is computed by considering individually for each criterion g_j the support it provides for the assertion of the outranking aS_jb, "alternative a is at least as good as alternative b". The partial concordance index is a fuzzy index computed as follows, for all $j = 1, \ldots, n$:

$$c_j(a,b) = \begin{cases} 1, \text{ if } g_j(a) \geq g_j(b) - q_j(g_j(b)), \\ 0, \text{ if } g_j(a) < g_j(b) - p_j(g_j(b)), \\ \frac{g_j(a)+p_j\left(g_j(b)\right)-g_j(b)}{p_j\left(g_j(b)\right)-q_j\left(g_j(b)\right)}, \text{ otherwise.} \end{cases}$$

After computing the partial concordance indices, a comprehensive concordance index is calculated as follows,

$$c(a,b) = \sum_{j=1}^{n} w_j c_j(a,b).$$

If veto thresholds are used, a *discordance index* can be defined also. For more information on pseudo-criteria based models, see Roy (1996).

4.1 Outranking aggregation procedure (SMAA-3)

SMAA-3 (Hokkanen et al., 1998) method is a variant of the original SMAA that applies, instead of the utility function, ELECTRE type pseudo-criteria and a maximin choice procedure (see Pirlot, 1995). According to this procedure, an alternative becomes the preferred one (not necessary unique) if the following set of constraints hold:

$$\min_{l=1,\ldots,m,l\neq i} c(x_i, x_l) \geq \min_{l=1,\ldots,m,l\neq k} c(x_k, x_l),$$
$$k = 1, \ldots, m, k \neq i.$$

Based on this the favourable weights of an alternative are defined as

$$W_i = \{w \in W : \min_{l=1,\ldots,m,l\neq i} \sum_{j=1}^{n} w_j c_j(x_i, x_l)$$
$$\geq \min_{l=1,\ldots,m,l\neq k} \sum_{j=1}^{n} w_j c_j(x_k, x_l),$$
$$k = 1, \ldots, k, k \neq i\}.$$

Based on these, the analysis is done in a way similar to SMAA, with the exception that the criteria measurements are considered to be deterministic (no integration over χ is done), and therefore no confidence factors are computed. It should be noted that now the central weight vector can lie outside the space of favourable weights of an alternative, because this preference model is non-linear. In this kind of (easily detectable) situations a favourable weight vector is chosen with a minimal distance to the central weight vector.

In the literature there exists simulation-tests of SMAA against SMAA-3. In these tests the results of SMAA-3 were found to be quite unstable with respect to the indifference threshold (Lahdelma and Salminen, 2002). Therefore, when SMAA-3 is applied in practice, great care should be put into choosing the thresholds. These test results are confirmed in [IV].

4.2 SMAA-TRI

All the SMAA variants described until here are for ranking or choosing problem statements. ELECTRE TRI (Yu, 1992a) is a method for sorting problem statements, and SMAA-TRI extends it to allow ignorance on the parameter values. There exists a large amount of work on parameter inference and robustness analysis for ELECTRE TRI, see Dias and Clímaco

(1999, 2000); Dias and Mousseau (2006); Dias et al. (2002); Mousseau et al. (2004, 2003, 2001); Mousseau and Słowiński (1998); Mousseau et al. (2000); Ngo The and Mousseau (2002).

ELECTRE TRI uses concordance and discordance indices for sorting the alternatives into pre-defined and ordered categories. Let us denote the categories in ascending preference order $C_1, \ldots, C_h, \ldots, C_k$ (C_1 is the "worst" category). These categories are defined by upper and lower profiles that consist of measurements for all criteria. In the assignment procedure alternatives are iteratively compared with the profiles. The profiles are denoted $p_1, \ldots, p_h, \ldots, p_{k-1}$. p_h is the upper limit of category C_h and the lower limit of category C_{h+1}. The profiles have to be strictly ordered, that is, they have to satisfy

$$p_1 \,\Delta\, p_2 \,\Delta\, \ldots \,\Delta\, p_{k-2} \,\Delta\, p_{k-1}, \tag{4.1}$$

where Δ is the dominance relation ($p_1 \Delta p_2$ means that p_2 dominates p_1). This dominance relation needs to be interpreted in a wide sense, because domination depends not only on the values of components of the two profiles, but also on the threshold values. We will not describe here the assignment procedure. It requires an additional technical parameter, the lambda cutting level, to be defined. The interested reader should refer to [III].

SMAA-TRI is developed for parameter stability analysis of ELECTRE TRI, and consists of analyzing finite spaces of arbitrarily distributed parameter values in order to describe for each alternative the share of parameter values that assign it to different categories. It analyzes the stability of weights, profiles, and the cutting level.

The input for ELECTRE TRI in SMAA-TRI is the following:

1. Uncertain or imprecise profiles are represented by stochastic variables ϕ_{hj} with joint density function $f_\Phi(\phi)$ in the space $\Phi \subseteq R^{(k-1)\times n}$. The joint density function must be such that all possible profile combinations satisfy (4.1). Usually the category profiles are defined to be independently distributed, and in this case the distributions must not overlap. For example, if the profile values for a criterion are Gaussian distributed, the distributions must have tails truncated as shown by the vertical lines in Figure 4.1.

2. The lambda cutting level is represented as a stochastic variable Λ with density function $f_L(\Lambda)$ defined within the valid range [0.5,1].

3. The weights and criteria measurements are represented as in SMAA-2.

4. The data and other parameters of ELECTRE TRI are represented by the set $T = \{M, q, p, v\}$. These components are considered to have deterministic values.

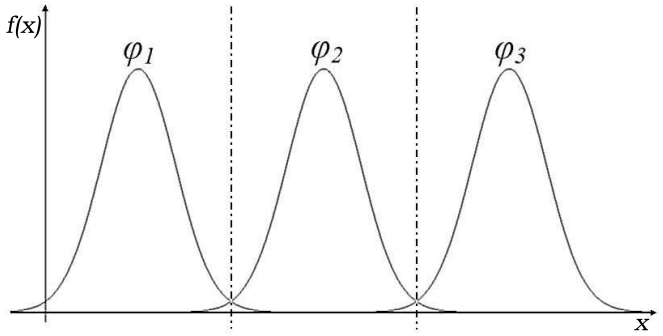


Figure 4.1: Probability distribution functions for three Gaussian distributed profile values (for a single criterion). The vertical lines show where the tails of the distributions must be truncated.

SMAA-TRI produces category acceptability indices for all pairs of alternatives and categories. The category acceptability index π_i^h describes the share of possible parameter values that have an alternative x_i assigned to category C_h. Let us define a *categorization function* that evaluates the category index h to which an alternative x_i is assigned by ELECTRE TRI:

$$h = K(i, \Lambda, \phi, w, T),$$

and a category membership function

$$m_i^h(\lambda, \phi, w, T) = \begin{cases} 1, \text{ if } K(i, \Lambda, \phi, w, T) = h, \\ 0, \text{ otherwise,} \end{cases}$$

which is applied in computing the category acceptability index numerically as a multi-dimensional integral over the finite parameter spaces as

$$\pi_i^h = \int_{0.5}^{1} f_L(\Lambda) \int_{\Phi} f_\Phi(\phi) \int_{W} f_W(w) m_i^h(\Lambda, \phi, w, T)\, dw\, d\phi\, d\Lambda.$$

The category acceptability index measures the stability of the assignment, and it can be interpreted as a fuzzy measure or a probability for membership in the category. If the parameters are stable, the category acceptability indices for each alternative should be 1 for one category, and 0 for the others. In this case the assignments are said to be robust with respect to the imprecise parameters.

The software presented in Chapter 6 implements SMAA-TRI, but also allows imprecision in all parameters. This means that the category acceptability indices are computed by integrating through spaces of all feasible parameter values instead of only the spaces of feasible values for lambda, profiles, and weights.

4.3 SMAA-III

ELECTRE III (Roy, 1978) is designed for solving a discrete ranking problem as it was defined for SMAA-2. Similarly to the other ELECTRE family methods, ELECTRE III is based on two phases. In the first phase, an outranking relation between pairs of alternatives is formed. The second phase consists of exploiting this relation, producing a final partial pre-order and a median pre-order.

The exploitation of the outranking relation consists of two phases. In the first phase, two complete pre-orders, Z_1 (descending) and Z_2 (ascending) are constructed with the so-called distillation procedures. In the second phase, a final partial pre-order or a complete median pre-order is computed based on these two pre-orders. In the original ELECTRE III, a median pre-order is computed based on the two complete pre-orders, Z_1 and Z_2, and the final partial pre-order.

In SMAA-III, the weights are represented as in the other SMAA methods. Imprecise thresholds are represented by stochastic functions $\alpha_j(\cdot)$, $\beta_j(\cdot)$, and $\gamma_j(\cdot)$, corresponding to the deterministic thresholds $p_j(\cdot)$, $q_j(\cdot)$, and $v_j(\cdot)$, respectively. To simplify the notation, we define a 3-tuple of thresholds $\tau = (\alpha, \beta, \gamma)$. It has a joint density function f_T in the space of possible values defining the functions. It should be noted that all feasible combinations of thresholds must satisfy $q_j(x_i) < p_j(x_i) < v_j(x_i)$.

Traditionally the thresholds in ELECTRE models have been used to model preferences of the DMs (e.g. differences deemed significant) as well as imprecision in the data. But it has been shown that the indifference threshold does not correspond to a linear imprecision interval (Lahdelma and Salminen, 2002). Therefore, in SMAA-III thresholds are used only to model preferences (together with weights). Imprecision in the criteria measurements is modelled with stochastic variables as in SMAA-2 (see Section 3.1).

Incomparabilities between alternatives can be present in the final results of ELECTRE III. This is one of the main features of ELECTRE methods in comparison with the methods applying classical multi-attribute utility theory (see Keeney and Raiffa, 1976). Incomparability is considered by the researchers and practitioners of ELECTRE methods as one of the strongest points of the methodology because it avoids to force comparison of very heterogenous alternatives. In the late seventies, it was considered a very important theoretical advance. But when dealing with practical situations, incomparabilities in the final result are sometimes inconvenient. This aspect was soon observed (Roy et al., 1986) and complete pre-orders and median pre-orders were proposed to be used in side of the partial pre-orders. SMAA-III applies median pre-orders in computing rank acceptability indices. The only information lost in using the median pre-order as the

primary measure of ranking is the incomparability. It will be retained by representing it with another index.

Monte Carlo simulation is used in SMAA-III to compute three types of descriptive measures: rank acceptability indices, pair-wise winning indices, and incomparability indices. In order to compute these indices, let us define a *ranking function* that evaluates the rank r of the alternative x_i with the corresponding parameter values:

$$rank(i, w, \xi, \tau).$$

The evaluation of this function corresponds to executing ELECTRE III and returning rank of the corresponding alternative in the resulting median pre-order.

4.3.1 Rank acceptability index

The rank acceptability index, b_i^r, measures the share of feasible weights that grant alternative x_i rank r in the median pre-order by taking into account simultaneously imprecisions in all parameters and criterion evaluations. It represents the share of all feasible parameter combinations that make the alternative acceptable for a particular rank, and it is most conveniently expressed percentage-wise.

The rank acceptability index b_i^r is computed numerically as a multidimensional integral over the spaces of feasible parameter values as

$$b_i^r = \int_{W:rank(i,w,\xi,\tau)=r} f_W(w) \int_X f_X(\xi) \int_T f_T(\tau)\, dT\, dw\, d\xi.$$

The rank acceptability index has the same meaning as in SMAA-2.

4.3.2 Pair-wise winning index

The pair-wise winning index (Leskinen et al., 2006), o_{ik}, describes the share of weights that place alternative x_i on a better rank than alternative x_k. An alternative x_i that has $o_{ik} = 1$ for some k always obtains a better rank than alternative x_k, and can thus be said to *dominate* it.

The pair-wise winning index o_{ik} is computed numerically as a multidimensional integral over the space of weights that give alternative a lower rank than for another.

$$o_{ik} = \int_{w\in W:rank(i,w,\xi,\tau)<rank(k,w,\xi,\tau)} f_W(w) \int_X f_X(\xi) \int_T f_T(\tau)\, dT\, dw\, d\xi.$$

The pair-wise winning indices are especially useful when trying to distinguish between the ranking differences of two alternatives. Because the

number of ranks in the median pre-order of different simulation runs varies, two alternatives might obtain similar rank acceptabilities although one is in fact inferior. In these cases looking at the pair-wise winning indices between this pair of alternatives can help to determine whether one of the alternatives is superior to the other or if they are equal in "goodness".

4.3.3 Incomparability index

Because median pre-orders are used in computing the rank acceptability indices, it is not anymore possible to model incomparability. As some DMs might be accustomed to make decisions also based on incomparabilities, another index is introduced. Incomparability index ρ_{ik} measures the share of feasible parameter values that cause alternatives x_i and x_k to be incomparable. For this reason, we define the incomparability function:

$$R(i,k,\xi,\tau) = \begin{cases} 1, \text{ if alternatives } x_i \text{ and } x_k \text{ are judged incomparable,} \\ 0, \text{ if not.} \end{cases}$$

This function corresponds to a run of ELECTRE III with the given parameter values and checking if the alternatives are judged incomparable in the final partial pre-order. In practice we do not compute the final partial pre-order, because this information can be extracted from the two partial pre-orders Z_1 and Z_2 as shown in [IV]. By using the incomparability function, the incomparability index is computed numerically as a multidimensional integral over the feasible parameter spaces as

$$\rho_{ik} = \int_W f_W(w) \int_X f_X(\xi) \int_T f_T(\tau) R(i,j,\xi,\tau)\, dT\, dw\, d\xi.$$

Chapter 5

Framework

We define now a SMAA framework to decide a method to choose on a specific decision making context. The first question to ask is whether we are dealing with a ranking or a sorting problem. If we are dealing with a sorting one, the only method of the SMAA family we can use is SMAA-TRI. With ranking problems, we have to choose the type of preference model we have: whether it is based on weights or on reference points. If we have a weight-based model, we have to choose the type of aggregation procedure: utility function or outranking method. With the reference point approach we use Ref-SMAA (see [V] or Lahdelma et al. (2005)). For utility function we use SMAA-2. With outranking model we can choose between SMAA-3 and SMAA-III. With all this information, we can choose whether to apply SMAA-2, SMAA-3, SMAA-III, or Ref-SMAA for the ranking problem. Depending on the method to apply, we obtain as output different descriptive measures that can be used to derive "second-order" aggregate measures. Choice of the method is presented as a decision-tree in Figure 5.1.

Other way to choose the method for a ranking problem is to question what kind of information is not available. Are the DMs willing to provide a shape for the utility function? If not, SMAA-2 can not be applied. Same type of questions can be posed with respect to other parameters of the methods in order to find out which method would be the most suitable.

In the context of this framework, we should notice that all other methods than Ref-SMAA, which is based on reference points, can be used with arbitrary weight information. This means that we can apply them with no preference information at all, as well as with mixed information of ordinal and cardinal types. In practice, the most useful ones are (partial) ordinal information and cardinal weight constraints. Complex weight constraints might be hard for the DMs to understand, and therefore by using more complex distributions the possibility for the information to contain uncertainty increases. If the DMs have problems understanding the underlying

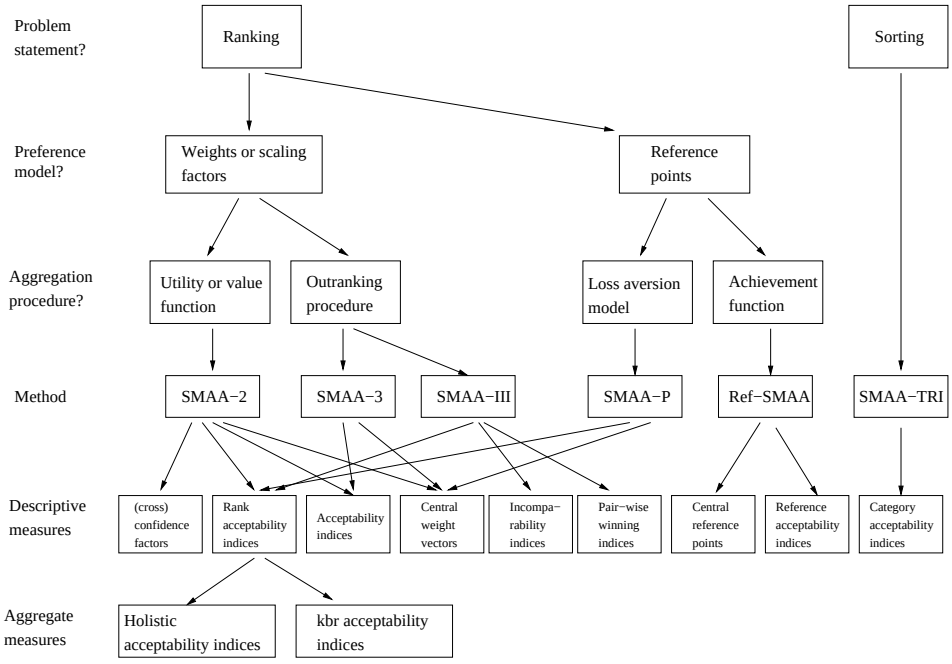


Figure 5.1: Decision-tree to choose the SMAA variant.

preference model, the achievement function based approach (Ref-SMAA) might be more suitable.

The shortcoming of the utility-function based approach (SMAA-2) is that the scaling has large effect on the results, and the meaning of the weights is based on the scale. Therefore, if the shape of the utility function is hard to define, it might be more suitable to use SMAA-3 or SMAA-III instead.

Arbitrarily distributed imprecise or uncertain criteria can be applied in all methods of the family except SMAA-3 that requires criteria measurements to have imprecision defined through thresholds. It should be noted that SMAA-O is not a stand-alone method, but rather a computational technique to handle ordinal criteria measurements. The possibility of using external sampling and the following generalisation to use SMAA with external methods can be considered a great advantage. For example, the approach applied in SMAA-TRI and SMAA-III can probably be applied to other methods as well, to use them with ignorance on the parameter values in order to analyze the stability of the results.

Chapter 6

Software

A user-friendly software is of crucial importance if an MCDM method is to enjoy a wide audience. A software was developed to allow users less accustomed in the field of numerical computation to use the new methods developed in this thesis. It was programmed in the C++ language and utilizes the gtkmm graphical user interface library (http://www.gtkmm.org) to be portable to various operating systems. Currently there exists versions for Linux, Mac OS X, and Windows XP.

The software implements SMAA-TRI and SMAA-III methods. It allows the SMAA-III model to be defined with uniform distributed, Gaussian distributed, or ordinal criteria. Ordinal criteria are not allowed for SMAA-TRI models. Ordinal criteria are modelled in SMAA-III through discrete rank values and setting indifference and preference thresholds to 0 and 1, respectively. Thresholds for cardinal criteria can have exact values or can be defined as intervals that can be absolute or a percentage of the criterion measurement in question. Criteria input screen is shown in Figure 6.1.

Criterion measurements and criteria uncertainties input screens are shown in Figures 6.2 and 6.3, respectively. The software allows to automatically set uncertainties to 5, 10, or 20 percentages of the corresponding measurement values. This allows an easy way to set up the model when using the method for an automated robustness or parameter stability analysis. The software allows three types of preferences: exact ones (expressed as exact weight values), upper and lower bounds for weights, and ordinal preferences (ranking of the criteria). The results are presented in a tabular form. While the software computes the various indices, the progress is shown interactively. Figure 6.4 presents an example of results from the model used in the case study of [III].

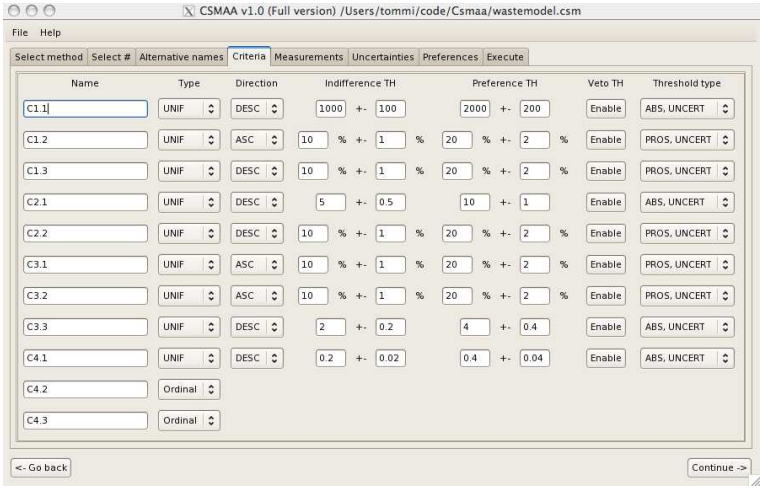


Figure 6.1: Criteria input screen in the software.

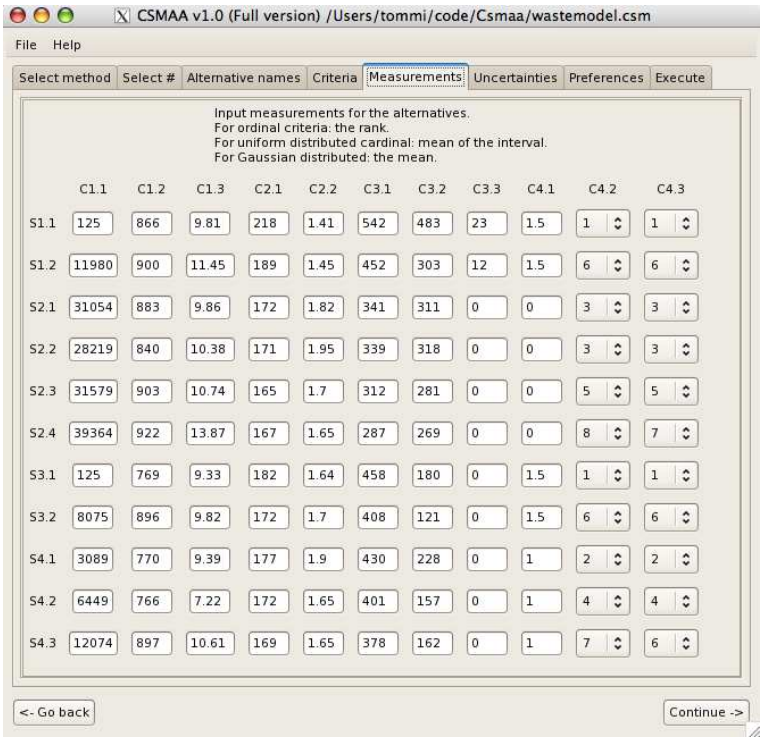


Figure 6.2: Criteria measurements input screen in the software.

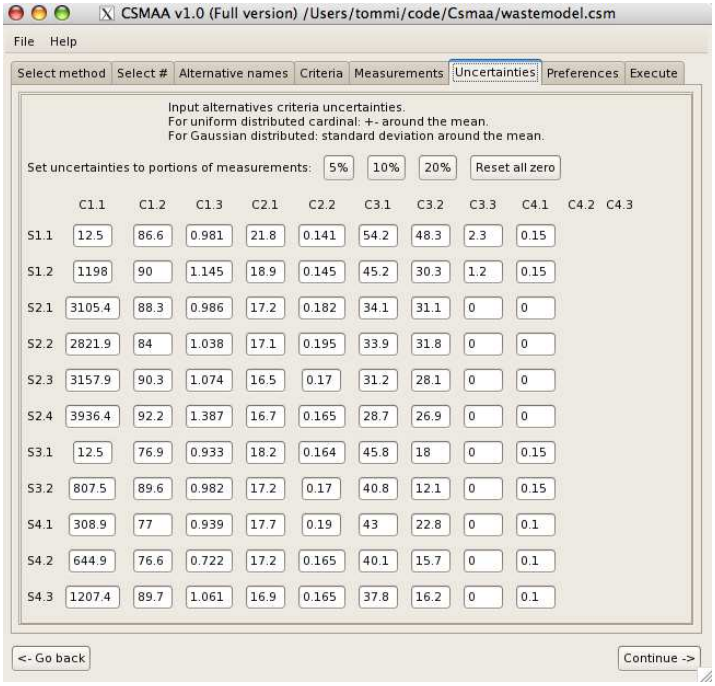


Figure 6.3: Criteria uncertainties input screen in the software.

Figure 6.4: Results screen for SMAA-TRI in the software.

Chapter 7

Summary of publications

In publication [I], we present efficient methods for performing the SMAA computations. We analyze the complexity and assess the accuracy of the presented algorithms. We perform empirical efficiency tests as well. These tests show that our SMAA implementation is fast enough to analyze typical sized discrete problems interactively within seconds, if tight upper bounds for weights are not applied.

In publication [II], SMAA-2 is applied in elevator planning. This formulates a ranking problem, in which different elevator configurations are to be ranked with respect to both performance and non-performance criteria. We compare 10 feasible elevator group configurations for a 20-floor building. We evaluate the criteria related to the service level in different traffic situations using the KONE Building Traffic Simulator, and use analytical models and expert judgements for other criteria. The performance criteria are represented by a multivariate Gaussian distribution, others by deterministic values and ordinal information.

In publication [III], a new method, SMAA-TRI, is introduced. SMAA-TRI aims to analyze the stability of ELECTRE TRI results and to derive robust conclusions when SMAA-TRI is applied. SMAA-TRI allows ELECTRE TRI to be used with imprecise, arbitrarily distributed values for weights, profiles, and the lambda cutting level. The method computes for each alternative the share of parameter values that have it assigned to different categories. We illustrate application of SMAA-TRI by re-analyzing a case study in the field of risk assessment.

In publication [IV], we present a new method, SMAA-III. It allows ELECTRE III to be applied with imprecise parameter values. By allowing imprecise values, the method also allows an easily applicable robustness analysis. In SMAA-III, simulation is used and descriptive measures are computed to characterize stability of the results. We present a software implementing the methodology and illustrate its usage by re-analyzing an

existing case study.

In publication [V], a complete survey of SMAA methodology is presented. Methods of this family allow solving MCDA problems of various types. Even though the methods have been applied in the past in various real-life decision-making situations, the structure of a unified SMAA framework has not been studied. This publication describes the methods of the family, and defines a unified SMAA framework. We also point out the key points in the methodology for future research.

Chapter 8

Concluding remarks

Decision support with Multiple Criteria Decision Making (MCDM) methods has become increasingly important for organizations of various sizes, because modern decision making situations often oblige Decision Makers (DMs) to consider several aspects of the problem and the trade-offs between them. Sometimes there are also multiple DMs whose opinions have to be taken into account. The problem settings often contain various types of uncertainties. Therefore methods that allow modelling of uncertainty in the parameters and possibly clashing preferences are needed.

Stochastic Multicriteria Acceptability Analysis (SMAA) family of methods include ways to handle various types of uncertainties and imprecision. Uncertain weights or other preference parameters can be used to model clashing or missing preferences. In this thesis, we showed how the SMAA approach can extend third-party MCDM methods to use imprecise parameters. This allows to perform an automated parameter stability analysis in addition to solving the two above-mentioned problems.

Although the possibility of defining uncertain parameters facilitates the elicitation process, the weight information should be consistent with the underlying preference model. For example, utility-theory based SMAA models should not use intervals for weights except for stability analysis. Imprecise trade-off ratios should be used instead for the weights to be consistent with the preference model. However, there does not yet exist an efficient weight generation technique for them. Future research should address this subject, and new efficient algorithms for generating weights with various types of constraints should be developed. In addition to being of importance for the SMAA methodology, they can be used with other Monte Carlo simulation applications within MCDM as well as in other disciplines.

We presented the basic SMAA method and its most important extensions. We also analyzed complexity of the algorithms and presented an application in the field of elevator planning. This application shows how

the methodology can be used to solve problems traditionally beyond the scope of MCDM. Following this, we presented the two new SMAA methods: SMAA-TRI and SMAA-III. These extend ELECTRE TRI and ELECTRE III, respectively, to allow imprecise parameter values. For an MCDM method to enjoy widespread accentance, a user-friendly software is needed. Part of the work leading to this thesis composed of programming a software implementing the SMAA-III and SMAA-TRI methods. This software was briefly presented in this thesis. Free demo versions of the software can be obtained from the author.

The comprehensive decision making process as supported by SMAA or the more traditional decision making methods differs in many aspects. In the traditional methods, the model has to be defined with exact values straight in the beginning, and elicitation of the preference parameters from DMs is usually slow. In many cases these parameters do not change dramatically with time. On the contrary, SMAA models can be defined with no preference information, and the model iterated until sufficiently precise results are obtained. This can help for a more dynamic decision making process with more space for discussion. For example, in the context of multiple DMs, usually most of the preferred alternatives of different DMs obtain some first rank acceptability. This can stimulate further discussion for redefining the parameters more precisely and for finding good compromise alternatives.

Even though SMAA methods allow flexible decision making process, they should not be used in automated decision making. The results are always somewhat vague and need to be interpreted as such. This is an important difference between the SMAA model and many other MCDM models allowing imprecise values. Although the results are more imprecise than of other methods, they explicitly show the uncertainties present in the parameters. This can lower the possibility of accepting an "incorrect" model. This is somehow the main idea of SMAA philosophy – Monte Carlo simulation is used to bring visible the consequences implicated by the uncertain data, but inside SMAA are still the traditional MCDM methods.

The current state of research in SMAA methodology is quite young and the proposed new directions in this thesis are the initial steps in diversing the methodology. This thesis has tried to bring together the somewhat heterogenous parts of the SMAA methodology. Although being applied in various real-life cases, the theoretical basis needs to be defined firmer and the different SMAA methods bound together in a consistent way. Future research should concentrate in this direction instead of developing new methods to the family.

This thesis is comprised of smaller works in various areas, mainly because of having been completed in different universities under supervision of professors from different fields. Somehow this characterizes the whole field

of MCDM: it is a synthesis of various disciplines, and for succesful research, we need economists to tell what is needed, mathematicians to provide the theoretical basis for it, as well as computer scientists for providing tools and methods to achieve the goals.

Bibliography

Abbas, A., 2006. Maximum entropy utility. Operations Research 54 (2), 277–290.

Bana e Costa, C. A., 1986. A multicriteria decision aid methodology to deal with conflicting situations on the weights. European Journal of Operational Research 26, 22–34.

Bana e Costa, C. A., 1988. A methodology for sensitivity analysis in three-criteria problems: A case study in municipal management. European Journal of Operational Research 33, 159–173.

Barney, G. C., Cooper, D. A., Inglis, J., 1998. Elevator & Escalator Micropedia. The International Association of Elevator Engineers, UK.

Barney, G. C., dos Santos, S. M., 1985. Elevator Traffic Analysis Design and Control. Peter Peregrinus Ltd, London, UK.

Basset, J., 1923. The probable number of stops made by an elevator. GE Rev 26 (8).

Belton, V., Stewart, T. J., 2002. Multiple Criteria Decision Analysis - An Integrated Approach. Kluwer Academic Publishers, Dordrecht.

Beynon, M., 2002. DS/AHP method: A mathematical analysis, including an understanding of uncertainty. European Journal of Operational Research 140, 148–164.

Beynon, M., Cosker, D., Marshall, D., 2001a. An expert system for multi-criteria decision making using Dempster Shafer theory. Expert Systems with Applications 20, 357–367.

Beynon, M., Cosker, D., Marshall, D., 2001b. An expert system for multi-criteria decision making using Dempster Shafer theory. Expert Systems with Applications 20, 357–367.

Beynon, M., Curry, B., Morgan, P., 2000. The Dempster-Shafer theory of evidence: An alternative approach to multicriteria decision modelling. Omega 28, 37–50.

Brans, J., Mareschal, B., 2005. PROMETHEE methods. In: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Verlag, Boston, Dordrecht, London, pp. 163–196.

Browne, J. J., Kelly, J. J., 2 1968. Simulation of elevator system for world's tallest buildings. Transportation Science 2 (1), 35–36.

Charnetski, J., 1973. The multiple attribute problem with partial information: The expected value and comparative hypervolume methods. Ph.D. thesis, University of Texas at Austin.

Charnetski, J., Soland, R., 1978. Multiple-attribute decision making with partial information: The comparative hypervolume criterion. Naval Research Logistics Quarterly 25, 279–288.

David, H. A., 1970. Order Statistics. Wiley and Sons, New York.

Dias, L., Clímaco, J., 1999. On computing ELECTRE's credibility indices under partial information. Journal of Multi-Criteria Decision Analysis 8 (2), 74–92.

Dias, L., Clímaco, J., 2000. ELECTRE TRI for groups with imprecise information on parameter values. Group Decision and Negotiation 9 (5), 355–377.

Dias, L., Mousseau, V., 2006. Inferring Electre's veto-related parameters from outranking examples. European Journal of Operational Research 170, 172–191.

Dias, L., Mousseau, V., Figueira, J., Clímaco, J., 2002. An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI. European Journal of Operational Research 138, 332–348.

Durbach, I., 2006. A simulation-based test of stochastic multicriteria acceptability analysis using achievement functions. European Journal of Operational Research 170 (3), 923–934.

Figueira, J., Greco, S., Ehrgott, M. (Eds.), 2005. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science+Business Media, Inc., New York.

Greco, S., Matarazzo, B., Słowiński, R., 1999. Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems. Lecture notes in artificial intelligence 1711, 146–157.

Greco, S., Matarazzo, B., Słowiński, R., 2000. Extension of the rough set approach to multicriteria decision support. INFOR 38 (3), 161–195.

Greco, S., Matarazzo, B., Słowiński, R., 2001. Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 129, 1–47.

Greco, S., Matarazzo, B., Słowiński, R., 2002. Rough sets methodology for sorting problems in presence of multiple attributes and criteria. European Journal of Operational Research 138, 247–259.

Hakonen, H., 2003. Simulation of building traffic and evacuation by elevators. Licenciate thesis, System Analysis Laboratory, Helsinki University of Technology.

Hokkanen, J., Lahdelma, R., Miettinen, K., Salminen, P., 1998. Determining the implementation order of a general plan by using a multicriteria method. Journal of Multi-Criteria Decision Analysis 7 (5), 273–284.

Hokkanen, J., Lahdelma, R., Salminen, P., 1999. A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor. Socio-Economic Planning Sciences 33, 1–23.

Hokkanen, J., Lahdelma, R., Salminen, P., 2000. Multicriteria decision support in a technology competition for cleaning polluted soil in Helsinki. Journal of Environmental Management 60 (4), 339–348.

Jessop, A., 1999. Entropy in multiattribute problems. Journal of Multi-Criteria Decision Analysis 8, 61–70.

Kangas, A., Kangas, J., Lahdelma, R., Salminen, P., 2006. Using SMAA-2 method with dependent uncertainties for strategic forest planning. Forest Policy and Economics 9, 113–125.

Kangas, J., Hokkanen, J., Kangas, A., Lahdelma, R., Salminen, P., 2003a. Applying stochastic multicriteria acceptability analysis to forest ecosystem management with both cardinal and ordinal criteria. Forest Science 49 (6), 928–937.

Kangas, J., Kangas, A., 2003. Multicriteria approval and SMAA-O method in natural resources decision analysis with both ordinal and cardinal criteria. Journal of Multi-Criteria Decision Analysis 12, 3–15.

Kangas, J., Kurttila, M., Kajanus, M., Kangas, A., 2003b. Evaluating the management strategies of a forestland estate – the S-O-S approach. Journal of Environmental Management 69, 349–358.

Kangas, J., Store, R., Kangas, A., 2005. Socioecological landscape planning approach and multicriteria acceptability analysis in multiple-purpose forest management. Forest Policy and Economics 7, 603–614.

Keeney, R., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Wiley & Sons, New York.

Lahdelma, R., Hokkanen, J., Salminen, P., 1998. SMAA – stochastic multiobjective acceptability analysis. European Journal of Operational Research 106 (1), 137–143.

Lahdelma, R., Makkonen, S., Salminen, P., 2006a. Multivariate Gaussian criteria in SMAA. European Journal of Operational Research 170 (3), 957–970.

Lahdelma, R., Makkonen, S., Salminen, P., 2006b. Two ways to handle dependent uncertainties in multi-criteria decision problems. Omega (in press).

Lahdelma, R., Miettinen, K., Salminen, P., 2003. Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA). European Journal of Operational Research 147 (1), 117–127.

Lahdelma, R., Miettinen, K., Salminen, P., 2005. Reference point approach for multiple decision makers. European Journal of Operational Research 164 (3), 785–791.

Lahdelma, R., Salminen, P., 2001. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Operations Research 49 (3), 444–454.

Lahdelma, R., Salminen, P., 2002. Pseudo-criteria versus linear utility function in stochastic multi-criteria acceptability analysis. European Journal of Operational Research 141 (2), 454–469.

Lahdelma, R., Salminen, P., 2003. Loss aversion in SMAA. Tech. Rep. 269, University of Jyväskylä, School of Business and Economics.

Lahdelma, R., Salminen, P., 2006a. Classifying efficient alternatives in SMAA using cross confidence factors. European Journal of Operational Research 170 (1), 228–240.

Lahdelma, R., Salminen, P., 2006b. Stochastic multicriteria acceptability analysis using the data envelopment model. European Journal of Operational Research 170 (1), 241–252.

Lahdelma, R., Salminen, P., Hokkanen, J., 2002. Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. European Journal of Operational Research 142 (2), 345–356.

Leinonen, R., 1999. Building traffic simulator. Master's thesis, Helsinki University of Technology.

Leskinen, P., Kangas, A. S., Kangas, J., 2004. Rank-based modelling of preferences in multi-criteria decision making. European Journal of Operational Research 158, 721–733.

Leskinen, P., Viitanen, J., Kangas, A., Kangas, J., 2006. Alternatives to incorporate uncertainty and risk attitude in multicriteria evaluation of forest plans. Forest Science 52 (3), 304–312.

Linkov, I., Tkachuk, A., Levchenko, A., Seager, T. P., Tuler, S. P., Kay, R., Figueira, J. R., Tervonen, T., 2007. An MCDA approach for establishing GPRA metrics: oil spill response. Working Paper 6/2007 of the Centre for Management Studies, Instituto Superior Técnico, Technical University of Lisbon, Portugal.

Milton, J. S., Arnold, J. C., 1995. Introduction to probability and statistics, 3rd Edition. Probability and statistics. McGraw-Hill International editions.

Morley, C. G., July 1962. Installations, mechanical: lifts, escalators, paternosters. Architect's Journal 136, 233–258.

Mousseau, V., Dias, L., Figueira, J., 2004. Dealing with inconsistent judgements in multiple criteria sorting models. Research Report 17/2004 of The Institute of Systems Engineering and Computers (INESC-Coimbra), Coimbra, Portugal, http://www.inescc.pt.

Mousseau, V., Figueira, J., Dias, L., Gomes da Silva, C., Clímaco, J., 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. European Journal of Operational Research 147, 72–93.

Mousseau, V., Figueira, J., Naux, J., 2001. Using assignment examples to infer weights for ELECTRE TRI method: Some experimental results. European Journal of Operational Research 130 (2), 263–275.

Mousseau, V., Słowiński, R., 1998. Inferring an ELECTRE TRI model from assignment examples. Journal of Global Optimization 12 (2), 157–174.

Mousseau, V., Słowiński, R., Zielniewicz, P., 2000. A user-oriented implementation of the ELECTRE-TRI method integrating preference elicitation support. Computers & Operations Research 27 (7-8), 757–777.

Mustajoki, J., Hämäläinen, R. P., Lindstedt, M. R. K., 2006. Using interval for global sensitivity and worst case analyses in multiattribute value trees. European Journal of Operational Research 174, 278–292.

Mustajoki, J., Hämäläinen, R. P., Salo, A., 2005. Decision support by interval SMART/SWING - incorporating imprecision in the SMART and SWING methods. Decision Sciences 36 (2), 317–339.

Ngo The, A., Mousseau, V., 2002. Using assignment examples to infer category limits for the ELECTRE TRI method. Journal of Multi-Criteria Decision Analysis 11, 29–43.

Parlow, H., March 1966. Lift operation and computers, a simulation of performance. Architect's Journal, 747–753.

Pawlak, Z., Słowiński, R., 1994. Decision analysis using rough sets. International Transactions in Operational Research 1 (1), 107–114.

Phillips, R. S., 1966. Electric lifts, 5th Edition. Pitman, London.

Pinfold, W. J., 1966. Lift design and user requirements. In: Symposium on lifts in large buildings. Building Research Station, Watford.

Pirlot, M., 1995. The characterization of 'min' as a procedure for exploiting valued preference relations and related results. Journal of Multi-Criteria Decision Analysis 4 (1), 37–57.

Rietveld, P., 1980. Multiple Objective Decision Methods and Regional Planning. North-Holland, Amsterdam.

Rietveld, P., Ouwersloot, H., 1992. Ordinal data in multicriteria decision making, a stochastic dominance approach to siting nuclear power plants. European Journal of Operational Research 56, 249–262.

Roubens, M., 1997. Fuzzy sets and decision analysis. Fuzzy Sets and Systems 90, 199–206.

Roy, B., 1968. Classement et choix en présence de points de vue multiples: La méthode ELECTRE. Revue Francaise d'Informatique et de Recherche Opérationnelle 8, 57–75.

Roy, B., 1971. La méthode ELECTRE II. Tech. rep., METRA, Direction Scientifique, Note de Travail n. 142.

Roy, B., 1978. ELECTRE III: Un algorithme de classement fondé sur une représentation floue des préferences en présense de critères multiples. Cahiers du Centre d'Etudes de Recherche Operationnelle 20 (1), 3–24.

Roy, B., 1996. Multicriteria Methodology for Decision Analysis. Kluwer Academic Publishers, Dordrecht.

Roy, B., Bertier, P., 1973. La méthode ELECTRE II - une application au média-planning. In: Ross, M. (Ed.), OR'72. North-Holland Publishing Company, pp. 291–302.

Roy, B., Hugonnard, J., 1982. Classement des prolongements de lignes de métro en banlieue parisienne (présentation d'une méthode multicritère originale). Cahiers du CERO 24 (2-3-4), 153–171.

Roy, B., Présent, M., Silhol, D., 1986. A programming method for determining which Paris metro stations should be renovated. European Journal of Operational Research 24, 318–334.

Roy, B., Skalka, J., 1984. ELECTRE IS : Aspects méthodologiques et guide d'utilisation. Tech. rep., Document du LAMSADE N^o 30, Université Paris-Dauphine, Paris.

Siikonen, M.-L., Leppälä, J., 1991. Elevator traffic pattern recognition. In: Proceedings of the Fourth World Congress of the International Fuzzy Systems Association, Brussels, Belgium. pp. 195–198.

Strakosch, G. R., 1967. Vertical Transportation. Wiley, New York.

Tervonen, T., Figueira, J., Lahdelma, R., Salminen, P., 2004a. Modelling MCDA group preferences for public human resource management: Evaluating the quality of education at the department of information technology, the University of Turku (Finland). Research Report 22/2004 of The Institute of Systems Engineering and Computers (INESC-Coimbra), Coimbra, Portugal, http://www.inescc.pt.

Tervonen, T., Lahdelma, R., Dias, J. A., Figueira, J., Salminen, P., 2007. SMAA-TRI: A parameter stability analysis method for ELECTRE-TRI. In: Linkov, I., Kiker, G. A., Wenning, R. J. (Eds.), Environmental Security in Harbors and Coastal Areas. Springer, pp. 217–231.

Tervonen, T., Lahdelma, R., Salminen, P., 2004b. A method for elicitating and combining group preferences for Stochastic Multicriteria Acceptability Analysis. Tech. Rep. 638, TUCS - Turku Center for Computer Science, http://www.tucs.fi.

Tregenza, P. R., 1971. The prediction of lift performance in multi-storey buildings. Ph.D. thesis, University of Nottingham.

Vincke, P., 1992. Multicriteria Decision-Aid. John Wiley & Sons, New York.

Xu, X., 2001. The SIR method: A superiority and inferiority ranking method for multiple criteria decision making. European Journal of Operational Research 131 (3), 587–602.

Yu, W., 1992a. Aide multicritère à la décision dans le cadre de la problématique du tri : Concepts, méthodes et applications. Ph.D. thesis, Université Paris-Dauphine.

Yu, W., 1992b. Aide multicritère à la décision dans le cadre de la problématique du tri: méthodes et applications. Ph.D. thesis, LAMSADE, Université Paris Dauphine, Paris.

Publication reprints

Publication I

Tervonen T., Lahdelma R., 2007. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research 178(2), 500–513.

Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 178 (2007) 500–513
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Decision Support

Implementing stochastic multicriteria acceptability analysis

Tommi Tervonen [a,b,c,*], Risto Lahdelma [a]

[a] University of Turku, Department of Information Technology, FIN-20520 Turku, Finland
[b] INESC—Coimbra, R. Antero de Quental 199, 3000-033 Coimbra, Portugal
[c] Faculdade de Economia da Universidade de Coimbra, Av. Dias da Silva 165, 3004-512 Coimbra, Portugal

Received 27 July 2004; accepted 29 December 2005
Available online 17 April 2006

Abstract

Stochastic multicriteria acceptability analysis (SMAA) is a family of methods for aiding multicriteria group decision making in problems with inaccurate, uncertain, or missing information. These methods are based on exploring the weight space in order to describe the preferences that make each alternative the most preferred one, or that would give a certain rank for a specific alternative. The main results of the analysis are rank acceptability indices, central weight vectors and confidence factors for different alternatives. The rank acceptability indices describe the variety of different preferences resulting in a certain rank for an alternative, the central weight vectors represent the typical preferences favouring each alternative, and the confidence factors measure whether the criteria measurements are sufficiently accurate for making an informed decision.

The computations in SMAA require the evaluation of multidimensional integrals that must in practice be computed numerically. In this paper we present efficient methods for performing the computations through Monte Carlo simulation, analyze the complexity, and assess the accuracy of the presented algorithms. We also test the efficiency of these methods empirically. Based on the tests, the implementation is fast enough to analyze typical-sized discrete problems interactively within seconds. Due to almost linear time complexity, the method is also suitable for analysing very large decision problems, for example, discrete approximations of continuous decision problems.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Stochastic multicriteria acceptability analysis; Simulation; Multiple criteria analysis; Complexity analysis

1. Introduction

Stochastic multicriteria acceptability analysis (SMAA) methods have been developed for discrete multicriteria decision aiding (MCDA) problems, where criteria measurements are uncertain or inaccurate and where it is for some reason difficult to obtain accurate or any preference information from the decision makers (DMs) (Lahdelma and Salminen, 2001).

Usually in MCDA problems the preference information is modelled by determining importance weights for criteria. The SMAA methods are based on exploring the weight space in order to describe

* Corresponding author. Address: INESC—Coimbra, R. Antero de Quental 199, 3000-033 Coimbra, Portugal. Tel.: +351 96 529 1325; fax: +351 23 982 4692.
E-mail addresses: tommi.tervonen@it.utu.fi (T. Tervonen), risto.lahdelma@cs.utu.fi (R. Lahdelma).

0377-2217/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2005.12.037

the preferences that would make each alternative the most preferred one, or that would give a certain rank for a specific alternative. The main results of the analysis are rank acceptability indices, central weight vectors and confidence factors for different alternatives. The rank acceptability indices describe the variety of different preferences resulting in a certain rank for an alternative, the central weight vectors represent the typical preferences favouring each alternative, and the confidence factors measure whether the criteria measurements are sufficiently accurate for making an informed decision.

In MCDA literature outside SMAA, there is a long history of methodologies that allow decision aiding under uncertain and/or imprecise information. See e.g. Dias and Clímaco (2000), Dias et al. (2002), Fishburn (1965), Hazen (1986), Kirkwood and Sarin (1985), Mousseau et al. (2000, 2003), and for more general information on this subject, see Figueira et al. (2005). Although this area has been studied for three decades, the SMAA methods are the first ones allowing both preference information and criteria measurements to be expressed as arbitrarily distributed stochastic variables. The SMAA approach has also recently been applied to extend other MCDA methods to allow using them with imprecise information (see Tervonen et al., 2005).

The SMAA methods are based on inverse weight space analysis, which has also been considered in the works of Charnetski Soland (1978) and Bana e Costa (1986). In the original SMAA method by Lahdelma et al. (1998) the weight space analysis is performed based on an additive utility or value function and stochastic criteria measurements. The SMAA-2 method (Lahdelma and Salminen, 2001) generalized the analysis to a general utility or value function, to include various kinds of preference information and to consider holistically all ranks. The SMAA-3 method (Lahdelma and Salminen, 2002) applies ELECTRE III type pseudo-criteria in the analysis. The SMAA-O method (Lahdelma et al., 2003) extends SMAA-2 for treating mixed ordinal and cardinal criteria in a comparable manner. The SMAA-A method (or Ref-SMAA method) models the preferences using reference points and achievement scalarizing functions (Lahdelma et al., 2005). Durbach (2006) has also developed a variant of the SMAA-A method using achievement functions.

SMAA methods are applicable in many real-life problem types for a number of reasons. Firstly, the inverse weight space approach is suitable for many group decision-making problems, where the DMs are unable or unwilling to provide preference information, or it is difficult to reach consensus over the preferences. In such cases the preference information can be expressed as weight intervals including preferences of all DMs, or with some other weight distribution accepted by all DMs. SMAA can then be used to compute descriptive information about the acceptability of different alternatives, and this can help the DMs to identify commonly acceptable compromise solutions. Secondly, SMAA supports a very general and flexible way to model different kinds of uncertain or inaccurate preference and criteria information through stochastic distributions. Thirdly, as demonstrated in this paper, the SMAA computations can be implemented very efficiently through numerical methods, making it possible to use the method in many different decision-making contexts, including interactive decision processes. As a consequence, SMAA methods have been successfully applied in a number of real-life decision problems in Finland. For applications of SMAA, see e.g. Hokkanen et al. (1998, 1999, 2000), Kangas et al. (2003, in press), Kangas and Kangas (2003), Lahdelma and Salminen (2006), Lahdelma et al. (2001, 2002).

In this paper we describe how the basic computations of the SMAA-2 and SMAA-O methods can be implemented efficiently through Monte Carlo simulation. We have chosen to present the computations of these two methods, because they form the basis for all other SMAA variants. In particular, we present the algorithms for computing the rank acceptability indices, central weight vectors, and confidence factors. We begin by introducing the SMAA-2 and SMAA-O methods in Section 2. In Section 3, we describe the implementation of the algorithms and discuss techniques for handling preference information. Following this, we analyze the complexity of the algorithms theoretically in Section 4. We assess the accuracy of the computations in Section 5, and present results from empirical efficiency tests in Section 6. We end this paper with conclusions in Section 7.

2. The SMAA-2 and SMAA-O methods

2.1. The basic SMAA-2 method

The SMAA-2 method (Lahdelma and Salminen, 2001) has been developed for discrete stochastic

multicriteria decision-making problems with multiple DMs. SMAA-2 applies inverse weight space analysis to describe for each alternative what kind of preferences make it the most preferred one, or place it on any particular rank. The decision problem is represented as a set of m alternatives $\{x_1, x_2, \ldots, x_m\}$ that are evaluated in terms of n criteria. The DMs' preference structure is represented by a real-valued utility or value function $u(x_i, w)$. The value function maps the different alternatives to real values by using a weight vector w to quantify DMs' subjective preferences. SMAA-2 has been developed for situations where neither criteria measurements nor weights are precisely known. Uncertain or imprecise criteria are represented by stochastic variables ξ_{ij} with joint density function $f_X(\xi)$ in the space $X \subseteq R^{m\times n}$. The DMs' unknown or partially known preferences are represented by a weight distribution with joint density function $f_W(w)$ in the feasible weight space W. Total lack of preference information is represented in 'Bayesian' spirit by a uniform weight distribution in W, that is, $f_W(w) = 1/\mathrm{vol}(W)$. The weight space can be defined according to needs, but typically, the weights are non-negative and normalized, that is; the weight space is an $n-1$-dimensional simplex in n-dimensional space:

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

Fig. 1 presents the feasible weight space of a three-criterion problem as the shaded triangle with corner points $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$.

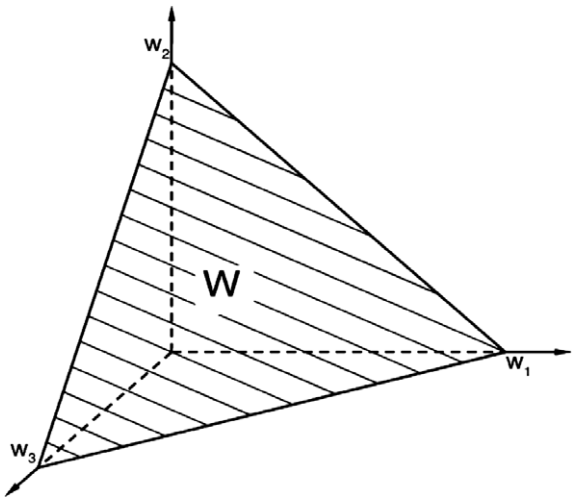


Fig. 1. Feasible weight space of a three-criterion problem.

The value function is used to map the stochastic criteria and weight distributions into value distributions $u(\xi_i, w)$. Based on the value distributions, the rank of each alternative is defined as an integer from the best rank (=1) to the worst rank (=m) by means of a ranking function

$$\mathrm{rank}(i, \xi, w) = 1 + \sum_{k=1}^{m} \rho(u(\xi_k, w) > u(\xi_i, w)), \qquad (2)$$

where $\rho(true) = 1$ and $\rho(false) = 0$. SMAA-2 is then based on analysing the stochastic sets of favourable rank weights

$$W_i^r(\xi) = \{w \in W : \mathrm{rank}(i, \xi, w) = r\}. \qquad (3)$$

Any weight $w \in W_i^r(\xi)$ results in such values for different alternatives, that alternative x_i obtains rank r.

The first descriptive measure of SMAA-2 is the *rank acceptability index* b_i^r, which measures the variety of different preferences (weights) that grant alternative x_i rank r. It is the share of all feasible weights that make the alternative acceptable for a particular rank, and it is most conveniently expressed percentage-wise. The rank acceptability index b_i^r is computed numerically as a multidimensional integral over the criteria distributions and the favourable rank weights as

$$b_i^r = \int_{\xi \in X} f_X(\xi) \int_{w \in W_i^r(\xi)} f_W(w)\, \mathrm{d}w\, \mathrm{d}\xi. \qquad (4)$$

The most acceptable (best) alternatives are those with high acceptabilities for the best ranks. Evidently, the rank acceptability indices are in the range $[0,1]$, where 0 indicates that the alternative will never obtain a given rank and 1 indicates that it will obtain the given rank always with any choice of weights.

Favourable rank weights and rank acceptability indices are illustrated in Fig. 2. The figure represents a deterministic two-criterion, three-alternative problem with linear value function. The favourable first rank weights (W_i^1) are shown in light gray, bordered by the favourable second rank weights (W_i^2) in dark gray. First and second rank acceptability indices (b_i^1, b_i^2) correspond in this figure to the distances spanned by the favourable rank weights. When the problem contains multiple criteria and alternatives, the rank acceptability indices can be better visualized by a three-dimensional column chart. Fig. 3 shows the rank acceptability indices from the Helsinki Harbour case with 13 alternatives and 11 criteria (Hokkanen et al., 1999).

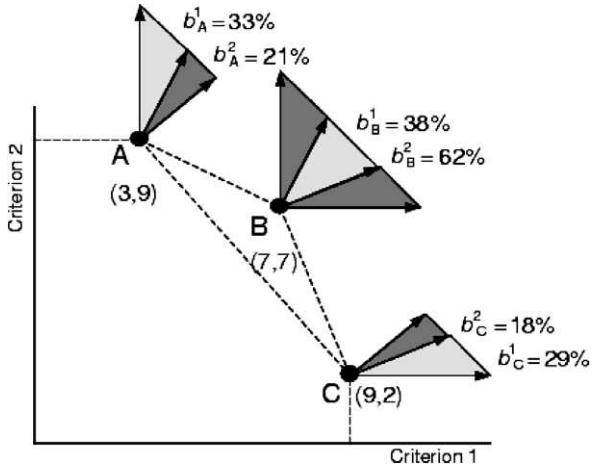


Fig. 2. First and second rank acceptabilities in a deterministic two-criterion problem with linear value function (Lahdelma and Salminen, 2001).

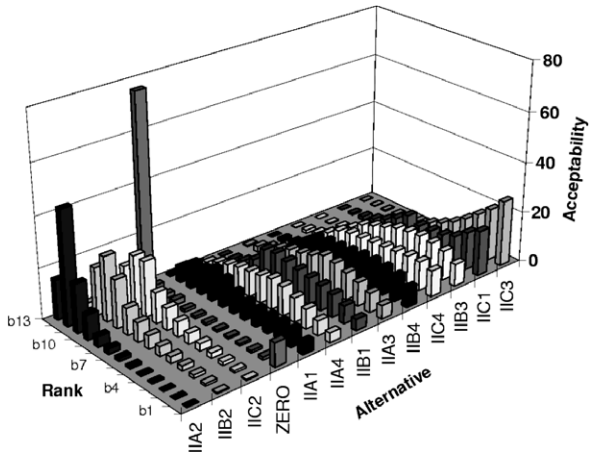


Fig. 3. Rank acceptability indices (b_i) from the Helsinki Harbour decision-making problem.

The first rank acceptability index b_i^1 is called the *acceptability index* a_i. The acceptability index is particularly interesting, because it is non-zero for stochastically efficient alternatives (alternatives that are efficient with some values for the stochastic criteria measurements) and zero for inefficient alternatives. The acceptability index not only identifies the efficient alternatives, but also measures the strength of the efficiency considering the uncertainty in criteria and DMs' preferences.

The *central weight vector* w_i^c is the expected centre of gravity (centroid) of the favourable first rank weights of an alternative. The central weight vector represents the preferences of a 'typical' DM supporting this alternative. The central weights of different alternatives can be presented to the DMs in order to help them understand how different weights correspond to different choices with the assumed preference model. The central weight vector w_i^c is computed numerically as a multidimensional integral over the criteria distributions and the favourable first rank weights using

$$w_i^c = \int_{\xi \in X} f_X(\xi) \int_{w \in W_i^1(\xi)} f_W(w) w \, \mathrm{d}w \, \mathrm{d}\xi / a_i. \tag{5}$$

Fig. 4 presents a sample chart of central weight vectors from the Helsinki Harbour decision-making problem.

The *confidence factor* p_i^c is the probability for an alternative to obtain the first rank when the central weight vector is chosen. The confidence factor is computed as a multidimensional integral over the criteria distributions using

$$p_i^c = \int_{\xi \in X : \mathrm{rank}(i,\xi,w_i^c)=1} f_X(\xi) \mathrm{d}\xi. \tag{6}$$

Confidence factors can similarly be calculated for any given weight vectors. The confidence factors measure whether the criteria measurements are accurate enough to discern the efficient alternatives.

2.2. Ordinal criteria

The SMAA-O method (Lahdelma et al., 2003) extends SMAA-2 to handle ordinal criteria measurements. In SMAA-O, the criteria may be ordinal, cardinal, or mixed. In the mixed case some of the criteria are measured on cardinal (interval) scales and others on ordinal scales. For an ordinal criterion, each alternative is measured by assigning it a *rank level*. The rank level x_{ij} is an integer from the best rank level 1 to the worst rank level m_j. Observe that multiple alternatives may obtain the same rank level, in which case $m_j < m$. The idea in SMAA-O is to map the ordinal criteria measurements into cardinal scales before they are used in the computations. The mapping is implemented by a function $g_j(\cdot)$ that preserves the ordinal information:

$$x_{ij} \succ x_{kj} \iff g_j(x_{ij}) > g_j(x_{kj}) \quad \forall i,k \in \{1,\ldots,m\}. \tag{7}$$

Without loss of generality we can assume that the mapping is scaled to interval $[0,1]$. Otherwise the

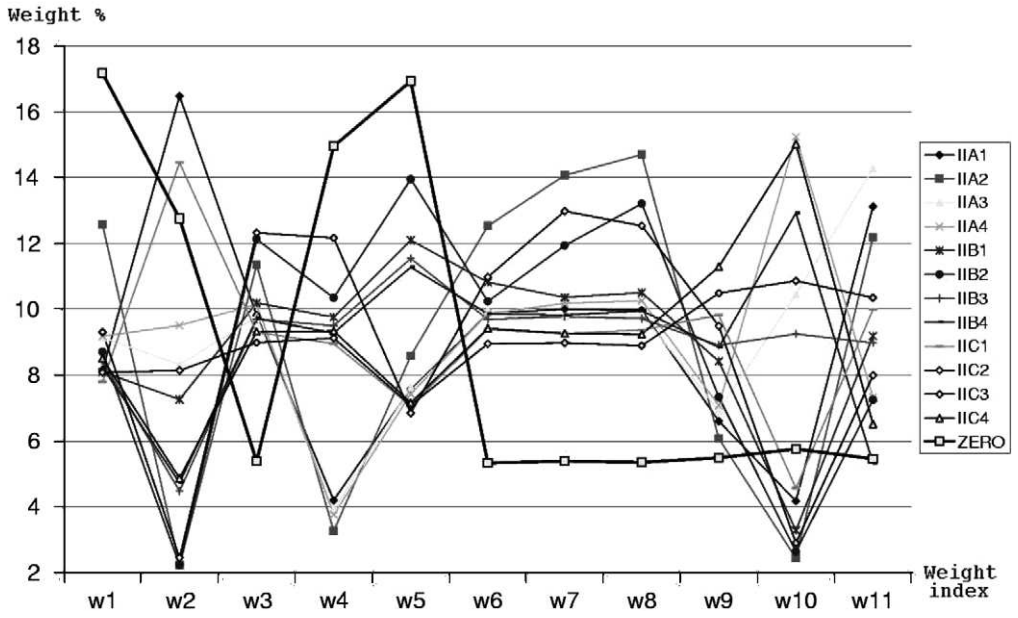


Fig. 4. Central weight vectors from the Helsinki Harbour decision-making problem.

shape of the mapping is unknown. SMAA-O simulates numerically all such mappings that preserve the ordinal criteria information.

2.3. Preference information

There are several different ways to handle partial preference information in SMAA methods. In this paper we focus on two ways that are applicable when the value function is additive (Lahdelma and Salminen, 2001):

- interval constraints for weights, and
- complete ranking of the criteria.

The preference information might also be mixed: there might be exact numerical values for some weights, ranking for a set of weights, and interval constraints for some weights. Mixed preference information is not considered in this paper.

Interval constraints for weights are given in the form

$$0 \leqslant w_j^{\min} \leqslant w_j \leqslant w_j^{\max} \leqslant 1, \quad \text{where } j \in \{1, \ldots, n\}. \tag{8}$$

The intervals $[w^{\min}, w^{\max}]$ can be defined so that they contain the preferences of the DMs (and other interest groups). The DMs can express their preferences either as precise weights or as weight intervals. The weight space analysis of SMAA is then performed in the restricted weight space

$$W' = \{w \in W | w_j^{\min} \leqslant w_j \leqslant w_j^{\max},\ j = 1, \ldots, n\}. \tag{9}$$

This means that the uniform weight distribution $f_W(w)$ is redefined as

$$f_W(w) = \begin{cases} 1/\mathrm{vol}(W') & \text{if } w \in W', \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Fig. 5 illustrates the restricted feasible weight space of a three-criterion problem with lower and upper bounds for w_1.

Complete ranking of the criteria is expressed as a sequence of inequality constraints for the weights

$$w_{j_1} \geqslant w_{j_2} \geqslant \cdots \geqslant w_{j_n}. \tag{11}$$

Such a ranking can be obtained by asking the DMs to identify the most important, second most important, etc. criterion. When judging mutual importance of the criteria, the DMs should consider the difference between the best and worst value for each criterion. If the DMs consider two criteria equally important, this can be represented by an equality constraint between those criteria in (11). Fig. 6 illustrates the feasible weight space for a three-criterion problem with the ranking $w_1 \geqslant w_2 \geqslant w_3$.

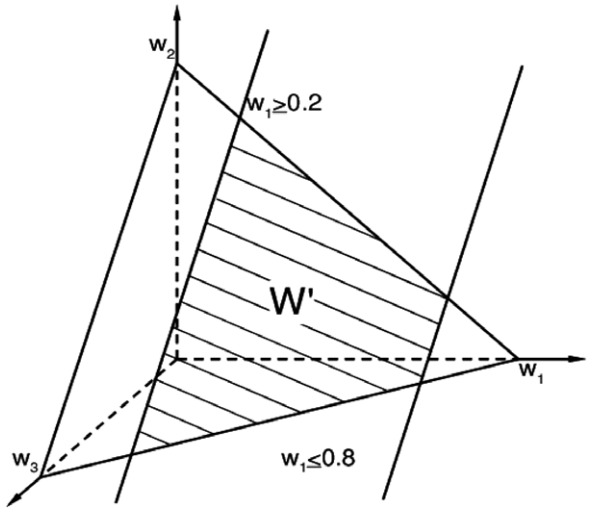


Fig. 5. Feasible weight space of a three-criterion problem with lower and upper bounds for w_1.

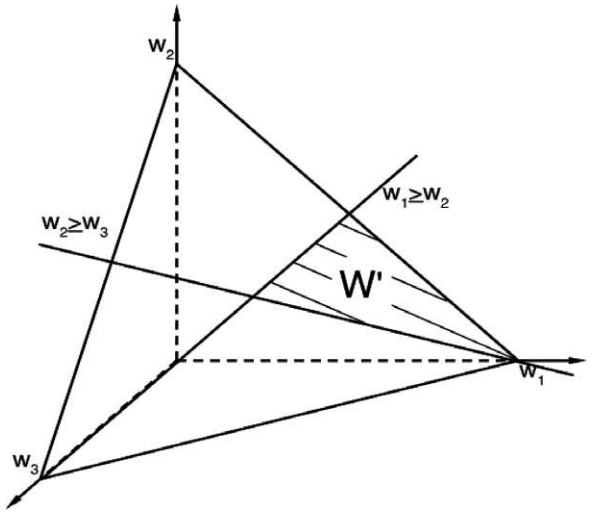


Fig. 6. Feasible weight space of a three-criterion problem with ranking of the criteria.

3. Description of the SMAA algorithm

The multidimensional integrals (4)–(6) of SMAA computations are in practice impossible to compute analytically, because the distributions f_X and f_W vary according to the application and can be arbitrarily complex. Straightforward integration techniques based on discretizing the distributions with respect to each dimension are infeasible, because the integrals have a very high dimension, and the required effort depends exponentially on the number of dimensions. For example, in a problem with eight criteria and 10 alternatives, the dimension of the integral for computing rank acceptability indices is 88, because in (4) the outer integration is through the eight-dimensional criteria space, and the inner one through the space of all criteria measurements for all alternatives ($8 \times 10 = 80$ dimensions). However, due to the nature of the problem, we do not need an answer with very high precision. Monte Carlo simulation is a well-established method for computing approximative values for high-dimensional integrals. In Monte Carlo simulation the required number of iterations is inversely proportional to the square of the desired accuracy, but does not significantly depend on the dimensionality of the problem (Fishman, 1996). Thus, Monte Carlo simulation can be used to obtain a precision of a few decimal places with moderate effort.

The algorithm is described in four parts. We first describe the method for generating a criterion matrix with cardinal and ordinal criteria. Secondly, we describe the applied weight generation technique and how preference information is handled. Before we describe the actual SMAA algorithm, we observe that the confidence factors (6) depend both on the central weight vectors and on the acceptability indices. As a consequence, the algorithm must consist of two phases. *Phase 1* consists of computation of the rank acceptability indices and the central weight vectors. The confidence factors are computed in *Phase 2*.

The following symbols are used in Algorithms 1–4:

- h_i^j number of times alternative i is evaluated into rank j in Monte Carlo simulations of Phase 1 (hits for rank j of alternative i)
- K_w number of iterations in Phase 1 (rank acceptability index and central weight vector computation)
- K_c number of iterations in Phase 2 (confidence factor computation)
- m_j number of rank levels for ordinal criterion j
- $r = [r_1, \ldots, r_m]$ vector of ranks of the alternatives
- $t = [t_1, \ldots, t_m]$ vector of value function values of the alternatives

The algorithms also use the following functions and subroutines:

- $\mathrm{RAND}_{U[0,1]}()$ function returning a uniformly distributed random number from the interval $[0, 1]$
- $\mathrm{RAND}_X()$ function returning a random criterion matrix from criteria distribution f_X

$\mathrm{RAND}_W()$ function returning a random weight vector from weight distribution f_W

$\mathrm{RANK}(t)$ function returning a vector of ranks corresponding to the vector of value function values t

$\mathrm{SORT}_{\mathrm{asc}}(s)$ subroutine sorting the components of vector s into ascending order

$\mathrm{SORT}_{\mathrm{desc}}(s)$ subroutine sorting the components of vector s into descending order

3.1. Generation of the criteria measurement matrix

The $\mathrm{RAND}_X()$ function generates a random criterion matrix of size $m \times n$ from the given criteria distribution. Each row of the matrix contains criteria measurements of a certain alternative. Cardinal criteria measurements follow a joint distribution, or independent distributions. We do not consider joint distribution for cardinal criteria in this paper. Refer to Lahdelma et al. (2004). Independent criteria measurements are generated separately from their corresponding distributions. Their distributions may have an arbitrary shape (e.g.uniform, normal, ...).

If some of the criteria are measured on ordinal scales, then the ordinal to cardinal mapping must be simulated for those criteria each time a new criterion matrix is created. The ordinal to cardinal mapping is simulated using the following method: first, $m_j - 2$ uniformly distributed random numbers from the interval $]0,1[$ are generated and sorted into descending order (m_j is the number of rank levels). Then, 1 is inserted as the first number and 0 as the last number. The simulated cardinal value for rank level j is then the jth of these numbers. Thus the simulated cardinal value for the best rank level is 1, and the simulated cardinal value for the worst rank level is 0. The simulated cardinal values for other rank levels should be unique and in the interval $]0,1[$. Because the majority of pseudo-random number generators will not produce duplicate floating point values except after a very long sequence, it is in practice unnecessary to have any special treatment for duplicate values. The procedure for generating the simulated cardinal values is defined as pseudo-code in Algorithm 1. Complexity of this procedure is due to sorting $O(m\log(m))$. The procedure must be executed once for each ordinal criterion when a new criterion matrix is generated. Fig. 7 illustrates a possible mapping with 11 rank levels generated by this procedure.

Algorithm 1. Generation of m_j simulated cardinal values $(q_1, \ldots, q_{m_j})$.

Output: q
1: **for** $j \leftarrow 2$ **to** $m_j - 1$ **do**
2: $\quad q_j \leftarrow \mathrm{RAND}_{U[0,1]}()$
3: **end for**
4: $\mathrm{SORT}_{\mathrm{desc}}(q)$
5: $q_1 \leftarrow 1$
6: $q_{m_j} \leftarrow 0$

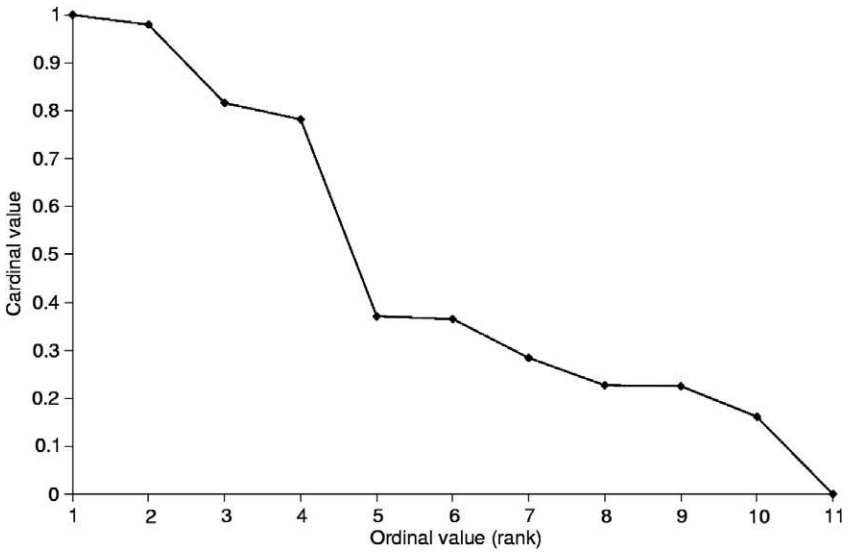


Fig. 7. A sample ordinal to cardinal mapping with $m_j = 11$ rank levels.

3.2. Generation of weights and handling preference information

The $\mathrm{RAND}_W()$ function generates the weights from the given weight distribution. We describe weight generation corresponding to three different types of preference information:

- absent preference information,
- interval constraints for weights, and
- complete ranking of the criteria.

In case of absent preference information the weights are generated from a uniform distribution in the normalized weight space (1). Because the weights must sum to unity, the n weights w_j are generated according to the following method: first $n-1$ independent random numbers are generated from the uniform distribution in interval $[0,1]$, and sorted into ascending order $(q_1, q_2, \ldots, q_{n-1})$. After that, 1 is inserted as the last number ($q_n = 1$) and 0 as the first number ($q_0 = 0$). Uniformly distributed normalized weights are then obtained as intervals between the consecutive numbers ($w_j = q_j - q_{j-1}$) (David, 1970). The procedure for generating the n uniformly distributed normalized weights is defined in Algorithm 2. Complexity of this procedure is $O(n\log(n))$ due to sorting.

When preference information is available, the weight generation process must be modified a little. Upper and lower bounds for weights (and in principle also more complex weight constraints) can be implemented by the rejection technique. After a vector of uniformly distributed normalized weights has been generated, the weights are tested against their bounds. If any of the constraints is not satisfied, the entire set is rejected and the weight generation is repeated. A problem with the rejection technique is that it may cause a very large share of the weights to be rejected and a very small share of them to be accepted. A small acceptance rate not only slows down the computation, but may also cause problems with the quality of the generated pseudo-random numbers that pass the rejection test. However, upper and lower bounds affect the acceptance rate differently. As can be seen from Fig. 5, upper bounds cut off the tip of the simplex, but lower bounds cut off the base. In a high-dimensional weight space, the volume of the tip is very small, but the volume of the base is large in relation to the entire weight space.

To estimate how large a share of the weight vectors need to be rejected due to upper bounds, we assume that all weights have a common upper bound $w^{\max}$. Then the probability for the largest of the generated weights to exceed the upper bound is

$$\begin{aligned} P[\max\{w_j\} > w^{\max}] &= n(1-w^{\max})^{n-1} - \binom{n}{2}(1-2w^{\max})^{n-1} \\ &\quad + \cdots (-1)^{k-1}\binom{n}{k}(1-kw^{\max})^{n-1} \cdots, \end{aligned} \tag{12}$$

where the series continues as long as $1 - kw^{\max} > 0$ (David, 1970). For example, if there are $n = 5$ weights with upper bound $w^{\max} = 0.4$, the rejection percentage is 63.2%. If we are applying both lower and upper bounds, it might be the case, that the lower bounds render some upper bounds redundant. Consider a three-criterion case with a lower bound of 0.3 for all weights. Then the maximum value that any weight may obtain is $1 - 0.3 - 0.3 = 0.4$. Therefore all upper bound weight constraints of 0.4 or higher are redundant.

The rejection technique can be very inefficient for weights with lower bounds in high-dimensional problems. Lower bounds can be treated efficiently by using a simple transformation technique. With lower bounds the feasible weight space becomes

$$W' = \left\{ w \in R^n | w_j \geqslant w_j^{\min} \quad \text{and} \quad \sum_{j=1}^{n} w_j = 1 \right\}, \tag{13}$$

which has the same simplex shape as the original weight space W, but is smaller. By substituting $w'_j = w_j - w_j^{\min}$ the restricted weight space becomes

$$W' = \left\{ w' \in R^n | w'_j \geqslant 0 \quad \text{and} \quad \sum_{j=1}^{n} w'_j = 1 - C \right\}, \quad \text{where } C = \sum_{j=1}^{n} w_j^{\min}. \tag{14}$$

The shifted weights w'_j can now be generated by a modification of Algorithm 2 where the weights are generated to sum to $1 - C$ instead of 1. Then the lower bounded weights w_j are obtained by substituting back $w_j = w'_j + w_j^{\min}$. Lower bounds do therefore not increase the complexity of weight generation.

Preference information presented in form of a complete ranking of the criteria is handled using a similar technique that was used for simulating the

ordinal to cardinal mapping in the previous section. First we generate a set of weights from a uniform distribution in the normalized weight space as in the case of absent preference information. Then we sort the weights into a consistent order according to the ranking of the criteria. This does not increase the complexity of weight generation.

Algorithm 2. Generation of n uniformly distributed random weights from the interval $[0,1]$ $(w_1,\ldots,w_n)$ which sum to unity.

Output: w
1: **for** $j \leftarrow 1$ **to** $n-1$ **do**
2: $\quad q_j \leftarrow \mathrm{RAND}_{U[0,1]}()$
3: **end for**
4: $\mathrm{SORT}_{\mathrm{asc}}(q)$
5: $q_0 \leftarrow 0$
6: $q_n \leftarrow 1$
7: **for** $j \leftarrow 1$ **to** n **do**
8: $\quad w_j \leftarrow q_j - q_{j-1}$
9: **end for**

3.3. Phase 1. Computation of b_i^r and w_i^c

To compute the rank acceptability indices b_i^r and the central weight vectors w_i^c for each alternative i, we must integrate over the criteria and weight distributions. Straightforward computation of the rank acceptability indices (4) would require executing Monte Carlo simulation $m \cdot n$ times, once for each index. Similarly, computing the central weight vector (5) for each alternative would require m executions. We can speed up the computation remarkably by observing that all rank acceptability indices and central weight vectors can be computed in a single simulation run. To do this, we generate during each iteration a random criterion matrix and a random weight vector from their corresponding distributions. Then we compute statistics on the ranks that different alternatives obtained and update the central weight vector of the most preferred alternative. Phase 1 is described as pseudocode in Algorithm 3.

Algorithm 3 uses the function RANK(t). RANK(t) returns a vector of ranks for alternatives based on their values in vector t. For example, if $t = [0, 0.5, 0.2]$, the resulting rank vector is $[3,1,2]$. This function is implemented efficiently by sorting the alternatives into descending order by their values and then assigning consecutive ranks from 1 to m to the sorted alternatives. Sometimes two or more alternatives may have the same values, and they should thus be assigned the same rank. Rank assignment should be implemented to handle such cases properly. However, shared ranks will be extremely rare when the criteria measurements are stochastic and independent. Complexity of this procedure is due to sorting $O(m\log(m))$.

Observe in Algorithm 3 that the central weight vector (w_i^c) is defined only when the acceptability index is non-zero, or, equivalently, when hits for the first rank (h_i^1) is greater than zero.

Algorithm 3. Monte Carlo simulation to compute the central weight vectors (w_i^c's) and the acceptability indices (b_i^r's).

Output: w_i^c's, b_i^r's
1: // Initialization of w_i^c and hit count
2: **for** $i \leftarrow 1$ **to** m **do**
3: $\quad w_i^c \leftarrow 0$
4: $\quad$ **for** $j \leftarrow 1$ **to** m **do**
5: $\quad\quad h_i^j \leftarrow 0$
6: $\quad$ **end for**
7: **end for**
8: // Main loop
9: **for** $k \leftarrow 1$ **to** K_w **do**
10: $\quad w \leftarrow \mathrm{RAND}_W()$
11: $\quad x \leftarrow \mathrm{RAND}_X()$
12: $\quad$ **for** $i \leftarrow 1$ **to** m **do**
13: $\quad\quad t_i \leftarrow u(x_i, w)$
14: $\quad$ **end for**
15: $\quad r \leftarrow \mathrm{RANK}(t)$
16: $\quad$ **for** $i \leftarrow 1$ **to** m **do**
17: $\quad\quad h_i^{r_i} \leftarrow h_i^{r_i} + 1$
18: $\quad\quad$ **if** $r_i = 1$ **then**
19: $\quad\quad\quad w_i^c \leftarrow w_i^c + w$
20: $\quad\quad$ **end if**
21: $\quad$ **end for**
22: **end for**
23: // Computation of w_i^c and b_i^r
24: **for** $i \leftarrow 1$ **to** m **do**
25: $\quad$ **if** $h_i^1 > 0$ **then**
26: $\quad\quad w_i^c \leftarrow w_i^c / h_i^1$
27: $\quad$ **end if**
28: $\quad$ **for** $j \leftarrow 1$ **to** m **do**
29: $\quad\quad b_i^j \leftarrow h_i^j / K_w$
30: $\quad$ **end for**
31: **end for**

3.4. Phase 2. Computation of p_i^c

To compute the confidence factors p_i^c from (6) we must integrate over the criteria distribution with respect to the different central weight vectors. Naive implementation of the computation would require repeating the simulation m times, once for each alternative. Again, we can device a way to compute all integrals simultaneously. To do this, we first generate during each iteration a random criterion matrix from the appropriate criteria distribution. After that, we evaluate for each alternative whether that alternative is the most preferred one using its central weight vector and the random criterion matrix. This technique decreases the number of generated criterion matrices by a factor of m. However, to evaluate if the alternative is the most preferred one, we still need to evaluate the value function a maximum of m times in an inner loop. We shall see later a surprising result on the expected complexity of this algorithm. The algorithm for Phase 2 is presented as pseudo-code in Algorithm 4.

Algorithm 4. Monte Carlo simulation to compute the confidence factors (p_i^c's).

Output: p_i^c's
1: **for** $i \leftarrow 1$ **to** m **do**
2: $p_i^c \leftarrow 0$
3: **end for**
4: **for** $j \leftarrow 1$ **to** K_c **do**
5: $x \leftarrow \mathrm{RAND}_X()$
6: **for** $i \leftarrow 1$ **to** m **do**
7: $t \leftarrow u(x_i, w_i^c)$
8: **for all** $k \in \{1,\ldots,m\}\backslash\{i\}$ **do**
9: **if** $u(x_k, w_i^c) > t$ **then**
10: **goto** worse
11: **end if**
12: **end for**
13: $p_i^c \leftarrow p_i^c + 1$
14: worse:
15: **end for**
16: **end for**
17: **for** $i \leftarrow 1$ **to** m **do**
18: $p_i^c \leftarrow p_i^c / K_c$
19: **end for**

4. Complexity of the SMAA algorithm

If all of the criteria are measured on cardinal scales, the complexity of the algorithm for Phase 1 (Algorithm 3) is $O(K_w \cdot \phi_W + K_w \cdot \phi_X + K_w \cdot m \cdot n + K_w \cdot m\log(m) + m^2)$, where ϕ_W is the complexity of generating a weight vector from the weight distribution and ϕ_X is the complexity of generating a criterion matrix from the criteria distribution. In many applications the weights are generated from a uniform distribution following the method described before (Algorithm 2). In practice the number of iterations $K_w \gg m$, because K_w is fairly large (10^4–10^6) to obtain sufficient accuracy and m is fairly small. With these assumptions the complexity can be written as $O(K_w \cdot (n\log(n) + m \cdot n + \phi_X + m\log(m)))$. If criteria measurements are independent, the complexity is $O(K_w \cdot (n\ \log(n) + m \cdot n + m\log(m)))$. In many practical decision-making problems the term $m \cdot n$ dominates, and the complexity can thus be written as $O(K_w \cdot m \cdot n)$. If some of the criteria are measured on ordinal scales, the ordinal to cardinal mapping is required for those criteria, and in that case the total complexity of Algorithm 2 is $O(K_w \cdot n \cdot m\log(m))$.

The complexity of the algorithm for Phase 2 (Algorithm 4) is $O(K_c \cdot (\phi_X + m^2 \cdot n))$, if all of the criteria are measured on cardinal scales. During each iteration, the algorithm first generates a criterion matrix for all alternatives, and then uses these when computing the values with different central weight vectors. This decreases the number of criterion matrices generated during Phase 2 of the algorithm from $K_c \cdot m$ to K_c and affects the running time remarkably. If the criteria measurements are independent, the complexity of the algorithm for Phase 2 can be written as $O(K_c \cdot m^2 \cdot n)$.

The algorithm for Phase 2 which has squared worst-case complexity with respect to m, has in fact quite low typical-case complexity. The squared complexity is a consequence of the inner loop comparing values of the alternatives. However, that loop is almost never executed completely. Normally, when the criteria measurements are independent stochastic variables, the values of the alternatives will be distinct. If we assume that during each Monte Carlo iteration one of the alternatives will have the best value with respect to its central weight vector, one the second best, etc., and if the alternatives are compared in a random order, then the

expected number of value function evaluations during each Monte Carlo iteration is

$$(1+(m-1))+\left(1+\frac{m}{2}\right)+\left(1+\frac{m}{3}\right)+\cdots+\left(1+\frac{m}{m}\right) = m-1+m\sum_{i=1}^{m}\frac{1}{i} = m-1+m\cdot H_m. \quad (15)$$

In this formula, the mth subsum of the harmonic series is known as the harmonic number H_m. H_m grows very slowly, so the typical-case complexity of the algorithm is not squared with respect to m, but rather close to linear. The typical-case complexity can be written as $O(K_c \cdot H_m \cdot m \cdot n)$. Again, if criteria are measured on ordinal scales, the complexity of the ordinal to cardinal mapping procedure increases the total typical-case complexity of Algorithm 4 to $O(K_c \cdot H_m \cdot n \cdot m\log(m))$.

5. Accuracy of the SMAA computations

The accuracy of the results can be calculated by considering the Monte Carlo simulations as point estimators for b_i^r and p_i^c. By the central limit theorem we can conclude that b_i^r and p_i^c are normally distributed, if the numbers of iterations (K_w, K_c) are large enough (>25) (Milton and Arnold, 1995). In practical SMAA computations the number of iterations is typically 10^4–10^6.

If we want to achieve accuracy d_b with 95% confidence for b_i^r, we need the following number of Monte Carlo iterations K_w (Milton and Arnold, 1995):

$$K_w = \frac{1.96^2}{4d_b^2}. \quad (16)$$

For example, if we want to achieve error limits of 0.01 for b_i^r, that can be accomplished with 95% confidence by performing $K_w = 9604$ Monte Carlo iterations.

The accuracy of p_i^c depends on the accuracy of the central weight vectors and the criteria distribution in a complex manner. In theory, an arbitrarily small error in a central weight vector may cause an arbitrarily large error in a confidence factor. If we disregard this error source for the confidence factors, then the same accuracy analysis applies for the confidence factors as for the rank acceptability indices, that is, $K_c = K_w$ yields the same precision for the confidence factors as for the rank acceptability indices.

The accuracy of w_i^c does not depend on the total number of Monte Carlo iterations, but rather on the number of iterations that contribute to the computation of that central weight vector. To achieve an accuracy of d_w with 95% confidence for w_i^c, the required number of iterations is

$$K_c = \frac{1.96^2}{a_i \cdot 4d_w^2}. \quad (17)$$

Thus, alternatives with small acceptability indices require more iterations to compute their central weight vectors with a given accuracy. In practice, we are normally not interested in central weight vectors for alternatives with extremely low acceptability indices.

6. Empirical tests

We have performed empirical tests to measure the running time of the algorithm separately for Phases 1 and 2. The tests were performed on a GNU/Linux personal computer with one 2.6 GHz Pentium-4 processor and no significant extra load during the tests.

Our test problems include all combinations of the number of alternatives m and number of criteria n, where

$$m \in \{4, 6, 8, 10, 15, 25, 50, 100, 150, 200\} \quad \text{and}$$
$$n \in \{4, 8, 16, 32\}.$$

For each problem size we generated six sample problems: three with uniformly distributed cardinal criteria measurements (SMAA-2), and three with ordinal criteria measurements (SMAA-O). The cardinal criteria measurements were uniformly distributed in the intervals $[x_{ij} - 0.2, x_{ij} + 0.2]$ where the mean x_{ij} was chosen randomly from the interval $[0, 1]$. The ordinal measurements had distinct random rank levels for all criteria. For both SMAA-2 and SMAA-O, one problem contained no preference information (NOP), another included complete ranking of the criteria (ORDP), and the third contained preference information in form of weight intervals $[1/(n*2), 0.5]$ for all criteria (INTP). Thus, we have a total of $10 \cdot 4 \cdot 6 = 240$ test problems.

For each test run we used 10 000 Monte Carlo iterations ($K_w = K_c = 10\,000$). Based on Eq. (16) this yields an accuracy of $d < 0.01$ for the rank acceptability indices, which is sufficient in many real-life applications.

We executed the SMAA algorithm 20 times for all test problems. After that, we calculated the mean of the running time for each model. Total running times are presented in Fig. 8. The times for test runs with weight intervals are not shown in the figure, because including weight intervals had no observable impact on the running time when compared with no preference information. From Fig. 8 we can see that when the problem size grows, SMAA-O is clearly slower than SMAA-2. Still, both methods are fast enough to be used in practical decision-making situations. Inclusion of ordinal preference information does not significantly increase the running time of SMAA-O, and imposes only a minor increase to the running time of SMAA-2.

To analyze the complexity of Phases 1 and 2 in more detail, we have computed the running times in milliseconds divided by the product of number of alternatives and criteria. We have plotted these times as stacked columns for SMAA-2 in Fig. 9 and for SMAA-O in Fig. 10. Note that some combinations of (m,n) result in duplicate labels on the x-axis. From these figures it can be seen, that the

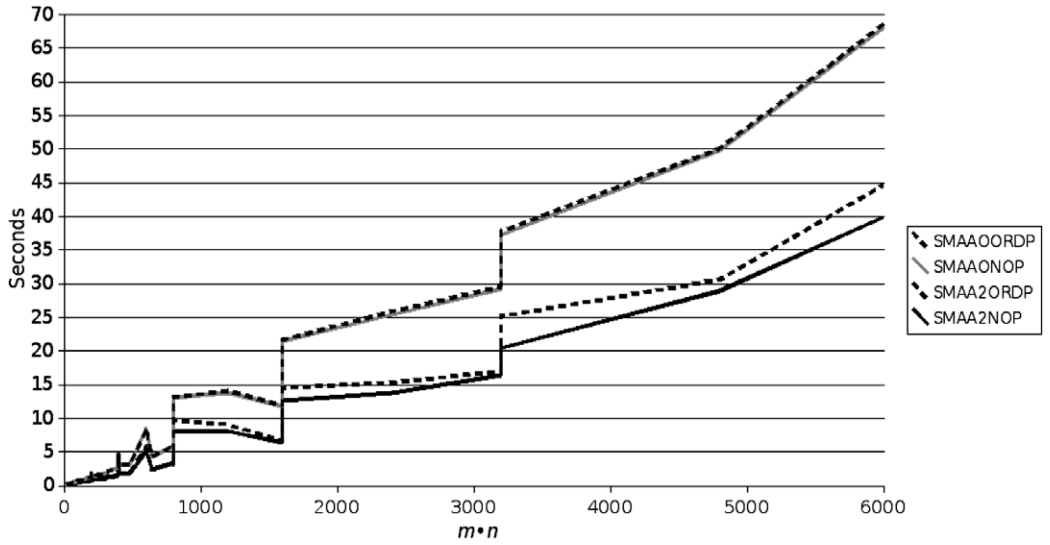


Fig. 8. Total running times of tests.

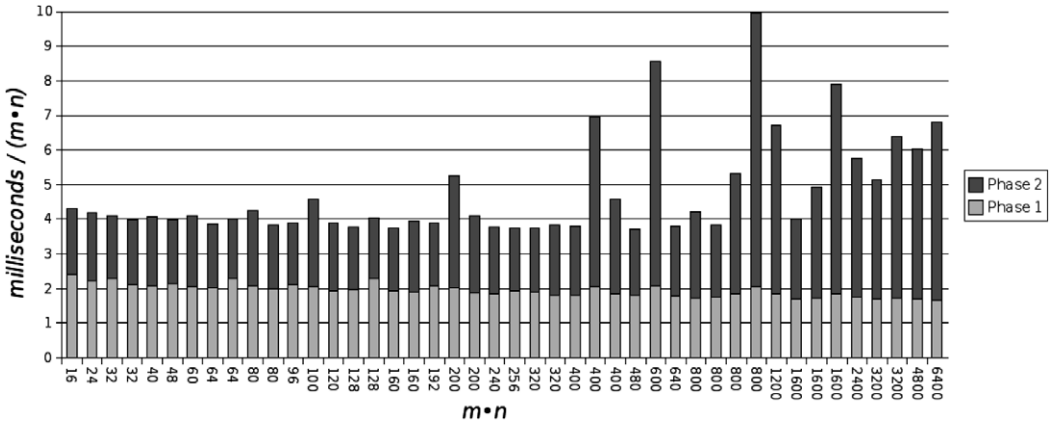


Fig. 9. Ratio of running time of SMAA-2 (in milliseconds) and product of number of alternatives and criteria.

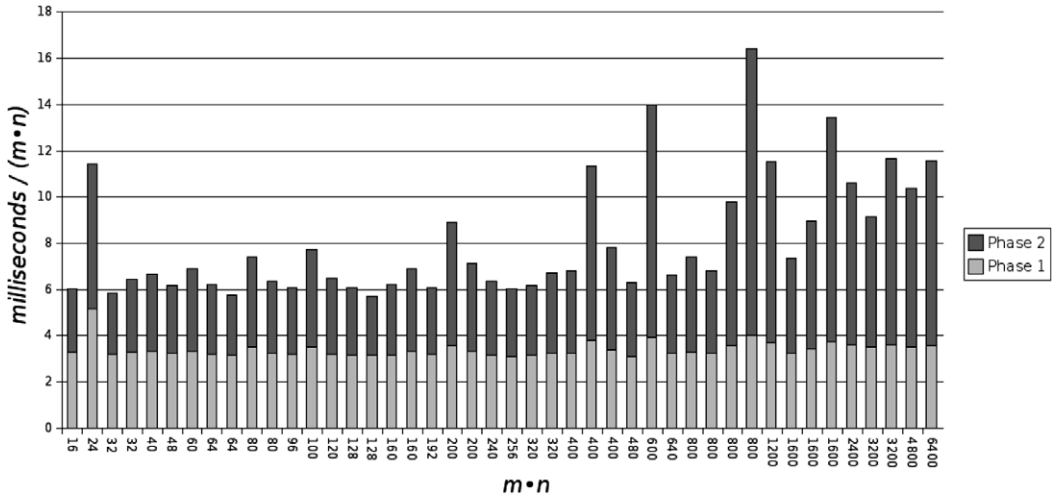


Fig. 10. Ratio of running time of SMAA-O (in milliseconds) and product of number of alternatives and criteria.

Phase 1 indeed has linear running time in respect to the number of criteria and alternatives. The running time for Phase 2 grows a little faster than the factor H_m would indicate. This is due to the fact that our implementation does not randomize the order in which the alternatives are compared in the innermost loop.

The empirical test results show that SMAA methods are applicable in MCDA problems with a large number of criteria and alternatives. In a typical MCDA ranking or choosing problem, there are under 20 alternatives and criteria. In this case, the execution of the algorithm with a personal computer takes only a few seconds. It should also be noted, that the total execution time grows almost linearly with respect to the number of alternatives and criteria. Therefore, these algorithms can be used also with very large decision-making problems (for example, over 100 criteria and over 1000 alternatives).

7. Conclusions

Stochastic multicriteria acceptability analysis is a family of methods for aiding multicriteria group decision making in problems with inaccurate, uncertain, or missing information. The multidimensional integrals which form core of the SMAA computations are in practice impossible to compute analytically. We have demonstrated that the computations can be implemented efficiently with sufficient accuracy using Monte Carlo simulation techniques. With cardinal criteria, the computation time is nearly proportional to $n \cdot m$. With ordinal criteria, the computation time is nearly proportional to $n \cdot m \cdot \log(m)$.

In a group decision-making process, it is common that new preference information is received and old information is adjusted as the process evolves. When new information is added to the model, the SMAA computations must be repeated. The empirical efficiency tests of the presented implementations show that the required time for computing a typical decision-making problem with 10 alternatives and eight criteria with a personal computer is less than a second. Thus, the effect of modified preference information on the results can be investigated interactively by the decision makers.

Some decision-making problems are continuous by nature, and the number of alternatives is thus in principle infinite. One approach to solve such problems is to form a large number of discrete decision alternatives and to evaluate them using discrete decision support methods. SMAA methods can help in such processes to filter out alternatives that are inefficient or otherwise inferior. The results in this paper show that SMAA methods are fast enough also to be used in such, fairly large problems.

Acknowledgements

The work of Tommi Tervonen was partially supported by the MONET research project (POCTI/GES/37707/2001) and grants from Turun Yliopistosäätiö and the Finnish Cultural Foundation.

References

Bana e Costa, C.A., 1986. A multicriteria decision aid methodology to deal with conflicting situations on the weights. European Journal of Operational Research 26, 22–34.

Charnetski, J., Soland, R., 1978. Multiple-attribute decision making with partial information: The comparative hypervolume criterion. Naval Research Logistics Quarterly 25, 279–288.

David, H.A., 1970. Order Statistics. Wiley, New York.

Dias, L., Clímaco, J., 2000. ELECTRE TRI for groups with imprecise information on parameter values. Group Decision and Negotiation 9 (5), 355–377.

Dias, L., Mousseau, V., Figueira, J., Clímaco, J., 2002. An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI. European Journal of Operational Research 138, 332–348.

Durbach, I., 2006. A simulation-based test of stochastic multicriteria acceptability analysis using achievement functions. European Journal of Operational Research 170 (3), 923–934.

Figueira, J., Greco, S., Ehrgott, M. (Eds.), 2005. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science+Business Media, Inc., New York.

Fishburn, P., 1965. Analysis of decisions with incomplete knowledge of probabilities. Operations Research 13, 217–237.

Fishman, G., 1996. Monte Carlo: Concepts, Algorithms, and Applications. Springer-Verlag, New York.

Hazen, G., 1986. Partial information, dominance, and potential optimality in multiattribute utility theory. Operations Research 34, 297–310.

Hokkanen, J., Lahdelma, R., Miettinen, K., Salminen, P., 1998. Determining the implementation order of a general plan by using a multicriteria method. Journal of Multi-Criteria Decision Analysis 7 (5), 273–284.

Hokkanen, J., Lahdelma, R., Salminen, P., 1999. A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor. Socio-Economic Planning Sciences 33, 1–23.

Hokkanen, J., Lahdelma, R., Salminen, P., 2000. Multicriteria decision support in a technology competition for cleaning polluted soil in Helsinki. Journal of Environmental Management 60, 339–348.

Kangas, A., Kangas, J., Lahdelma, R., Salminen, P., in press. Using SMAA-2 method with dependent uncertainties for strategic forest planning. Forest Policy and Economics.

Kangas, J., Kangas, A., 2003. Multicriteria approval and SMAA-O method in natural resources decision analysis with both ordinal and cardinal criteria. Journal of Multi-Criteria Decision Analysis 12, 3–15.

Kangas, J., Hokkanen, J., Kangas, A., Lahdelma, R., Salminen, P., 2003. Applying stochastic multicriteria acceptability analysis to forest ecosystem management with both cardinal and ordinal criteria. Forest Science 49 (6), 928–937.

Kirkwood, C., Sarin, R., 1985. Ranking with partial information: A method and an application. Operations Research 33, 38–48.

Lahdelma, R., Salminen, P., 2001. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Operations Research 49 (3), 444–454.

Lahdelma, R., Salminen, P., 2002. Pseudo-criteria versus linear utility function in stochastic multi-criteria acceptability analysis. European Journal of Operational Research 14, 454–469.

Lahdelma, R., Salminen, P., 2006. Classifying efficient alternatives in SMAA using cross confidence factors. European Journal of Operational Research 170 (1), 228–240.

Lahdelma, R., Hokkanen, J., Salminen, P., 1998. SMAA—Stochastic multiobjective acceptability analysis. European Journal of Operational Research 106, 137–143.

Lahdelma, R., Salminen, P., Simonen, A., Hokkanen, J., 2001. Choosing a reparation method for a landfill using the SMAA-O multicriteria method. In: Multiple Criteria Decision Making in the New Millenium. Lecture Notes in Economics and Mathematical Systems, vol. 507. Springer-Verlag, Berlin, pp. 380–389.

Lahdelma, R., Salminen, P., Hokkanen, J., 2002. Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. European Journal of Operational Research 142, 345–356.

Lahdelma, R., Miettinen, K., Salminen, P., 2003. Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA). European Journal of Operational Research 147, 117–127.

Lahdelma, R., Makkonen, S., Salminen, P., 2004. Treating dependent uncertainties in multi-criteria decision problems. In: Rao, M.R., Puri, M.C. (Eds.), Operational Research and Its Applications: Recent Trends, vol. II, New Delhi, pp. 427–434.

Lahdelma, R., Miettinen, K., Salminen, P., 2005. Reference point approach for multiple decision makers. European Journal of Operational Research 164 (3), 785–791.

Milton, J.S., Arnold, J.C., 1995. Introduction to Probability and Statistics. Probability and Statistics, 3rd ed. McGraw-Hill International editions.

Mousseau, V., Słowiński, R., Zielniewicz, P., 2000. A user-oriented implementation of the ELECTRE-TRI method integrating preference elicitation support. Computers & Operations Research 27 (7–8), 757–777.

Mousseau, V., Figueira, J., Dias, L., Gomes da Silva, C., Clímaco, J., 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. European Journal of Operational Research 147, 72–93.

Tervonen, T., Almeida-Dias, J., Figueira, J., Lahdelma, R., Salminen, P., 2005. SMAA-TRI: A parameter stability analysis method for ELECTRE TRI. Research Report 6/2005 of The Institute of Systems Engineering and Computers (INESC-Coimbra), Coimbra, Portugal. Available from: <http://www.inescc.pt>.

Publication II

Tervonen T., Hakonen H., Lahdelma R., 2008. Elevator planning with stochastic multicriteria acceptability analysis. Omega 36(3), 352–362.

Available online at www.sciencedirect.com
ScienceDirect
Omega 36 (2008) 352–362
omega
The International Journal of Management Science
www.elsevier.com/locate/omega

Elevator planning with stochastic multicriteria acceptability analysis☆

Tommi Tervonen[a,b,*], Henri Hakonen[b,c], Risto Lahdelma[b]

[a]CEG-IST, Centre for Management Studies, Instituto Superior Técnico, Departamento de Engenharia e Gestão, 1049-001 Lisbon, Portugal
[b]Department of Information Technology, University of Turku, FIN 20520 Turku, Finland
[c]Kone Oyj, Keilasatama 3, FIN 02150 Espoo, Finland

Received 29 September 2004; accepted 26 April 2006
Available online 27 November 2006

Abstract

Modern elevator systems in high-rise buildings consist of groups of elevators with centralized control. The goal in elevator planning is to configure a suitable elevator group to be built. The elevator group must satisfy specific minimum requirements for a number of standard performance criteria. In addition, it is desirable to optimize the configuration in terms of other criteria related to the performance, economy and service level of the elevator group. Different stakeholders involved in the planning phase emphasize different criteria. Most of the criteria measurements are by nature uncertain. Some criteria can be estimated by using analytical models, while others, especially those related to the service level in different traffic patterns, require simulations.

In this paper we formulate the elevator planning problem as a stochastic discrete multicriteria decision-making problem. We compare 10 feasible elevator group configurations for a 20-floor building. We evaluate the criteria related to the service level in different traffic situations using the KONE Building Traffic Simulator, and use analytical models and expert judgments for other criteria. The resulting decision problem contains mixed type criteria. Some criteria are represented by the multivariate Gaussian distribution, others by deterministic values and ordinal (ranking) information. To identify configurations that can best satisfy the goals of the stakeholders, we analyze the problem using the stochastic multicriteria acceptability analysis (SMAA) method.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Stochastic multicriteria acceptability analysis (SMAA); Elevator planning; Multicriteria; Simulation

1. Introduction

In modern high-rise buildings workers and inhabitants are transported between floors mainly by means of multiple elevators. Elevators are usually operated by elevator group control systems in order to provide efficient transportation. When a high-rise building is designed, a suitable configuration for the elevator group has to be designed. The decision makers (DMs) should consider performance as well as price and other non-performance criteria of alternative elevator group configurations. Because analytical methods are limited to the up-peak traffic situation and cannot evaluate the effect of a group control algorithm, the performance has to be measured using computer simulation, which produces stochastic measurements for the performance criteria of alternative configurations. The performance of an elevator group can be measured using several

☆ This paper was processed by Guest Editors Margaret M. Wiecek, Matthias Ehrgott, Georges Fadel and José Rui Figueira.

* Corresponding author. CEG-IST, Centre for Management Studies, Instituto Superior Técnico, Departamento de Engenharia e Gestão, 1049-001 Lisbon, Portugal.
E-mail addresses: tommi.tervonen@it.utu.fi (T. Tervonen), henri.hakonen@kone.com (H. Hakonen), risto.lahdelma@cs.utu.fi (R. Lahdelma).

0305-0483/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.omega.2006.04.017

criteria, such as the average waiting time (WT) or the average ride time of the passengers. The price and other non-performance criteria can usually be assessed with sufficient accuracy or by ranking the alternatives.

Different DMs may have different preferences for the criteria. For example, some DMs pay attention to the average WT while others think that the percentage of long WTs is more important since it represents the fairness in service. The builder may stress the amount of floor space used by the elevator system. There are usually some trade-offs and dependencies between criteria. The problem of elevator planning can thus be considered as a discrete multicriteria decision-making problem with multiple DMs and stochastic criteria measurements. We are interested in finding a compromise solution which takes into account different possible preferences of DMs, and thus we have chosen to analyze the problem using the SMAA method.

SMAA methods have been developed for discrete multicriteria decision-making problems, where criteria measurements are uncertain or inaccurate and where it is for some reason difficult to obtain accurate or any preference information from the DMs [1]. Usually the preference information is modelled by determining importance weights for criteria. 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. In the original SMAA method [2] the weight space analysis is performed based on an additive utility or value function and stochastic criteria measurements. The SMAA-2 method [1] generalized the analysis to a general utility or value function, to include various kinds of preference information and to consider holistically all ranks. The SMAA-O method [3] extends SMAA-2 for treating mixed ordinal and cardinal criteria in a comparable manner. SMAA is suitable for solving problems also when the uncertainties of criteria measurements are dependent [4].

Elevator planning research has a long history. The operative performance has been studied over decades [5,6]. The up-peak interval and the up-peak handling capacity has been analyzed in many publications in the 1960s, see, e.g. [7–9]. The patience of passengers and what should be considered good service in different types of buildings has been studied since 1940s according to [10]. Earliest applications of simulation to elevator planning are from the 1960s [11,12]. There are also more recent applications in all areas of elevator planning, but in practice normal elevator groups are still designed using methods from the 1960s. In this paper we present a multicriteria method that allows to use stochastic simulator output in the decision analysis. We consider a realistic elevator planning problem, which consists of a 20-floor building for which one of 10 possible elevator group configurations has to be chosen. We will analyze the alternative configurations using the KONE Building Traffic Simulator. Based on the output of the simulator, we form a multicriteria decision-making problem, which we analyze using SMAA. To our best knowledge, we are the first to apply a stochastic MCDA method in elevator planning. We have chosen to use SMAA in the decision analysis, because it is the only MCDA methodology that allows multivariate Gaussian distributed criteria measurements.

This paper is organized as follows: Section 2 introduces the reader to the area of elevator planning, and Section 3 to SMAA methods. In Section 4, we present the simulator that is used to generate the data, and the simulation results. We define the decision-making problem and present the SMAA analysis in Section 5. Section 6 ends this paper with conclusions.

2. Elevator planning

The goal in elevator planning is to find a suitable elevator group to serve the traffic of a high-rise building. Because the buildings do not exist at the planning stage, the traffic must be estimated by using the building specifications: the number of floors, their heights, the floor area and the building type. The travel height can be calculated from the number of floors and their heights, and the total population can be estimated according to the type of building and the floor area. Building types have characteristic *traffic profiles*. For example, office buildings typically have up-peak traffic in the morning when employees enter the building, intense two-way or inter-floor traffic during the lunch time, and down-peak traffic when employees exit the building [13].

The performance of a group of elevators is mainly determined by the number and size of the cars and their speed. Also acceleration, door types and the group control algorithm affect performance. Usual performance criteria are the handling capacity and the interval calculated in the up-peak situation. The *up-peak handling capacity* is the percentage of population per 5 min that can be transported from the lobby to the upper floors. It is assumed that elevators are filled to 80% of rated load (although it is possible to fill elevator up to rated load that does not happen in practice). The (up-peak) *interval* is an interval between two starts from the lobby. The interval is also related to the WT. The up-peak is used since it is the most demanding situation considering elevator handling capacity at least in office buildings,

and because there are analytical formulas for calculating the up-peak handing capacity and interval [14]. The usual recommendations state that the up-peak handling capacity for an office building should be 11–17% and interval 20–30s [15].

Non-performance criteria, such as cost and occupied floor area should also be considered. The *cost* of an elevator system consists of build and maintenance costs. The *floor area* occupied by the elevator group consists of the shaft space and the waiting area for passengers. In high-rise buildings the population is large and distances are long, thus the portion of shafts is large compared to the total floor area. This means more costs, since the rentable area is reduced. In some cases the building design constrains the occupied area, sometimes there is more freedom to use space. The elevator planning is not independent of building design; the architect should take advice from the elevator planner.

Instead of considering only up-peak traffic, we take into account the entire daily traffic and consider all criteria simultaneously. In the study presented in this paper we consider the following six criteria. The cost and area criteria take into account the building owners point of view. Passengers point of view is taken into account by WT, journey time (JT), the percentage of WTs exceeding 60s (WT60), and the percentage of JTs exceeding 120s (JT120). The WT is measured from the moment a passenger enters the waiting area to the moment he/she enters the elevator. The JT is the total time from entering the waiting area to exiting the elevator. The last two criteria measure unsatisfactory service, which may happen especially in intense traffic peaks.

3. The SMAA methods

The SMAA-2 method [1] has been developed for discrete stochastic multicriteria decision-making problems with multiple DMs. SMAA-2 applies inverse weight space analysis to describe for each alternative what kind of preferences make it the most preferred one, or place it on any particular rank. The decision problem is represented as a set of m alternatives $\{x_1, x_2, \ldots, x_m\}$ that are evaluated in terms of n criteria. The DMs' preference structure is represented by a real-valued utility or value function $u(x_i, w)$. The value function maps the different alternatives to real values by using a weight vector w to quantify DMs' subjective preferences. SMAA-2 has been developed for situations where neither criteria measurements nor weights are precisely known. Uncertain or imprecise criteria are represented by stochastic variables ξ_{ij} with joint density function $f_X(\xi)$ in the space $X \subseteq R^{m\times n}$. We denote the stochastic criteria measurements of alternative x_i with ξ_i. The DMs' unknown or partially known preferences are represented by a weight distribution with joint density function $f_W(w)$ in the feasible weight space W. Total lack of preference information is represented in 'Bayesian' spirit by a uniform weight distribution in W, that is, $f_W(w) = 1/\mathrm{vol}(W)$. The weight space can be defined according to needs, but typically, the weights are non-negative and normalized, that is; the weight space is an $n-1$ dimensional simplex in n dimensional space:

$$W = \left\{ w \in R^n : w \geqslant 0 \quad \text{and} \quad \sum_{j=1}^{n} w_j = 1 \right\}. \tag{1}$$

The value function is used to map the stochastic criteria and weight distributions into value distributions $u(\xi_i, w)$. Based on the value distributions, the rank of each alternative is defined as an integer from the best rank (=1) to the worst rank (=m) by means of a ranking function

$$\mathrm{rank}(i, \xi, w) = 1 + \sum_{k=1}^{m} \rho(u(\xi_k, w) > u(\xi_i, w)), \tag{2}$$

where $\rho(true) = 1$ and $\rho(false) = 0$. SMAA-2 is then based on analyzing the stochastic sets of favorable rank weights

$$W_i^r(\xi) = \{w \in W : \mathrm{rank}(i, \xi, w) = r\}. \tag{3}$$

Any weight $w \in W_i^r(\xi)$ results in such values for different alternatives, that alternative x_i obtains rank r.

The first descriptive measure of SMAA-2 is the *rank acceptability index* b_i^r, which measures the variety of different preferences that grant alternative x_i rank r. It is the share of all feasible weights that make the alternative acceptable for a particular rank, and it is most conveniently expressed percentage wise. The rank acceptability index b_i^r is computed numerically as a multidimensional integral over the criteria distributions and the favorable rank weights as

$$b_i^r = \int_{\xi\in X} f_X(\xi) \int_{w\in W_i^r(\xi)} f_W(w)\,\mathrm{d}w\,\mathrm{d}\xi. \tag{4}$$

The most acceptable (best) alternatives are those with high acceptabilities for the best ranks.

The *central weight vector* w_i^c is the expected center of gravity (centroid) of the favorable first rank weights of an alternative. The central weight vector represents

the preferences of a 'typical' DM supporting this alternative. The central weights of different alternatives can be presented to the DMs in order to help them understand how different weights correspond to different choices with the assumed preference model. The central weight vector w_i^c is computed numerically as a multidimensional integral over the criteria distributions and the favorable first rank weights using

$$w_i^c = \int_{\xi \in X} f_X(\xi) \int_{w \in W_i^1(\xi)} f_W(w) w \, \mathrm{d}w \, \mathrm{d}\xi / b_i^1. \quad (5)$$

The *confidence factor* p_i^c is the probability for an alternative to obtain the first rank when the central weight vector is chosen. The confidence factor is computed as a multidimensional integral over the criteria distributions using

$$p_i^c = \int_{\xi \in X : \mathrm{rank}(i, \xi, w_i^c) = 1} f_X(\xi) \, \mathrm{d}\xi. \quad (6)$$

Confidence factors can similarly be calculated for any given weight vectors. The confidence factors measure whether the criteria measurements are accurate enough to discern the efficient alternatives.

The uncertainty of the criteria measurements can be modelled very flexibly in SMAA methods by using an appropriate joint distribution $f_X(\xi)$. If the uncertainties are independent, then separate distributions $f_{ij}(\xi_{ij})$ can be used for each measurement. Simple parametric distributions, such as the uniform and normal distribution may be suitable in many applications. When the uncertainties of the criteria measurements are dependent, then the dependent parameters can be represented by a joint distribution. The multivariate Gaussian (normal) distribution is particularly suitable, because it is theoretically well understood and yet it approximates well many real-life phenomena. Use of the multivariate Gaussian distribution with SMAA is described in more detail in [4].

There are several different ways to handle partial preference information in SMAA methods [1]. In the decision-making problem considered in this paper, we apply interval constraints for weights. For more information about this technique, see [16].

4. Simulation model and simulation results

To obtain stochastic criteria measurements for the performance criteria, we executed simulations with the KONE Building Traffic Simulator [17,18]. The simulation model consists of the elevator model and traffic generation. Features of the model are:

- Floors have landing call buttons. Entering passenger gives a call to the (up/down) direction where he is heading.
- The group control algorithm allocates the call to the most suitable elevator. The algorithm is a genetic algorithm [19], which optimizes WTs. The group control has also a returning algorithm which sends the elevator back to the lobby to wait for a call. The returning algorithm is necessary in the incoming traffic situation.
- A stopping elevator opens doors, exiting passengers get out, entering passengers get in and the doors are closed. The simulator has delays related to door opening, entrance, exit and door closing.
- An elevator can take passengers up to the *maximum load*, which is about 80% of the cars rated load. If the load exceeds *bypass load* (about 80% of maximum load), an elevator does not accept new landing calls. The loads are expressed in persons.
- An elevator cannot reverse direction with passengers aboard.
- An elevator accelerates smoothly to the *rated speed*, provided that the distance is long enough. The smoothness is modelled by the acceleration derivative *jerk*, which is $1.6\,\mathrm{m/s^3}$. The deceleration is an inverse to the acceleration phase.
- The passengers arrive to different floors approximately according to a *Poisson process*. This means that the inter-arrival times follow the exponential distribution, $f(x) = \alpha e^{-\alpha x}$, where α is the arrival rate.
- There is one *entrance floor* and rest of the floors are *populated floors*. Traffic consists of three components: incoming, outgoing and inter-floor components. Incoming passengers travel from an entrance floor to populated floors, outgoing passengers from populated floors to the entrance floor and inter-floor passengers between populated floors. Intensity of traffic and the percentages of incoming, outgoing and inter-floor passengers are determined by traffic parameters.

The traffic profile determines the intensity and the portions of traffic components at each moment. The intensity is expressed as portion of population per time unit. The passengers are generated as follows:

(1) The simulator generates the expected number of passengers to a 5-min period and assigns them random entry times. The number of passengers is the total population multiplied by the traffic intensity.

Table 1
Characteristics of the simulated building

Characteristic	Value
Floors	20
Floor height (m)	4.2
Travel (m)	78
Floor area (m^2/floor)	1000
Rentable area (m^2/floor)	800
Persons per floor	60
Persons total	1140

(2) The traffic component of passenger (incoming, outgoing or inter-floor) is chosen randomly according to the traffic profile.
(3) The component determines whether the arrival and destination floors are entrance or populated floors:
   (a) If the floor is a populated floor, the probability of the floor is proportional to the floor population.
   (b) If independently generated arrival and destination floors happen to be equal (can happen with an inter-floor passenger), the floor generation is repeated.

Table 1 shows characteristics of the simulated building. The building has a lobby floor and 19 populated floors. The estimated number of people is 60 per floor.

Fig. 1 shows the intensities of incoming, outgoing and inter-floor passengers during the day from 7 a.m. to 7.15 p.m. The traffic profile is measured from an office building. The profile shows typical morning, lunch time and afternoon traffic peaks. When passengers are generated according to the traffic profile, the expected number of passengers are 11 502. Since total population of the building is uncertain, the traffic is varied between 80% and 120% of forecasted traffic. With these parameters, we generated 21 traffic situations according to the traffic profile. The same passengers were used for all 10 alternatives in order to reduce the covariance between the measurements of different alternatives.

Table 2
Alternative configurations of the elevator group

Name	Elevators	Rated load	Speed (m/s)	Acceleration (m/s^2)	Area (m^2)	Cost
E6L17S4	6	17	4.0	1.0	69.8	1
E6L21S4	6	21	4.0	1.0	77.4	2
E6L17S5	6	17	5.0	1.0	71.4	3
E6L24S4	6	24	4.0	1.0	87.2	4
E7L17S35	7	17	3.5	0.8	87.5	5
E7L17S4	7	17	4.0	1.0	87.5	6
E7L13S5	7	13	5.0	1.0	76.0	7
E7L17S5	7	17	5.0	1.0	89.5	8
E8L13S35	8	13	3.5	0.8	79.4	9
E8L17S35	8	17	3.5	0.8	93.5	10

Table 2 shows 10 alternative configurations. The number of elevators varies between 6 and 8, rated load from 13 to 24 and speed from 3.5 to 5 m/s. Area is the shaft space plus waiting area space. The exact costs are unknown. The costs are ranked from 1 to 10, where 1 is the cheapest and 10 is the most expensive. All alternatives are feasible with respect to up-peak handling capacity and interval.

Simulation results are presented in Figs. 2–5 as WTs, JTs, percentage of WT exceeding 60 s and percentage

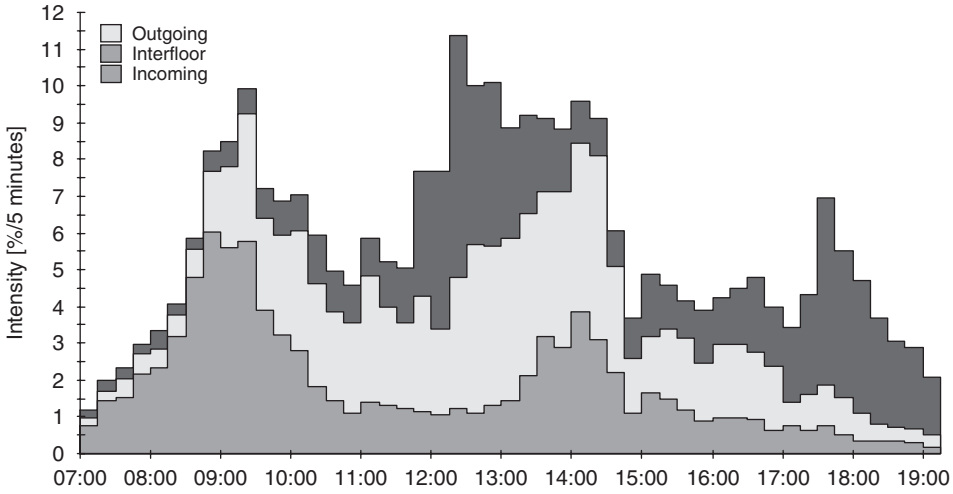


Fig. 1. Traffic profile of the simulated building [13].

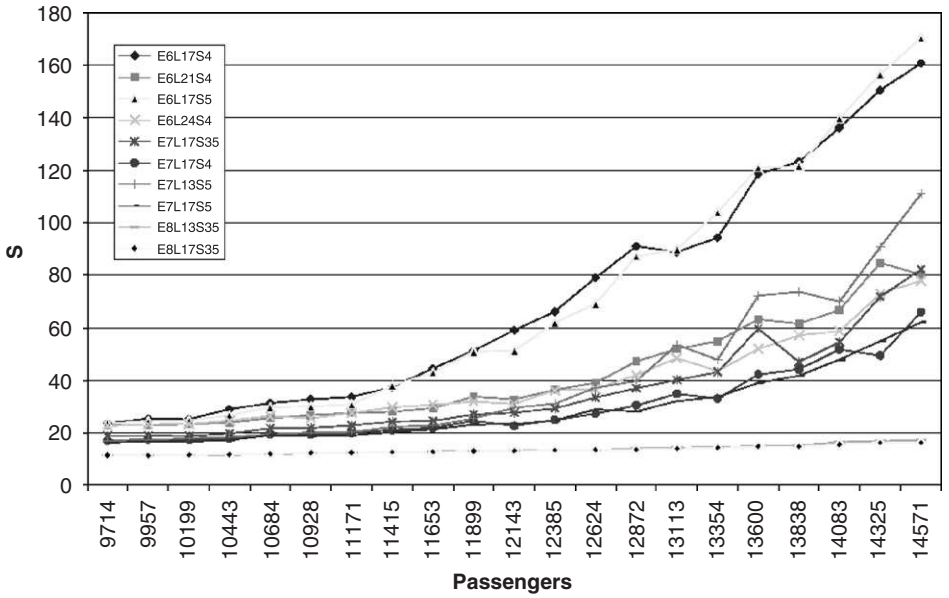


Fig. 2. Waiting times.

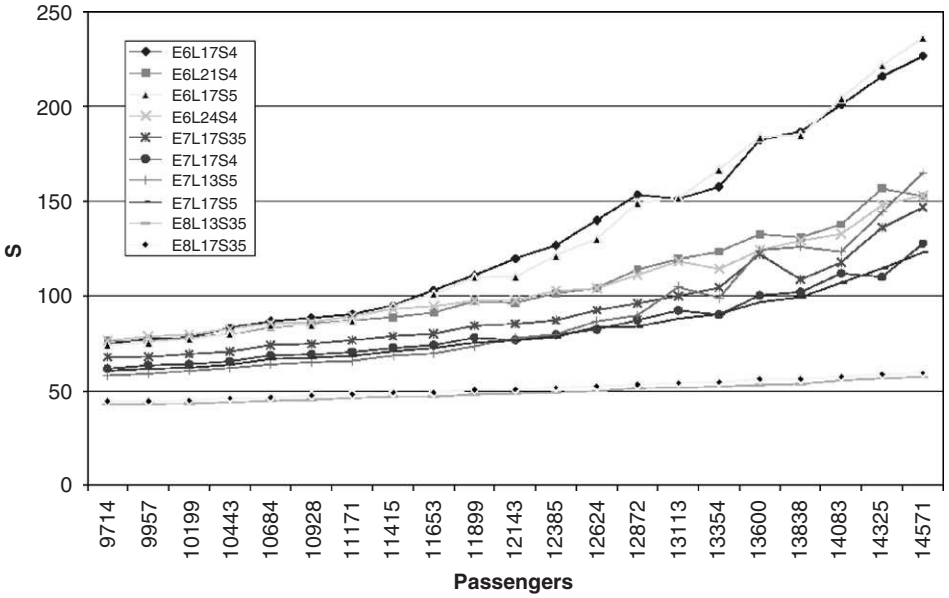


Fig. 3. Journey times.

of JTs exceeding 120 s. The passenger transfer times are closer to reality than the values normally used in elevator planning. This lowers the handling capacity and lengthens the waiting and JTs especially when the traffic is intensive. The horizontal axis shows the number of generated passengers in each of the 21 traffic situations. The figures seem quite similar, since the order of alternatives is approximately the same with all criteria and all intensities. The groups with eight elevators are more efficient than the smaller groups.

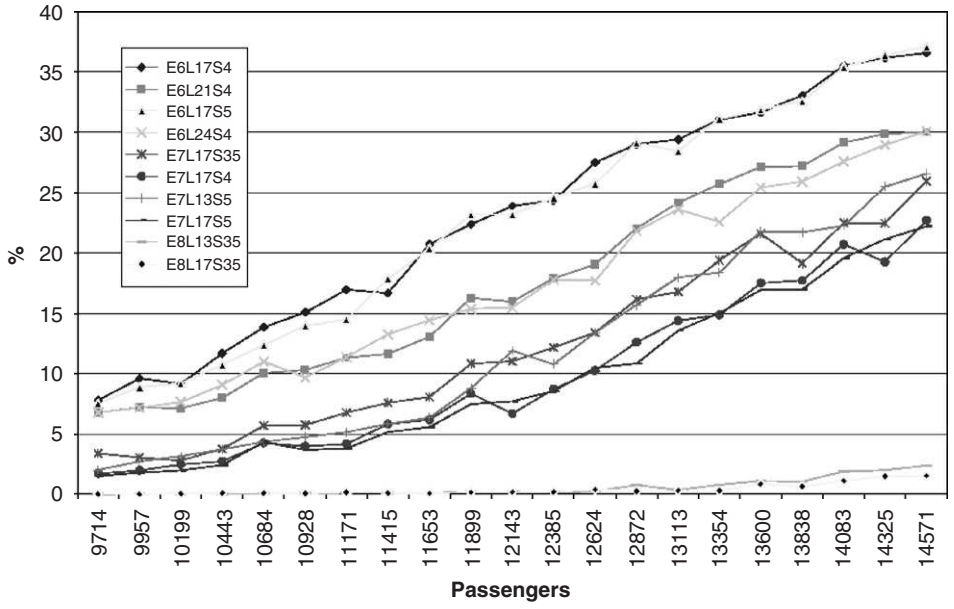


Fig. 4. Percentage of waiting times over 60 s.

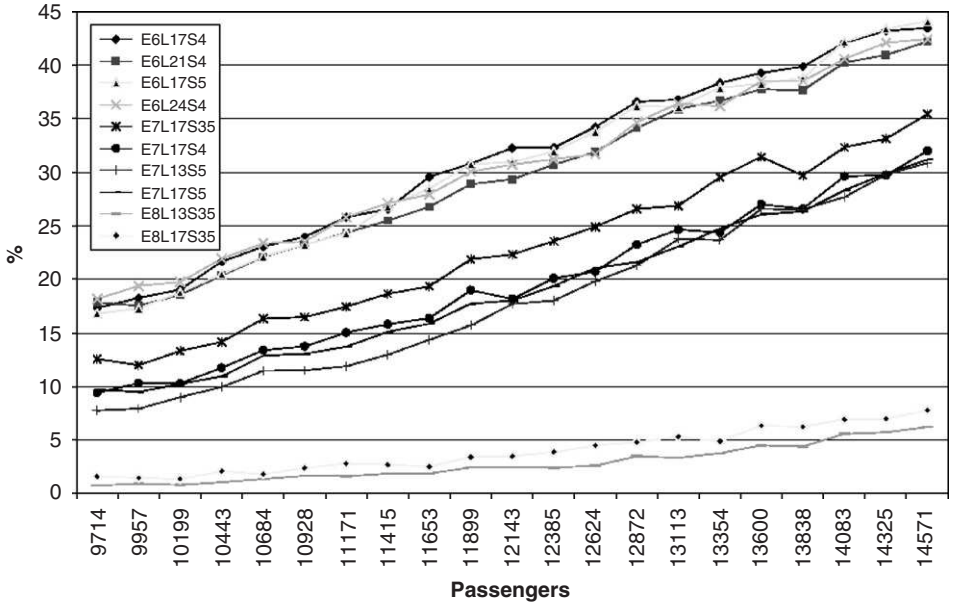


Fig. 5. Percentage of journey times over 120 s.

5. Decision-making problem and SMAA analysis

The uncertainties of the performance criteria were assessed based on the simulations for each of the 10 configurations. Based on the simulation results we estimated the parameters for a multivariate Gaussian distribution, i.e. the expected value of each criteria measurement and the covariance matrix (Λ) for the uncertainty dependencies. The uncertainties of the performance criteria were quite dependent, with multivariate correlations in the interval [0.8,1]. The cross-correlation matrix is not presented here because of its size (40×40).

Table 3
Values for performance criteria of alternatives (mean ± st.dev.)

Alternative	WT ($\mu \pm \sigma$)	JT ($\mu \pm \sigma$)	WT60 ($\mu \pm \sigma$)	JT120 ($\mu \pm \sigma$)
E6L17S4	71.52 ± 43.58	131.00 ± 47.63	22.95 ± 9.28	31.16 ± 8.29
E6L21S4	42.08 ± 19.12	105.53 ± 24.63	17.62 ± 8.11	29.64 ± 7.83
E6L17S5	71.19 ± 46.23	130.02 ± 50.46	22.57 ± 9.55	30.60 ± 8.52
E6L24S4	39.17 ± 16.01	104.87 ± 21.87	17.27 ± 7.45	30.48 ± 7.55
E7L17S35	35.46 ± 17.84	92.62 ± 22.24	12.30 ± 7.23	22.76 ± 7.19
E7L17S4	29.68 ± 13.35	83.42 ± 17.83	9.87 ± 6.61	19.59 ± 6.84
E7L13S35	40.84 ± 26.85	89.00 ± 30.29	12.05 ± 8.02	18.02 ± 7.37
E7L17S5	28.83 ± 12.99	81.65 ± 17.61	9.57 ± 6.63	18.98 ± 6.77
E8L13S35	13.58 ± 1.76	49.21 ± 4.43	0.56 ± 0.71	2.82 ± 1.64
E8L17S35	13.39 ± 1.47	51.40 ± 4.60	0.39 ± 0.46	3.98 ± 1.97

Table 4
Confidence factors (p^c) and rank acceptabilities (b_i) in percentages, sorted in decreasing order with respect to confidence factors

Alt	p^c	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
E8L13S35	87.12	68.36	6.45	4.77	5.46	5.30	2.36	2.37	3.21	1.72	0.00
E6L17S4	41.43	16.54	11.17	8.88	6.98	5.23	4.53	4.99	8.48	26.51	6.70
E6L21S4	40.01	6.54	12.70	15.88	14.26	14.34	10.75	19.05	4.87	1.60	0.02
E7L13S5	6.27	4.54	14.01	22.73	15.75	14.65	10.21	7.62	6.81	2.76	0.93
E6L17S5	5.68	3.31	8.44	10.59	7.84	6.40	4.85	5.02	6.92	13.47	33.16
E7L17S4	2.92	0.42	1.51	20.66	19.94	15.04	16.48	12.66	8.58	3.86	0.85
E7L17S35	2.32	0.28	0.85	3.04	8.80	18.62	27.61	18.30	11.56	8.24	2.71
E8L17S35	0.79	0.01	44.57	7.96	4.43	3.87	5.01	5.06	5.72	8.34	15.05
E6L24S4	0.48	0.00	0.29	0.92	3.18	4.06	7.28	12.45	29.99	15.20	26.63
E7L17S5	0.00	0.00	0.02	4.56	13.37	12.50	10.93	12.48	13.87	18.32	13.96

The cost was modelled as an ordinal criterion (see [3]), because exact price information was not available. The required floor area was measured on a cardinal scale with 5 m^2 uncertainty for all alternatives. The measurements for cost and area criteria are presented in Table 2. The criteria measurements for the performance criteria are presented in Table 3.

We defined a decision-making problem using the four dependent performance criteria, one ordinal criteria (price) and one cardinal criteria (area). We also added preference information in form of weight bounds to the model; weights for cost and shaft space were constrained to be in the interval [0.1,1]. The preference information was added to the model because of the strong dependencies between performance criteria, which shows that they all ultimately measure a single criterion, performance from the passengers point of view. Because of the additivity of weights, the performance would obtain too high significance in the analysis without balancing accomplished by using weight constraints.

We analyzed the model using 100 000 Monte-Carlo iterations, which gives error limits $\leqslant 0.01$ [16]. Results of the SMAA computations are presented in Tables 4 and 5. Acceptability indices are illustrated graphically in Fig. 6, and central weights as stacked columns in Fig. 7. Notice that central weight vector is not defined for alternative E7L17S5, because it has confidence factor of 0.

By inspecting the SMAA results in Table 4, we can see that the last three alternatives (E8L17S35, E6L24S4, and E7L17S5) can be rejected as feasible alternatives because of their near-zero confidence factors. From the remaining seven alternatives, E7L17S4 and E7L17S35 should also be rejected because they have reasonably small confidence factors. After this initial analysis, we have already eliminated half of the alternatives, and the rest have confidence factors in the range 5–87%.

Next we will examine the rank acceptabilities of the alternatives. Alternative E6L17S5 has low rank acceptability for the best ranks (3.31% for rank 1, 8.44% for rank 2), and high acceptability for the worst ranks (13.47% for rank 9, 33.16% for rank 10), and thus we choose to reject it. Alternative E8L17S35 has the highest first rank acceptability index (b_1), but it also is the most expensive of the four remaining alternatives. By looking at the central weights in Table 5, we can see

Table 5
Confidence factors (p^c) and central weights, sorted in decreasing order with respect to confidence factors

Alt	p^c	WT	JT	WT60	JT120	Cost	Area
E8L13S35	87.12	15.07	15.30	15.54	15.63	17.43	21.03
E6L17S4	41.43	9.79	9.92	9.65	9.90	33.96	26.79
E6L21S4	40.01	12.15	10.19	9.87	7.98	46.28	13.53
E7L13S5	6.27	14.74	14.47	15.54	14.37	13.71	27.17
E6L17S5	5.68	13.26	12.21	10.51	10.03	18.62	35.36
E7L17S4	2.92	13.70	11.86	15.46	10.99	42.52	5.46
E7L17S35	2.32	13.69	12.56	14.23	11.43	43.23	4.86
E8L17S35	0.79	26.61	9.97	27.20	7.52	28.21	0.48
E6L24S4	0.48	25.28	9.45	8.81	5.27	49.47	1.71

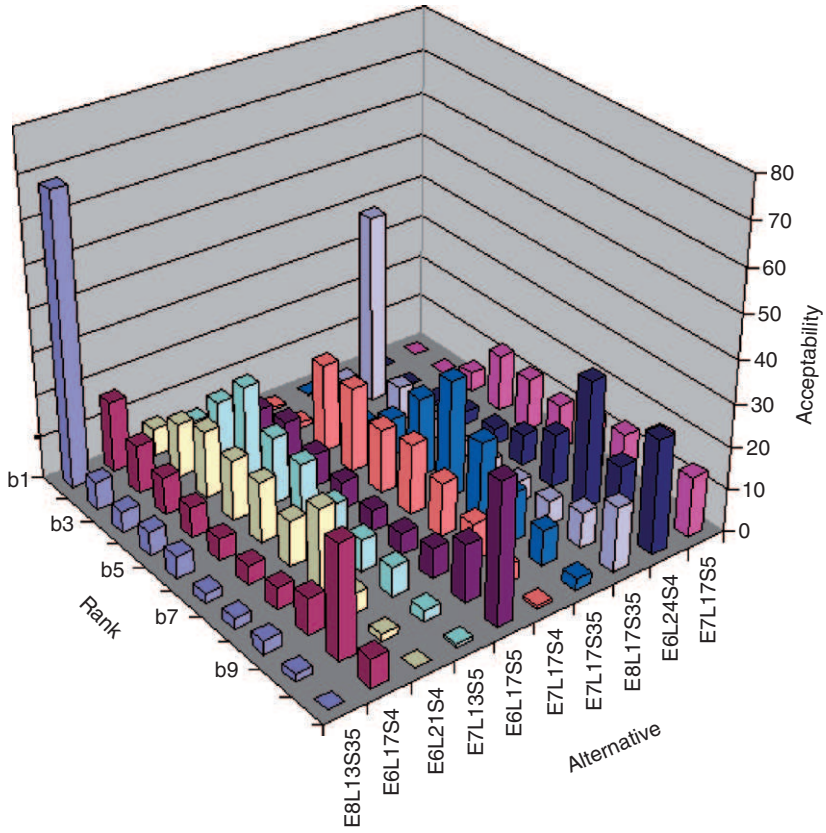


Fig. 6. Rank acceptability indices.

that it obtains the highest weight values for performance criteria, and low values for cost and area. If price and floor area usage are not important, this alternative is the most preferable. The high first rank acceptability index follows from the additivity of the weights; the performance criteria obtain relatively more weight than non-performance criteria, because they ultimately all measure the same criteria (performance from the passengers point of view).

For the other three alternatives, E6L17S4 and E6L21S4 have reasonably high confidence factors (40–42%) when compared to alternative E7L13S5 (6%), which is clearly the most compromising of the four remaining alternatives. E6L17S4 and E6L21S4 are the two cheapest and smallest of the configurations, which is also seen in their central weights (high values for cost and area). If these two criteria (cost and space) are considered important, either of these alternatives

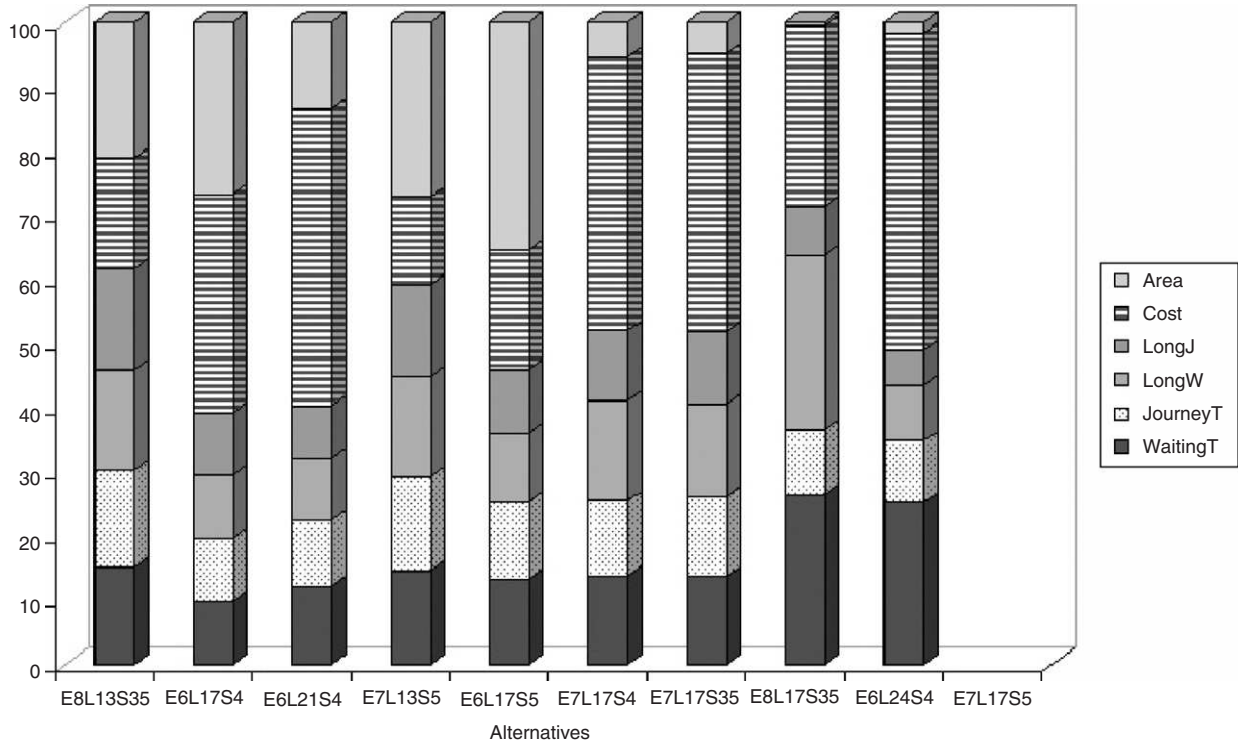


Fig. 7. Central weight vectors.

should be chosen. By looking at their central weights, the choice between these two could be based on importance of cost compared to the importance of area (E6L17S4 has smaller weight for cost, higher for area). When looking for a compromise solution, E6L17S4 should be avoided, because it has high acceptability, 26%, for rank 9.

Final conclusion of the analysis is, that the DMs should choose one from alternatives E8L17S35, E6L17S4, E6L21S4, and E7L13S5. If a compromise alternative is preferred, E7L13S5 should be chosen.

6. Conclusions

Stochastic multicriteria acceptability analysis (SMAA) has a history of real-life cases where it has helped in multicriteria decision-making problems. In this paper we presented a realistic elevator planning problem where SMAA helps in determining the elevator group to be built on a high-rise building. In elevator planning, the DMs might be representatives of the elevator company, consultants or customers. These groups of DMs usually have clashing preferences. For example, elevator company might favor expensive and effective elevator group configurations while the customer prefers a low-cost alternative. We presented the KONE Building Traffic Simulator and the traffic simulation results which were used in determining the SMAA model.

We analyzed 10 alternative configurations for an elevator group using SMAA. Based on our analysis of the results, from the original 10 alternatives, only four are acceptable, and if a compromise solution is preferred, there exists a single solution for the decision-making problem. Results of our analysis show, that SMAA is effective on recognising acceptable solutions in elevator planning, and furthermore on determining which of those are compromise solutions and which are favored by different groups of DMs.

Multivariate Gaussian distribution was used for modelling the dependent performance criteria. Future research should explore the effects of using different distributions and their effects on the SMAA results. We hope that this paper gives rise to the usage and research of SMAA methods in elevator planning.

Acknowledgments

The work of Tommi Tervonen was supported by grants from Turun Yliopistosäätiö, Finnish Cultural

Foundation, and the MONET research project (POCTI/GES/37707/2001, Fundação para Ciência e Tecnologia, Portugal). Tommi Tervonen was working a part of the time of this research in INESC-Coimbra, Portugal.

References

[1] Lahdelma R, Salminen P. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Operations Research 2001;49(3):444–54.
[2] Lahdelma R, Hokkanen J, Salminen P. SMAA—stochastic multiobjective acceptability analysis. European Journal of Operational Research 1998;106:137–43.
[3] Lahdelma R, Miettinen K, Salminen P. Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA). European Journal of Operational Research 2003;147:117–27.
[4] Lahdelma R, Makkonen S, Salminen P. Multivariate Gaussian criteria in SMAA. European Journal of Operational Research 2006;170(3):957–70.
[5] Basset J. The probable number of stops made by an elevator. GE Rev 1923;26(8).
[6] Strakosch GR. Vertical transportation. New York: Wiley; 1967.
[7] Morley CG. Installations, mechanical: lifts, escalators, paternosters. Architect's Journal 1962;136:233–58.
[8] Phillips RS. Electric lifts. 5th ed., London: Pitman; 1966.
[9] Pinfold WJ. Lift design and user requirements. In: Symposium on lifts in large buildings, Building Research Station, Watford; 1966.
[10] Tregenza PR. The prediction of lift performance in multi-storey buildings. PhD thesis, University of Nottingham; 1971.
[11] Browne JJ, Kelly JJ. Simulation of elevator system for world's tallest buildings. Transportation Science 1968;2(1):35–6.
[12] Parlow H. Lift operation and computers, a simulation of performance. Architect's Journal 1966;153:747–53.
[13] Siikonen ML, Leppälä J. Elevator traffic pattern recognition. In: Proceedings of the fourth world congress of the international fuzzy systems association, Brussels, Belgium; 1991. p. 195–8.
[14] Barney GC, dos Santos SM. Elevator traffic analysis design and control. London, UK: Peter Peregrinus Ltd; 1985.
[15] Barney GC, Cooper DA, Inglis J. Elevator & escalator micropedia. UK: The International Association of Elevator Engineers; 1998.
[16] Tervonen T, Lahdelma R. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research, 2006, to appear.
[17] Hakonen H. Simulation of building traffic and evacuation by elevators, Licenciate thesis, System Analysis Laboratory, Helsinki University of Technology; 2003.
[18] Leinonen R. Building traffic simulator. Master's thesis, Helsinki University of Technology; 1999.
[19] Tyni T, Ylinen J. Genetic algorithms in elevator car routing problem. In: Proceedings of the genetic and evolutionary computation conference (GECCO-2001); 2001. p. 1413–22.

Publication III

Tervonen T., Figueira J. R., Lahdelma R., Almeida Dias J., Salminen P., 2007. A stochastic method for robustness analysis in sorting problems. European Journal of Operational Research (in press).

Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research xxx (2007) xxx–xxx
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Decision Support

A stochastic method for robustness analysis in sorting problems

Tommi Tervonen [a,b,*], José Rui Figueira [a], Risto Lahdelma [b], Juscelino Almeida Dias [a], Pekka Salminen [c]

[a] Centre for Management Studies, Instituto Superior Técnico, Technical University of Lisbon, 2780-990 Porto Salvo, Portugal
[b] Department of Information Technology, University of Turku, FIN-20520 Turku, Finland
[c] School of Business and Economics, University of Jyväskylä, P.O. Box 35, FIN-40014 Jyväskylä, Finland

Received 9 February 2006; accepted 6 September 2007

Abstract

ELECTRE TRI is a multiple criteria decision aiding sorting method with a history of successful real-life applications. In ELECTRE TRI, values for certain parameters have to be provided. We propose a new method, SMAA-TRI, that is based on stochastic multicriteria acceptability analysis (SMAA), for analyzing the stability of such parameters. The stability analysis can be used for deriving robust conclusions. SMAA-TRI allows ELECTRE TRI to be used with uncertain, arbitrarily distributed values for weights, the lambda cutting level, and profiles. The method consists of analyzing finite spaces of arbitrarily distributed parameter values. Monte Carlo simulation is applied in this in order to describe for each alternative the share of parameter values that have it assigned to different categories. We show the real-life applicability by re-analyzing a case study in the field of risk assessment.
© 2007 Elsevier B.V. All rights reserved.

Keywords: MCDA; ELECTRE TRI; Robustness analysis; Monte Carlo simulation, Risk assessment

1. Introduction

Partitioning a set of objects into groups (clusters, classes, or categories) is among the most researched areas in various disciplines. The groups can be defined *a priori* or *a posteriori* and be ordered or not. In the case of *a priori* defined ordered groups the problem is called an *ordinal classification* or *sorting* problem, and the objects are assigned to *categories* based on upper and lower profiles, central objects or other norms (Doumpos and Zopounidis, 2002).

In the late seventies a trichotomy procedure for sorting problems based on the outranking approach was proposed by Moscarola and Roy (1977). Several years later, in order to help decision making in a large banking company faced with a problem of accepting or refusing credit requests, a new method with a name of ELECTRE A was developed and applied in 10 sectors of activity. Based on these earlier works, in 1992 a method called ELECTRE TRI (Yu, 1992) emerged. It is one of the most successful and applied methods for multiple criteria decision aiding (MCDA) sorting problems, subject to constant application and development (see Figueira et al., 2005).

ELECTRE TRI requires an input of numerous parameters. The parameters can be divided into preference parameters (relative importance coefficients of criteria or weights, thresholds, and category profiles) and a technical parameter (the lambda cutting level). The weight elicitation process in general is one of the most difficult problems in MCDA. There are numerous weight elicitation techniques proposed for ELECTRE methods, see e.g. Mousseau (1995), Hokkanen and Salminen (1997), Figueira and Roy (2002), Rogers and Bruen (1998), Mousseau et al.

* Corresponding author. Address: Centre for Management Studies, Instituto Superior Técnico, Technical University of Lisbon, 2780-990 Porto Salvo, Portugal. Tel.: +351 96 529 1326.
E-mail addresses: tommi.tervonen@it.utu.fi (T. Tervonen), figueira@ist.utl.pt (J.R. Figueira), risto.lahdelma@cs.utu.fi (R. Lahdelma), judias@fe.uc.pt (J.A. Dias), psalmine@econ.jyu.fi (P. Salminen).

0377-2217/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2007.09.008

(2001). All these techniques produce different values for weights, and therefore it is advisable to perform some kind of robustness analysis when they are applied (Roy, 2002, 2005).

Important research has been made about parameter inference and robustness analysis for ELECTRE TRI. The first advance on this topic was by Mousseau and Słowiński (1998), who presented a method for inferring indifference and preference thresholds, profiles and weights, from assignment examples (a set of training examples) through non-linear optimization. A linear programming method for inferring the weights from assignment examples was introduced by Mousseau et al. (2001), and implemented in a user-friendly software (see Mousseau et al., 2000). Dias and Clímaco (1999, 2000) presented a method for deriving robust conclusions with uncertain information on parameter values. These works were combined into a unified framework by Dias et al. (2002) that allows to infer the parameters and to derive robust conclusions based on assignment examples. Nevertheless, inconsistent judgements can appear when these methods are applied. Algorithms for solving interactively the inconsistencies were proposed by Mousseau et al. (2003, 2004). An approach for inferring category limits was introduced by Ngo The and Mousseau (2002). This methodology was complemented by Dias and Mousseau (2006), that introduced a partial inference procedure for inferring the veto-related parameters.

In this paper, we introduce the SMAA-TRI method for analyzing the robustness of ELECTRE TRI results based on a parameter stability analysis. A *parameter stability analysis* consists of analyzing a space of feasible parameters for possible changes in the output of the method. Stability analysis allows the model to include non-deterministic parameters and provides the DMs with more output than parameter inference. The main contribution of this work is in introducing a method for applying ELECTRE TRI with uncertain information that can be used for parameter stability analysis, but we also give meaning to the indices derived from simulations used in SMAA-TRI.

SMAA-TRI is based on Stochastic Multicriteria Acceptability Analysis (SMAA) (Lahdelma et al., 1998; Lahdelma and Salminen, 2001, 2002; Tervonen et al., 2007). SMAA is a family of decision support methods to aid decision makers (DMs) in discrete decision making problems. SMAA methods for the ranking problem statement are based on inverse weight space analysis that produces descriptive values characterizing the decision making problem. They have been applied in numerous real-life situations. For a survey on SMAA methodology, see Tervonen and Figueira (in press). SMAA-TRI is the first SMAA method for the sorting problem statement.

The rest of the paper is organized as follows. A description of ELECTRE TRI is presented in Section 2. SMAA-TRI is introduced in Section 3. We demonstrate the application of SMAA-TRI by re-analyzing a case study in Section 4. We end the paper with conclusions in Section 5.

2. ELECTRE TRI

ELECTRE TRI was designed to assign a set of alternatives $\{a_1,\ldots,a_i,\ldots,a_m\}$ to pre-defined and ordered categories $C_1,\ldots,C_h,\ldots,C_k$ based on a set of criteria $\{g_1,\ldots,g_j,\ldots,g_n\}$. Each category C_h is characterized by both a lower (b_{h-1}) and an upper (b_h) profile. The profiles are limits for the categories and are treated in a same manner as alternatives. The profiles must be connected with a dominance relation as follows:

$$b_k \Delta b_{k-1} \Delta \ldots \Delta b_h \Delta \ldots \Delta b_1 \Delta b_0, \tag{1}$$

meaning that all criteria evaluations must be lower for a dominated profile (“lower” profile). The assignment of an alternative to a certain category results from the comparison of the alternative with the profiles with respect to the outranking assertion “the alternative is at least as good as the profile (or vice versa)”. This means that the alternative is at least as good as the profile on a sufficient set of criteria (the concordance condition), and is not exceedingly worse on any criterion (the veto condition). In what follows, we will assume, without any loss of generality, that the scales of the criteria are ascending (that is, all criteria are to be maximized).

Fulfillment of the concordance condition with respect to individual criterion is measured taking into account two thresholds: the indifference and the preference threshold. These are used to model the fact that the preferences between two levels of criteria evaluations usually do not change abruptly and small differences can be judged as indifferent. Also, imprecision can be taken into account when defining thresholds.

Indifference threshold on criterion g_j (denoted as $p_j(\cdot)$) defines the maximum difference in scores for which an alternative and a profile are considered to be indifferent with respect to this criterion. The preference threshold on criterion g_j (denoted as $q_j(\cdot)$) defines the minimum difference in scores so that a profile is strictly preferred to the alternative (or vice versa). In this case the concordance condition is not fulfilled. Between these two points lies a zone in which the profile is preferred to the alternative, but in a less strict sense. The thresholds can be constant of defined based on either the score of an alternative or of a profile. Without loss of generality, we will assume constant thresholds throughout this paper.

The concordance condition with respect to a single criterion is measured with a partial concordance index. It is computed as follows for criterion g_j:

$$c_j(a_i,b_h)=\begin{cases}1, & \text{if } g_j(a_i)\geqslant g_j(b_h)-q_j(g_j(b_h)),\\ 0, & \text{if } g_j(a_i)<g_j(b_h)-p_j(g_j(b_h)),\\ \frac{g_j(a_i)+p_j(g_j(b_h))-g_j(b_h)}{p_j(g_j(b_h))-q_j(g_j(b_h))}, & \text{otherwise.}\end{cases} \tag{2}$$

The first part of the equation corresponds to the case where the difference is less than the indifference threshold or the alternative has a higher score. The second part corresponds to the case when the score of the profile exceeds the score of

Please cite this article in press as: Tervonen, T. et al., A stochastic method for robustness analysis in sorting problems, European Journal of Operational Research (2007), doi:10.1016/j.ejor.2007.09.008

the alternative with at least the indifference threshold. In the zone between strict preference and indifference, the index is calculated with a linear interpolation. Computation of the partial concordance index is illustrated graphically in Fig. 1.

The partial concordance indices are aggregated to a normalized weighted sum to form a comprehensive concordance index. The weights in ELECTRE TRI have a meaning of being relative importance coefficients or "votes" in favour of criteria. The higher the weight of a criterion is, the more important it is. The comprehensive concordance index is:

$$c(a_i, b_h) = \sum_{j \in J} w_j c_j(a_i, b_h). \tag{3}$$

The fulfillment of the veto condition is measured with a discordance index. This takes into account the veto thresholds. These define the minimum difference for a criterion to be totally discordant ("raises veto") with the assertion "a_i is at least as good as b_h". The discordance index is calculated as follows for criterion g_j:

$$d_j(a_i, b_h) = \begin{cases} 1, & \text{if } g_j(a_i) < g_j(b_h) - v_j(g_j(b_h)) \\ 0, & \text{if } g_j(a_i) \geqslant g_j(b_h) - p_j(g_j(b_h)) \\ \frac{g_j(b_h) - g_j(a_i) - p_j(g_j(b_h))}{v_j(g_j(b_h)) - p_j(g_j(b_h))}, & \text{otherwise.} \end{cases} \tag{4}$$

In the zone between complete and no veto, the index is calculated with linear interpolation, as illustrated in Fig. 2.

After the concordance and discordance indices are computed, the credibility index ρ of the outranking assertion can be computed as follows:

$$\rho(a_i, b_h) = \begin{cases} c(a_i, b_h) \prod_{j \in V} \frac{1 - d_j(a_i, b_h)}{1 - c(a_i, b_h)}, & \text{if } V \neq \emptyset, \\ c(a_i, b_h), & \text{otherwise,} \end{cases} \tag{5}$$

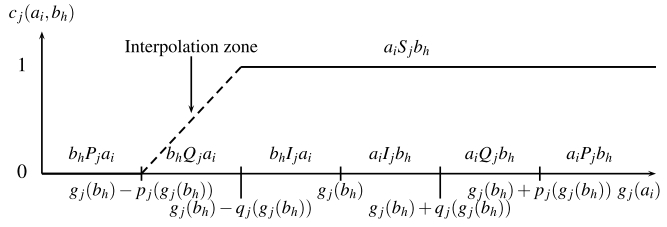


Fig. 1. The partial concordance index $c_j(a_i, b_h)$.

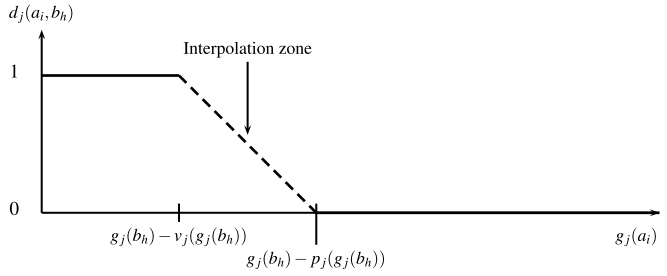


Fig. 2. The partial discordance index $d_j(a_i, b_h)$.

with

$$V = \{j \in \{1, \ldots, n\} : d_j(a_i, b_h) > c(a_i, b_h)\}. \tag{6}$$

Notice that when $d_j(a_i, b_h) = 1$ for any $j = 1, \ldots, n$, this implies that $\rho(a_i, b_h) = 0$. This means that one completely discordant criterion makes this assertion to be invalid.

After determining the credibility index a *λ-cutting level* has to be defined. The cutting level is used to transform the "fuzzy" outranking relation into a crisp one. It is defined as the smallest credibility index value compatible with the assertion $a_i S b_h$ so that if $\rho(a_i, b_h) \geqslant \lambda$, then $a_i S b_h$. λ should be in the range $[0.5, 1]$, and it describes the minimum sum of weights of criteria that must support the assertion $a_i S b_h$.

Based on the outranking relations between all pairs of alternatives and profiles, two different exploitation procedures can be applied for assigning the alternatives into the categories: the pessimistic and the optimistic procedure. In the pessimistic procedure, an alternative a_i is successively compared with $b_k, b_{k-1}, \ldots$, until $a_i S b_{k-1}$. Then a_i is assigned to the best category C_h such that $a_i\ S b_{h-1}$. For description of the optimistic procedure and it is prerequisites, see Figueira et al. (2005).

3. SMAA-TRI

The fundamental idea of SMAA is to use Monte Carlo simulation for exploring the weight space in order to provide DMs with values characterizing the problem. The SMAA methodology has been developed for discrete stochastic MCDA problems with multiple DMs. The SMAA-2 method (Lahdelma and Salminen, 2001) applies inverse weight space analysis to describe the preferences that make an alternative the most preferred one, or place it on any particular rank. In SMAA, the criteria evaluations can be generated based on arbitrary distributions, or they can be sampled from an external source.

SMAA-TRI is developed for parameter stability analysis of ELECTRE TRI and consists of analyzing finite spaces of arbitrarily distributed parameter values in order to describe the share of parameter values that assign an alternative to different categories. We analyze the stability of weights, profiles, and the cutting level, and consider the remaining parameters to have deterministic values for easier comprehensibility. The method can easily be extended to consider non-deterministic values for thresholds.

The approach we are taking is not the only possible one for uncertain parameter values. There also exists entropy methods for this, see Jessop (1999) and Abbas (2006).

For analyzing ELECTRE TRI we will denote the input for ELECTRE TRI in SMAA-TRI as follows:

(1) Uncertain profiles are represented by stochastic variables ϕ_{hj} with a joint density function $f_\Phi(\phi)$ in the space $\Phi \subseteq R^{k-1 \times n}$. The joint density function must be such that all possible profile combinations satisfy

Please cite this article in press as: Tervonen, T. et al., A stochastic method for robustness analysis in sorting problems, European Journal of Operational Research (2007), doi:10.1016/j.ejor.2007.09.008

(1). Usually the category profiles are defined to be independently distributed and in this case the distributions must not overlap. For example, if the profile values for a criterion are Gaussian distributed, the tails of the distributions must be truncated as shown by the vertical lines in Fig. 3.

(2) Lambda cutting level is presented by a stochastic variable Λ with a density function $f_L(\Lambda)$ defined within the valid range $[0.5, 1]$.

(3) Weights are represented by a weight distribution with a joint density function $f_W(w)$ in the feasible weight space W. Total lack of preference information is represented by a uniform weight distribution in W, that is, $f_W(w) = 1/\mathrm{vol}(W)$. The weights are non-negative and normalized: the weight space is an $n-1$ dimensional simplex in an n dimensional space:

$$W = \left\{ w \in R^n : w \geqslant 0 \text{ and } \sum_{j=1}^{n} w_j = 1 \right\}. \tag{7}$$

(4) Data and other parameters of ELECTRE TRI are represented by the set $T = \{M, q, p, v\}$. These components are considered to have deterministic values.

There has been related work on representing uncertain parameters through joint distributions. For more information on it, see for example Rao and Sobel (1980) and Barron and Barret (1996). In practice the parameters of SMAA-TRI are usually represented with Gaussian distributions or uniform distributions in a certain interval (that is, with an upper and a lower bound).

SMAA-TRI produces category acceptability indices for all pairs of alternatives and categories. The category acceptability index π_i^h describes the share of possible parameter values that assigns an alternative a_i to category C_h. It is most conveniently expressed percentage-wise. The index is a generalization of the rank acceptability index of SMAA-2. Let us define a *categorization function* that gives the category index h to which an alternative a_i is assigned by ELECTRE TRI:

$$h = K(i, \Lambda, \phi, w, T), \tag{8}$$

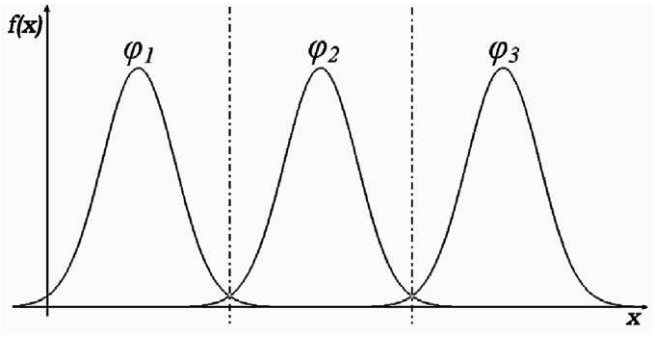


Fig. 3. Probability distribution functions for three Gaussian distributed profile values (for a single criterion). The horizontal lines show where the tails of the distributions must be truncated.

and a category membership function

$$m_i^h(\lambda, \phi, w, T) = \begin{cases} 1, & \text{if } K(i, \Lambda, \phi, w, T) = h, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The categorization function is applied in computing the category acceptability index π_i^h numerically as a multidimensional integral over the finite parameter spaces as

$$\pi_i^h = \int_{0.5}^{1} f_L(\Lambda) \int_{\phi \in \Phi} f_\Phi(\phi) \int_{w \in W} f_W(w) m_i^h(\Lambda, \phi, w, T) \mathrm{d}w \mathrm{d}\phi \mathrm{d}\Lambda. \tag{10}$$

The category acceptability index measures the stability of the assignment and it can be interpreted as a probability for membership in the category. Evidently, the category acceptability indices are within the range $[0, 1]$, where 0 indicates that the alternative will never be assigned to the category, and 1 indicates that it will be assigned to the category with any combination of feasible parameter values. The acceptabilities for each alternative sum to unity. If the parameters are stable, the category acceptability indices for each alternative should be 1 for one category, and 0 for the others. In this case the assignments are said to be robust with respect to the uncertain parameters.

The category acceptability indices provide a measure of uncertainty for the results of the sensitivity and robustness analyses as they were considered in ELECTRE TRI before. While traditional way to perform sensitivity analysis in ELECTRE TRI (not in line with the general definition of sensitivity analysis) is to consider only the extremes of what can be defined to be possible values for the uncertain parameters (Merad et al., 2004), the category acceptability indices consider the whole space that can be determined with arbitrary joint probability distributions. This type of complete sensitivity analysis has been considered with respect to multi-attribute utility theory by Butler et al. (1997), but SMAA-TRI is the first approach to apply it to ELECTRE methods.

Therefore, while robustness analysis for ELECTRE TRI (Dias et al., 2002) provides a result such as "depending on the parameter values, the alternative is assigned either to category 2 or 3", the SMAA-TRI provides the result as "the alternative is assigned to category 2 with 5% of the feasible parameter values, and to category 3 with 95% of the feasible parameter values".

There are three advantages gained with the additional information:

(1) The cognitive effort required in determining the extremes of the parameters considered in the sensitivity analysis is reduced, because the space can be determined to be, for example, uniformly distributed and thus small changes in the value intervals do not change the results dramatically.

(2) Quantifying the amount of parameter values that result in "unstable" assignment determines the risk related with uncertain parameters. This will later be demonstrated in the re-analysis of the case study.

Please cite this article in press as: Tervonen, T. et al., A stochastic method for robustness analysis in sorting problems, European Journal of Operational Research (2007), doi:10.1016/j.ejor.2007.09.008

(3) Weight elicitation techniques provide different weight values, and thus it seems more relevant to elicit the weights as uncertain values rather than as deterministic ones.

In addition to providing parameter stability analysis, SMAA-TRI also allows ELECTRE TRI to be applied when multiple DMs with conflicting preferences participate in the decision making processes. The method allows arbitrarily distributed weights, and therefore they can be defined, for example, as intervals containing the preferences of all DMs (Lahdelma and Salminen, 2001). In this case the results of the analysis can be used to find assignments accepted by majority of the DMs. Modelling group preferences with intervals containing the preferences of all DMs is somewhat more justified in SMAA-TRI than in utility theory-based models (SMAA, SMAA-2), because the weights are criteria importance coefficients and not scaling factors. Also the extremes of parameter combinations that assign alternatives to certain categories can be computed simultaneously with the parameter stability analysis.

The category acceptability indices are computed through Monte Carlo simulation quite similarly as in SMAA-2. The algorithms for SMAA-2 together with analyses of complexity and running times have been presented by Tervonen and Lahdelma (2007). Biggest part of complexity of the SMAA-TRI algorithm consists of the weight generation. Considering the nature of the weights (criteria importance coefficients), the most suitable weight constraints are ranking of the weights or upper/lower bounds for individual weights. These can be handled with quite efficient techniques. For details, see Tervonen and Lahdelma (2007). The SMAA-TRI simulation scheme is presented in Fig. 4.

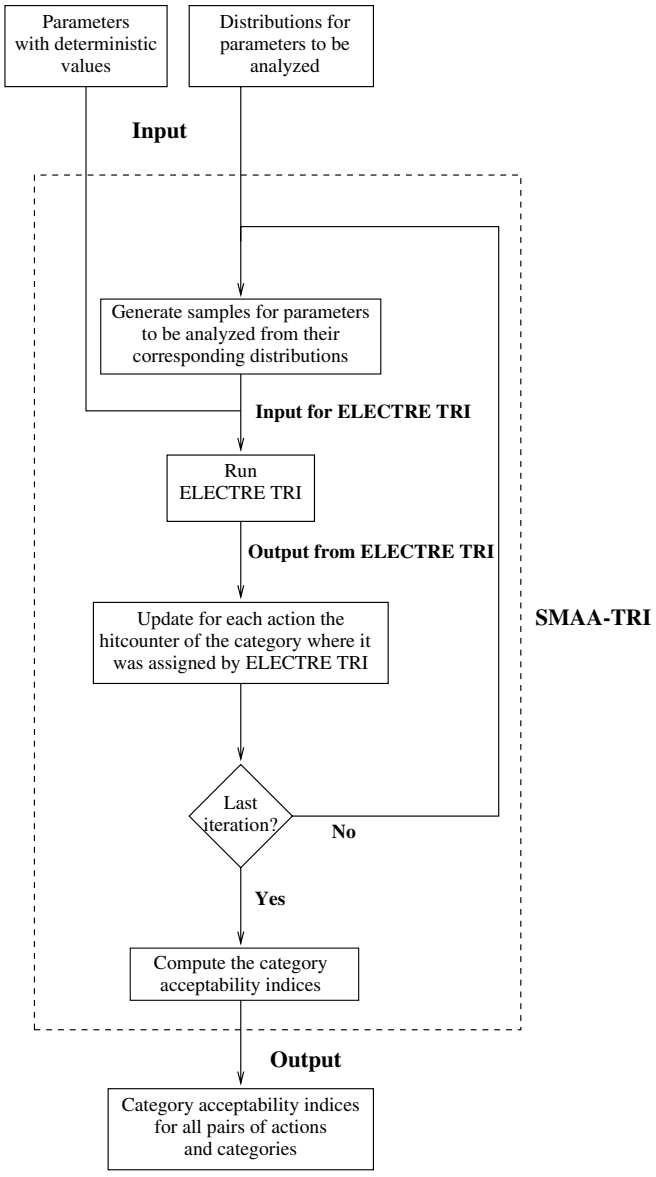


Fig. 4. The SMAA-TRI simulation scheme.

4. Case study: Experiments and results

In this case study we re-analyze a recent real-world application of ELECTRE TRI in the field of risk analysis. The original analysis is presented in Merad et al. (2004). The study concentrates on France's Lorraine region, where iron has been mined for more than a century. The underground mining tunnels have caused land subsidence, which has caused buildings to collapse. The object of this study was to partition the land into zones and assign these zones into predefined risk categories in order to decide which zones need constant surveillance. We will re-analyze the assignment procedure by using the data provided in the case study.

The assignment phase consists of 10 homogenous zones (alternatives), $a_1,\ldots,a_{10}$, that are evaluated in terms of 10 criteria, $g_1,\ldots,g_{10}$. There are 4 risk categories into which the zones are assigned, Category 1 is for zones with highest risk and Category 4 for lowest. The risk categories are separated by three profiles b_1, b_2, and b_3. We will omit the numerical values for alternative performances, profiles, and thresholds here for brevity. The interested reader should refer to the original study (Merad et al., 2004) for these values.

The authors used the Revised Simos' procedure by Figueira and Roy (2002) to elicit the criteria weights. These weights are presented in Table 1. This case study uses the pessimistic assignment rule.

The authors of the original case study used lambda cutting level of 0.65, but also analyzed the sensitivity of the results by altering the lambda to 0.7, 0.75, 0.8, and 0.85. The results including the sensitivity analysis are presented in Table 2.

Table 1
Weights of the case study

Weight	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
Non-normalized	5	1	1	1	5	1	1	20	1	10

We made a stability analysis to this case study using SMAA-TRI with deterministic profile values for the sake of consistency with the sensitivity analysis of the original study. We chose cutting level to be represented by a stochastic variable uniformly distributed in the range $[0.65, 0.85]$. The feasible weight space was defined with constraints provided in Table 3. These constraints are not probably the best constraints possible, as quantifying the imprecision should have been done along with the original case study.

SMAA-TRI was executed with 10,000 Monte Carlo iterations. The resulting category acceptability indices are presented in Table 4. Visualization of the results is important in SMAA methods, especially if there is a large amount of alternatives and/or criteria. Because the categories are ordered and therefore upwards inclusive, we have chosen to visualize them with stacked columns in Fig. 5. This allows the DM to see the amount of high risk in the assignment. For example, the acceptability of the assignment of a certain alternative to category 1 or 2.

The results of the re-analysis show the usefulness of SMAA-TRI. Although the stability analysis results are quite different from the ones by Merad et al. (2004), SMAA-TRI provides more information than the original analysis. This can be seen, for example, by comparing the sensitivity analysis results for Zone 5 in Table 2 and the category acceptability indices for the same zone in Table 4. The original sensitivity analysis gives information that Zone 5 can be assigned to risk categories 3 or 4, and with this information the DMs (especially if they are risk-aware) should treat the zone as it would be assigned to risk category 3, which is of higher risk than category 4. But with the information provided by the category acceptability indices a more informed decision can be made. By regarding our uncertain information about the parameters, we can quite safely (98% acceptability) place the zone in risk category 4. In a real-life case, this possibly induces savings as less safety measures (that are usually quite expensive) can be applied.

Table 2
Original results of the case study and sensitivity analysis

Zone	Result	Sensitivity analysis
a_1	Category 1	Categories 1 and 2
a_2	Category 1	Categories 1 and 2
a_3	Category 2	Stable
a_4	Category 2	Stable
a_5	Category 4	Categories 3 and 4
a_6	Category 1	Categories 1 and 2
a_7	Category 3	Categories 3 and 4
a_8	Category 4	Categories 3 and 4
a_9	Category 1	Categories 1 and 2
a_{10}	Category 4	Stable

Table 3
Weight constraints for the re-analysis

Weight	Lower bound	Upper bound
w_1	3	7
w_2	0	2
w_3	0	2
w_4	0	2
w_5	3	7
w_6	0	2
w_7	0	2
w_8	15	25
w_9	0	2
w_{10}	7	13

Table 4
Category acceptability indices

Zone	Category 1	Category 2	Category 3	Category 4
a_1	100	0	0	0
a_2	100	0	0	0
a_3	0	100	0	0
a_4	0	100	0	0
a_5	0	0	2	98
a_6	100	0	0	0
a_7	0	34	54	12
a_8	0	34	10	56
a_9	100	0	0	0
a_{10}	0	34	21	45

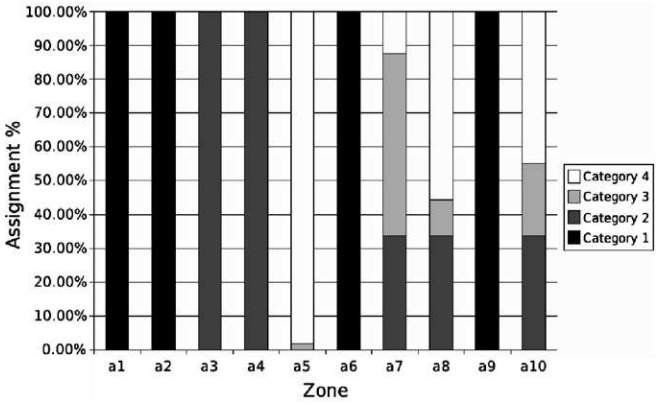


Fig. 5. The category acceptability indices.

In this re-analysis using uncertain weights provides some interesting results. The original sensitivity analysis considered the assignment of Zone 10 stable, but by considering the imprecision on the weights, the assignment of the zone is quite unstable. With only 45% of the feasible parameter values the zone is placed in risk category 4, and a quite large share of the feasible values (34%) places the zone in risk category 2. If the original case study would have been performed with uncertain weight values, the assignments of the alternatives would have probably been quite different.

5. Conclusions

In this paper, we presented the SMAA-TRI method that allows ELECTRE TRI to be applied with stochastic values for profiles, lambda cutting level, and weights. The SMAA-TRI analysis results in category acceptability indices for all pairs of alternatives and categories, and these can be used to analyze the stability of the parameters. The indices can be used also for deriving robust conclusions, or if not pos-

Please cite this article in press as: Tervonen, T. et al., A stochastic method for robustness analysis in sorting problems, European Journal of Operational Research (2007), doi:10.1016/j.ejor.2007.09.008

sible, to quantify the "amount of instability" in the results induced by the uncertain parameter values.

Defining parameter values for ELECTRE TRI model is not an easy task. Moreover, if there are multiple DMs with conflicting preferences, it might even be impossible to reach consensus about weight values. With our approach the possibility to define the model by using stochastic variables overcomes these difficulties: the lambda cutting level and the profiles can be defined with uncertain values, and the weights can be defined as intervals containing the preferences of all DMs. The method can be extended to include uncertain values for thresholds as well, but it is not certain that it would bring additional help to the decision making process.

We presented a re-analysis of the case study in which the usefulness of SMAA-TRI was shown. We compared the original robustness analysis and our re-analysis, and found that the added information about the sensitivity of the parameters allows more informed decisions to be made.

Acknowledgements

The work of Tommi Tervonen was supported by Grants from Turun Yliopistosäätiö and the Finnish Cultural Foundation. The work of Juscelino Almeida Dias was supported by the Grant SFRH/BM/18781/2004 (Fundação para a Ciência e Tecnologia, Portugal).

References

Abbas, A., 2006. Maximum entropy utility. Operations Research 54 (2), 277–290.

Barron, F.H., Barret, B.E., 1996. Decision quality using ranked attribute weights. Management Science 42, 1515–1523.

Butler, J., Dia, J., Dyer, J., 1997. Simulation techniques for the sensitivity analysis of multi-criteria decision models. European Journal of Operational Research 103 (3), 531–545.

Dias, L., Clímaco, J., 1999. On computing ELECTRE's credibility indices under partial information. Journal of Multi-Criteria Decision Analysis 8 (2), 74–92.

Dias, L., Clímaco, J., 2000. ELECTRE TRI for groups with imprecise information on parameter values. Group Decision and Negotiation 9 (5), 355–377.

Dias, L., Mousseau, V., 2006. Inferring Electre's veto-related parameters from outranking examples. European Journal of Operational Research 170, 172–191.

Dias, L., Mousseau, V., Figueira, J., Clímaco, J., 2002. An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI. European Journal of Operational Research 138, 332–348.

Doumpos, M., Zopounidis, C., 2002. In: Multicriteria Decision Aid Classification Methods, vol. 73. Kluwer Academic Publishers, Dordrecht.

Figueira, J., Mousseau, V., Roy, B., 2005. ELECTRE methods. In: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science+Business Media Inc. (Chapter 4).

Figueira, J., Roy, B., 2002. Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. European Journal of Operational Research 139, 317–326.

Hokkanen, J., Salminen, P., 1997. Choosing a solid waste management system using multicriteria decision analysis. European Journal of Operational Research 98, 19–36.

Jessop, A., 1999. Entropy in multiattribute problems. Journal of Multi-Criteria Decision Analysis 8, 61–70.

Lahdelma, R., Hokkanen, J., Salminen, P., 1998. SMAA – stochastic multiobjective acceptability analysis. European Journal of Operational Research 106, 137–143.

Lahdelma, R., Salminen, P., 2001. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Operations Research 49 (3), 444–454.

Lahdelma, R., Salminen, P., 2002. Pseudo-criteria versus linear utility function in stochastic multicriteria acceptability analysis. European Journal of Operational Research 141 (2), 454–469.

Merad, M., Verdel, T., Roy, B., Kouniali, S., 2004. Use of multi-criteria decision-aids for risk zoning and management of large area subjected to mining-indiced hazards. Tunnelling and Underground Space Technology 19, 125–138.

Moscarola, J., Roy, B., 1977. Procédure automatique d'éxamen de dossiers fondée sur une ségmentation trichotomique en présence de critères multiples. RAIRO Recherche Opérationnelle 11 (2), 145–173.

Mousseau, V., 1995. Eliciting information concerning the relative importance of criteria. In: Pardalos, Y., Siskos, C., Zopounidis, C. (Eds.), Advances in Multicriteria Analysis. Kluwer Academic Publishers, Dordrecht, pp. 17–43.

Mousseau, V., Dias, L., Figueira, J., 2004. Dealing with inconsistent judgements in multiple criteria sorting models. Research Report 17/2004 of The Institute of Systems Engineering and Computers (INESC-Coimbra), Coimbra, Portugal, <http://www.inescc.pt>.

Mousseau, V., Figueira, J., Dias, L., Gomes da Silva, C., Clímaco, J., 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. European Journal of Operational Research 147, 72–93.

Mousseau, V., Figueira, J., Naux, J., 2001. Using assignment examples to infer weights for ELECTRE TRI method: Some experimental results. European Journal of Operational Research 130 (2), 263–275.

Mousseau, V., Słowiński, R., 1998. Inferring an ELECTRE TRI model from assignment examples. Journal of Global Optimization 12 (2), 157–174.

Mousseau, V., Słowiński, R., Zielniewicz, P., 2000. A user-oriented implementation of the ELECTRE-TRI method integrating preference elicitation support. Computers & Operations Research 27 (7–8), 757–777.

Ngo The, A., Mousseau, V., 2002. Using assignment examples to infer category limits for the ELECTRE TRI method. Journal of Multi-Criteria Decision Analysis 11, 29–43.

Rao, J.S., Sobel, M., 1980. Incomplete Dirichlet integrals with applications to ordered uniform spacings. Journal of Multivariate Analysis 10, 603–610.

Rogers, M., Bruen, M., 1998. A new system for weighting environmental criteria for use within ELECTRE III. European Journal of Operational Research 107, 552–563.

Roy, B., 2002. Robustesse de quoi et *vis-à-vis* de quoi mais aussi robustesse pourquoi en aide à la décision? Newsletter of the European Working Group on Multiple Criteria Decision Aiding 6 (3).

Roy, B., 2005. A propos de robustesse en Recherche Opérationelle et Aide à la Décision. In: J. Billaut, A. Moukrim and E. Sanlaville (Eds.), Flexibilité et Robustesse en Ordonnancement, Hermes Science Publications, (Chapter 2).

Tervonen, T., Figueira, J., Lahdelma, R., Salminen, P., 2007. Towards robust ELECTRE III with simulation: theory and software of SMAA-III. Research report 1/2007 of CEG-IST, Technical University of Lisbon, Portugal.

Tervonen, T., Figueira, J.R., in press. A survey on stochastic multicriteria acceptability analysis methods. Journal of Multi-Criteria Decision Analysis.

Tervonen, T., Lahdelma, R., 2007. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research 178 (2), 500–513.

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.

Please cite this article in press as: Tervonen, T. et al., A stochastic method for robustness analysis in sorting problems, European Journal of Operational Research (2007), doi:10.1016/j.ejor.2007.09.008

Publication IV

Tervonen T., Figueira J. R., Lahdelma R., Salminen P., 2007. Towards robust ELECTRE III with simulation: theory and software of SMAA-III. Working paper 1/2007 of CEG-IST. Submitted to European Journal of Operational Research.

Towards robust ELECTRE III with simulation: theory and software of SMAA-III

Tommi Tervonen [a,b,*], José Rui Figueira [a], Risto Lahdelma [b], Pekka Salminen [c]

[a] *CEG-IST, Centre for Management Studies, Instituto Superior Técnico, Technical University of Lisbon, 2780-990 Porto Salvo, Portugal*

[b] *Department of Information Technology, University of Turku, FIN-20520 Turku, Finland*

[c] *School of Business and Economics, University of Jyväskylä, P.O. Box 35, FIN-40014 Jyväskylä, Finland*

Abstract

ELECTRE III is a well-established multiple criteria decision making method with a solid track of real-world applications. It requires precise values to be specified for the parameters and criteria measurements, which in some cases might not be available. In this paper we present a method, SMAA-III, that allows ELECTRE III to be applied with imprecise parameter values. By allowing imprecise values, the method also allows an easily applicable robustness analysis. In SMAA-III, simulation is used and descriptive measures are computed to characterize stability of the results. We present a software implementing the method and show the usage by re-analyzing an existing case study.

Key words: Decision support systems, Robustness and sensitivity analysis, Simulation, Multiple criteria analysis

1 Introduction

ELECTRE III is a well-established multiple criteria decision making (MCDM) method for ranking a discrete set of alternatives. It belongs to the ELECTRE family of methods that are based on constructing and exploiting an outranking relation

* Corresponding author.
Email addresses: tommi.tervonen@it.utu.fi (Tommi Tervonen), figueira@ist.utl.pt (José Rui Figueira), risto.lahdelma@cs.utu.fi (Risto Lahdelma), psalmine@econ.jyu.fi (Pekka Salminen).

(see Figueira et al., 2005). ELECTRE III has a long history of successful real-world applications in different areas. The inputs for ELECTRE III consist of criteria evaluations on a set of alternatives and preference information expressed as weights and thresholds.

ELECTRE III is a pseudo-criterion based model, and as such it uses a threshold to model indifference between pairs of alternatives. Although this threshold might be an easy concept for a common Decision Maker (DM) to understand, simulation studies have shown, that it causes the model be quite unstable with respect to changes in the indifference threshold value (Lahdelma and Salminen, 2002). Because of this unstability, robustness analysis should always be done by considering different values for the threshold.

Real-world decision-making problems in general include various types of uncertainties inherent in problem structuring and analysis (Belton and Stewart, 2002). Eliciting the DMs preferences in terms of relative criteria importance coefficients or weights is usually difficult. Such weights should always be considered imprecise, because humans usually do not think about preferences as exact numerical values, but as more vague concepts (Smets, 1991). In some cases, weight information may be entirely missing, which corresponds to extremely imprecise weights.

This work presents a tool for dealing with imperfect knowledge within the ELECTRE III method. It can be used either when information is poor and/or when a robustness analysis needs to be done. The way robustness analysis is conducted comprises intensity of exploration in the parameters space. This is achieved by applying simulation in such a way that the parameter space is explored with a high concentration of discrete values. In addition to this, the exploration is coherent with the model. This means that, for example, when exploring the weight space, the meaning of weight is taken into account. In ELECTRE III weights represent the amount of “votes” criteria have.

Capability to derive robust conclusions when applying MCDM methods is nowadays of uttermost importance. The main sources of imperfect knowledge that are present in complex and multifaceted decision-making situations require a careful observation of the results, and make them dependent of an exploration of the neighbourhood of the parameters used mainly to represent preferences or technical aspects of the problem. If an alternative occupies almost always the first position when changing simultaneously all the parameters in a certain neighbourhood, it means that it can be a good choice for future implementation; these are the kind of robust conclusions we are interested in.

The method presented in this paper is based on Stochastic Multicriteria Acceptability Analysis (SMAA) (Lahdelma and Salminen, 2001), that is a family of decision support methods for aiding DMs in discrete decision-making problems. For a survey on SMAA methods, see Tervonen and Figueira (2006). The proposed

method, SMAA-III, explores weight, criteria measurement, and threshold spaces, in order to describe which values result in certain ranks for the alternatives. It allows ELECTRE III to be used with different kinds of imprecise or partially missing information. This brings numerous advantages. Firstly, SMAA-III allows performing an initial analysis without preference information in order to eliminate "inferior" alternatives. Secondly, it allows DMs to express their preferences imprecisely, which can lower the DMs' cognitive effort compared to specifying precise weights. Thirdly, imprecise criteria measurements can be represented with arbitrary joint probability distributions, allowing to model imprecision in a coherent way not possible with ELECTRE III. Fourthly, it allows representing the preferences of a group of DMs. Fifthly, the method can be used for analyzing the robustness of the results by representing the imprecision of the elicited weights as constraints or as suitable probability distributions.

In SMAA-III, robustness is analyzed with respect to the weights, criteria measurements, and thresholds. Traditionally, robustness with ELECTRE methods is analyzed by considering discrete points in the weight space (see e.g. Rogers et al., 2000). But in the case of ELECTRE III this is not enough: weights between these points that might give contradictory results are missed. There are simulation techniques for robustness analysis also outside the SMAA methodology (see Butler et al., 1997), but to our best knowledge, they have never before been applied to ELECTRE III.

User-friendly software is of crucial importance for a MCDM method to be of pratical importance. We present the software for SMAA-III, and demonstrate its use by re-analyzing a real-life case study. The software presented is available for all major operating systems.

This paper is organized as follows: ELECTRE III is briefly introduced in Section 2. SMAA-III is presented in Section 3. The software and a re-analysis of a case-study are presented in Section 4. The paper ends with conclusions and avenues for future research in Section 5.

2 ELECTRE III

ELECTRE III is designed for solving a discrete ranking problem. It consists of m alternatives $a_1, \ldots, a_i, \ldots, a_m$, that are evaluated in terms of n criteria $g_1, \ldots, g_j, \ldots, g_n$. We denote by J the set of criterion indices. $g_j(a_i)$ is the evaluation of criterion g_j for alternative a_i. Without loss of generality, we assume that all criteria are to be maximized.

Similarly to the other ELECTRE family methods, ELECTRE III is based on two phases. In the first phase, an outranking relation between pairs of alternatives is

formed. The second phase consists of exploiting this relation, producing a final partial pre-order and a median pre-order. S denotes the *outranking* relation, that is, aSb denotes that "alternative a is at least as good as alternative b".

ELECTRE III applies pseudo-criteria in constructing the outranking relation. A pseudo-criterion is defined with two thresholds for modelling preferability: an indifference threshold $q_j(g_j(\cdot))$ for defining the difference in criterion g_j that the DM deems insignificant, and a preference threshold $p_j(g_j(\cdot))$ for the smallest difference that is considered absolutely preferred. Between these two is a zone of "hesitation" between indifference and strict preference. ELECTRE III also defines third threshold, a veto threshold $v_j(g_j(\cdot))$. It is the smallest (negative) difference that completely nullifies (raises "veto" against) the outranking relation. In addition to the thresholds, preferences are quantified through a weight vector $w = (w_1, \ldots, w_j, \ldots, w_n)$. Without loss of generality, we assume that $\sum j \in J w_j = 1$.

Exploitation of the outranking relation produces a partial pre-order, in which every pair of alternatives is connected with either indifference (I), incomparability (R), or preference ($\succ$) relation.

2.1 Constructing the outranking relation

The outranking relation between every pair of alternatives is constructed based on a comprehensive concordance index and partial discordance indices. The concordance index is computed by considering individually for each criterion g_j the support it provides for the assertion aS_jb "a outranks b with respect to criterion g_j". The partial concordance index is computed as follows, for all $j \in J$:

$$c_j(a,b) = \begin{cases} 1, \text{ if } g_j(b) - g_j(a) \leq q_j(g_j(a)), \\ \frac{g_j(a) + p_j\left(g_j(a)\right) - g_j(b)}{p_j\left(g_j(a)\right) - q_j\left(g_j(a)\right)}, \text{ if } q_j(g_j(a)) < g_j(b) - g_j(a) \leq p_j(g_j(a)), \\ 0, \text{ if } g_j(b) - g_j(a) > p_j(g_j(a)). \end{cases} \quad (1)$$

After computing the partial concordance indices, the comprehensive concordance index is computed as follows,

$$c(a,b) = \sum_{j \in J} w_j c_j(a,b). \quad (2)$$

The discordance of criterion g_j describes the veto effect this criterion imposes against the assertion aSb. The partial discordance indices are computed separately

for each criterion $j \in J$:

$$d_j(a,b) = \begin{cases} 1, \text{ if } g_j(b) - g_j(a) \geq v_j(g_j(a)), \\ \frac{g_j(b)-g_j(a)-p_j(g_j(a))}{v_j(g_j(a))-p_j(g_j(a))}, \text{ if } p_j(g_j(a)) \leq g_j(b) - g_j(a) < v_j(g_j(a)), \\ 0, \text{ if } g_j(b) - g_j(a) < p_j(g_j(a)). \end{cases} \quad (3)$$

By applying the pre-mentioned indices, the degree of credibility of the outranking assertion *aSb* is defined as

$$\rho(a,b) = \begin{cases} c(a,b) \prod_{j \in V} \frac{1-d_j(a,b)}{1-c(a,b)}, \text{ if } V \neq \emptyset, \\ c(a,b), \text{ otherwise,} \end{cases} \quad (4)$$

with

$$V = \{ j \in J : d_j(a,b) > c(a,b) \}. \quad (5)$$

Notice that when $d_j(a,b) = 1$ for any $j \in J$, this implies that $\rho(a,b) = 0$.

2.2 The exploitation procedure

The exploitation of the outranking relation consists of two phases. In the first phase, two complete pre-orders, Z_1 (descending) and Z_2 (ascending) are constructed with the so-called distillation procedures. In the second phase, a final partial pre-order or a complete median pre-order is computed based on these two pre-orders.

The distillation procedures work by iteratively cutting the fuzzy outranking relations with descending λ-cutting levels. With a given cutting level λ_*, alternative a outranks alternative b ($aS^{\lambda_*}b$) if the following holds:

$$aS^{\lambda_*}b \iff \begin{cases} \rho(a,b) > \lambda_*, \text{ and} \\ \rho(a,b) > \rho(b,a) + s(\rho(a,b)), \end{cases} \quad (6)$$

where $s(\cdot)$ is the distillation threshold, usually defined as (Belton and Stewart, 2002)

$$s(x) = 0.3 - 0.15x. \quad (7)$$

The pre-orders are constructed in an iterative manner. In each step the alternatives with the highest or lowest qualification scores are distillated, depending on whether the distillation is descending or ascending. The qualification score is computed as a difference between the number of alternatives that the selected alternative outranks and the number of alternatives that outrank it for a given cutting level. The

procedure is presented in Algorithm 1.

Algorithm 1 Distillations

1: Determine the maximum value of the credibility indices in the set under consideration. Assign this to λ.
2: Determine $\lambda_* = \max_{d(a,b)<\lambda-s(\lambda)} \{d(a,b)\}$, where (a,b) belong to the set under consideration.
3: If $\lambda_* = 0$, end this distillation.
4: Determine for each alternative its *qualification* score, that is: the difference between the number of alternatives it outranks and the number of alternatives that outrank it. Outranking is determined according to λ_*.
5: The set of alternatives having the largest (or smallest, if the distillation is ascending) qualification is the current distillate.
6: If the number of alternatives in current distillate is larger than 1, repeat the process from step 2 inside the distillate.
7: Form a new set under consideration by removing the distillated alternatives from the current one. If this set is not empty, repeat the process on the new set from step 1.
8: The final pre-orders are ranked so that the alternatives in the first distillate are given rank 1, in the second rank 2, etc.

In the original ELECTRE III, a median pre-order is computed based on the two complete pre-orders, Z_1 and Z_2, and the final partial pre-order. The final partial pre-order is computed as the intersection of the two complete pre-orders in such a way that the following relations hold:

$$\begin{aligned} a \succ b &\Longleftrightarrow (a \succ^{Z_1} b \wedge a \succ^{Z_2} b) \vee (aI^{Z_1}b \wedge a \succ^{Z_2} b) \vee (a \succ^{Z_1} b \wedge aI^{Z_2}b), \\ aIb &\Longleftrightarrow (aI^{Z_1}b \wedge aI^{Z_2}b), \\ aRb &\Longleftrightarrow (a \succ^{Z_1} b \wedge b \succ^{Z_2} a) \vee (b \succ^{Z_1} a \wedge a \succ^{Z_2} b). \end{aligned} \tag{8}$$

After this, the median pre-order can be computed by removing the incomparabilities with calculating the differences of ranks of an alternative in the two complete pre-orders.

2.3 Robustness analysis for weights

There are numerous weight elicitation techniques proposed for ELECTRE methods, the following being among the most recent and popular ones:

(1) DIVAPIME by Mousseau (1995) produces intervals for weights.
(2) Hokkanen and Salminen (1997) used two different weight elicitation procedures, and found that the normalized sets of weights had minor differences.
(3) SRF by Figueira and Roy (2002) allows weight elicitation in a user-friendly

manner by using a technique based on a pack of "playing cards" to determine the relative importance criteria coefficients. It can produce interval weights and was also designed to support multiple DMs.

(4) The approach proposed by Rogers and Bruen (1998) uses pairwise comparisons to elicit the weights.

The first three techniques that produce intervals or two weight sets that may be used to define intervals, can directly be used in robustness analysis. When using the fourth weight elicitation technique, intervals (such as $\pm 10\%$) could be defined around the original weights.

Traditionally the robustness analysis for ELECTRE methods has been an *ad hoc* investigation into the effect of changing values (Belton and Stewart, 2002). This type of investigation typically considers only discrete points (for example, extreme points) of the feasible weight space (e.g. weight intervals). The procedure of building the pre-orders is based on exploiting the fuzzy outranking relation, which is non-linear and discontinuous by nature. Therefore, instead of just a few discrete points, it is important to analyze the entire continuum of the weight space.

3 SMAA-III

In order to overcome the limitations of ELECTRE III, SMAA-III applies simulation and studies the effect of changing parameter values and criterion evaluations on the results. The imprecision is quantified through joint density functions in the corresponding spaces.

The weights are represented by a weight distribution with joint density function $f_W(w)$ in the feasible weight space W. The weights are non-negative and normalized: the weight space is an $n-1$ dimensional simplex,

$$W = \left\{ w \in \mathbb{R}^n : w \geq 0 \text{ and } \sum_{j \in J} w_j = 1 \right\}. \tag{9}$$

Completely missing preference information is represented by a uniform (constant) weight distribution in W, that is,

$$f_W(w) = 1/\text{vol}(W). \tag{10}$$

If some kind of preference information is available, different weight distributions can be applied (see Lahdelma and Salminen, 2001). In practice, the preferences can usually be elicited as interval constraints for weights. In this case, a uniform distribution in the space bounded by the constraints is used. Figure 1 illustrates the restricted feasible weight space of a 3-criterion problem with lower and up-

per bounds for w_1. In this paper the focus is on weight information provided as intervals, because:

(1) if there are multiple DMs whose preferences need to be taken into account, the weight intervals in general can be determined to contain the preferences of all DMs (see Lahdelma and Salminen, 2001), and
(2) weight intervals allow simple robustness analysis also when only deterministic weights are available, by specifying, for example, a $\pm 10\%$ interval for each weight.

It should be observed that other forms of easily elicitable preference information can be used as well, such as ranking of the criteria. A ranking can be obtained by asking the DMs to identify their most important, second most important, etc. criterion. Figure 2 illustrates the feasible weight space for a three-criterion problem with the ranking $w_1 \geq w_2 \geq w_3$.

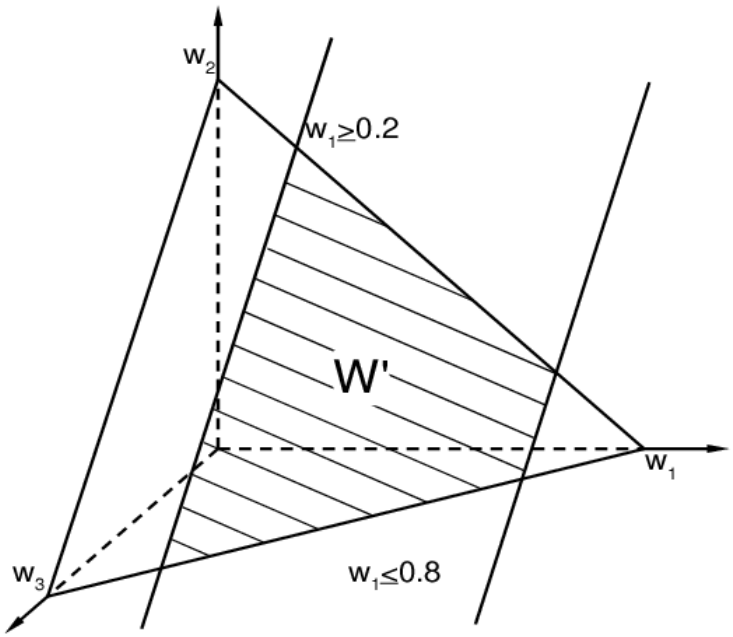


Figure 1. Feasible weight space of a 3-criterion problem with lower and upper bounds for w_1.

Imprecise thresholds are represented by stochastic functions $\alpha_j(\cdot)$, $\beta_j(\cdot)$, and $\gamma_j(\cdot)$, corresponding to the deterministic thresholds $p_j(\cdot)$, $q_j(\cdot)$, and $v_j(\cdot)$, respectively. To simplify the notation, we define a 3-tuple of thresholds $\tau = (\alpha, \beta, \gamma)$. It has a joint density function f_T in the space of possible values defining the functions. It should be noted that all feasible combinations of thresholds must satisfy $q_j(a_i) < p_j(a_i) < v_j(a_i)$.

Traditionally the thresholds in ELECTRE models have been used to model preferences of the DMs (e.g. differences deemed significant) as well as imprecision

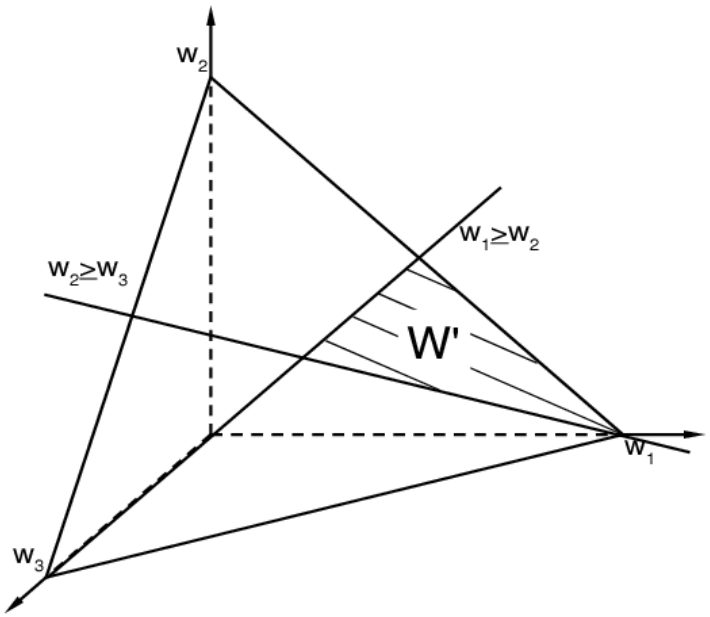


Figure 2. Feasible weight space of a 3-criterion problem with ranking of the criteria.

in the data. But it has been shown that the indifference threshold does not correspond to a linear imprecision interval (Lahdelma and Salminen, 2002). Therefore, in SMAA-III thresholds are used only to model preferences (together with weights). Imprecision in the criteria measurements is modelled with stochastic variables.

These stochastic variables are denoted with ξ_{ij} corresponding to the deterministic evaluations $g_j(a_i)$. They have a density function $f_X(\xi)$ defined in the space $X \subseteq R^{m \times n}$. In principle, arbitrary distributions can be used, but in practice a uniform distribution in a certain interval or a Gaussian distribution is used.

Incomparabilities between alternatives can be present in the final results of ELECTRE III. This is one of the main features of ELECTRE methods in comparison with the methods applying classical multi-attribute utility theory (see Keeney and Raiffa, 1976). In the late seventies, it was considered a very important theoretical advance. But, in reality when dealing with practical situations, incomparabilities in the final result are inconvenient. This aspect was soon observed (Roy et al., 1986) and partial pre-orders were replaced by complete pre-orders or median pre-orders. This is quite logical, as why would DMs want to apply a tool that tells the alternatives are incomparable: this was already the starting position. We apply median pre-orders in computing rank acceptability indices. The only information lost in using the median pre-order as the primary measure of the ranking is the incomparability. As our method is also aimed to help analysts accustomed to ELECTRE III, we will later present another index to measure incomparability.

Since only the median pre-order is needed in SMAA-III, the intermediate step of constructing the final partial pre-order can be skipped; the median pre-order is con-

structed directly from the two complete pre-orders in the following way:

$$a \succ b \Longleftrightarrow \begin{cases} (a \succ^{Z_1} b \wedge a \succ^{Z_2} b) \vee (aI^{Z_1}b \wedge a \succ^{Z_2} b) \vee (a \succ^{Z_1} b \wedge aI^{Z_2}b) \\ (a \succ^{Z_1} b \wedge b \succ^{Z_2} a) \wedge (|r^{Z_1}(a) - r^{Z_2}(a)| < |r^{Z_1}(b) - r^{Z_2}(b)|) \\ (b \succ^{Z_1} a \wedge a \succ^{Z_2} b) \wedge (|r^{Z_1}(a) - r^{Z_2}(a)| < |r^{Z_1}(b) - r^{Z_2}(b)|) \end{cases},$$
$$aIb \Longleftrightarrow \neg(b \succ a) \wedge \neg(a \succ b),$$

where $r(\cdot)$ is the ranking of an alternative in the superscripted pre-order.

Monte Carlo simulation is used in SMAA-III to compute three types of descriptive measures: rank acceptability indices, pair-wise winning indices, and incomparability indices. In order to compute these indices, let us define a *ranking function* that evaluates the rank r of the alternative a_i with the corresponding parameter values:

$$rank(i, w, \xi, \tau). \tag{11}$$

The evaluation of this function corresponds to executing ELECTRE III and returning rank of the corresponding alternative in the resulting median pre-order. We will next introduce the three indices. Interpretation of their values is presented in Section 4 through various re-analyses.

3.1 Rank acceptability index

The rank acceptability index, b_i^r, measures the share of feasible weights that grant alternative a_i rank r in the median pre-order by taking into account simultaneously imprecisions in all parameters and criterion evaluations. It represents the share of all feasible parameter combinations that make the alternative acceptable for a particular rank, and it is most conveniently expressed percentage-wise.

The rank acceptability index b_i^r is computed numerically as a multidimensional integral over the spaces of feasible parameter values as

$$b_i^r = \int_{W:rank(i,w,\xi,\tau)=r} f_W(w) \int_X f_X(\xi) \int_T f_T(\tau)\, dT\, dw\, d\xi. \tag{12}$$

The most acceptable ("best") alternatives are those with high acceptabilities for the best ranks. Evidently, the rank acceptability indices are within the range [0,1], where 0 indicates that the alternative will never obtain a given rank and 1 indicates that it will obtain always the given rank with any feasible choice of parameters. Thus, the rank acceptability indices are a measure of robustness.

Using the rank acceptability indices as measures of robustness is quite straightforward. More caution should be put on interpreting the results when these indices are

computed without weight information to characterize the problem. If an alternative obtains a low score for the first rank acceptability, it does not necessarily mean that it is "inferior". The DMs' actual preferences may well lie within the corresponding (small) set of favourable first rank weights.

3.2 Pair-wise winning index

The pair-wise winning index (Leskinen et al., 2006), o_{ik}, describes the share of weights that place alternative a_i on a better rank than alternative a_k. An alternative a_i that has $o_{ik} = 1$ for some k always obtains a better rank than alternative a_k, and can thus be said to *dominate* it.

The pair-wise winning index o_{ik} is computed numerically as a multidimensional integral over the space of weights that give alternative a lower rank than for another.

$$o_{ik} = \int_{w \in W : rank(i,w,\xi,\tau) < rank(k,w,\xi,\tau)} f_W(w) \int_X f_X(\xi) \int_T f_T(\tau)\, dT\, dw\, d\xi. \qquad (13)$$

The pair-wise winning indices are especially useful when trying to distinguish between the ranking differences of two alternatives. Because the number of ranks in the median pre-order of different simulation runs varies, two alternatives might obtain similar rank acceptabilities although one is in fact inferior. In these cases looking at the pair-wise winning indices between this pair of alternatives can help to determine whether one of the alternatives is superior to the other or if they are equal in "goodness".

3.3 Incomparability index

Because median pre-orders are used in computing the rank acceptability indices, it is not anymore possible to model incomparability. As some DMs might be accustomed to make decisions also based on incomparabilities, another index is introduced. Incomparability index ρ_{ik} measures the share of feasible parameter values that cause alternatives a_i and a_k to be incomparable. For this reason, we define the incomparability function:

$$R(i,k,\xi,\tau) = \begin{cases} 1, \text{ if alternatives } a_i \text{ and } a_k \text{ are judged incomparable,} \\ 0, \text{ if not.} \end{cases} \qquad (14)$$

This function corresponds to a run of ELECTRE III with the given parameter values and checking if the alternatives are judged incomparable in the final partial

pre-order. In practice we do not compute the final partial pre-order, because this information can be extracted from the two partial pre-orders Z_1 and Z_2 as shown in (8). By using the incomparability function, the incomparability index is computed numerically as a multidimensional integral over the feasible parameter spaces as

$$\rho_{ik} = \int_W f_W(w) \int_X f_X(\xi) \int_T f_T(\tau) R(i,j,\xi,\tau)\, dT\, dw\, d\xi. \tag{15}$$

3.4 Computation

All of the indices mentioned above are computed with Monte Carlo simulation. The procedure is similar to that presented and analyzed by Tervonen and Lahdelma (2007). SMAA-III differs in the sense that it applies the ELECTRE III procedure to derive the descriptive values instead of a utility function.

In each simulation iteration, sample parameter values are generated from their corresponding distributions, and ELECTRE III is executed with these values. Then the corresponding hit-counters are updated as with the original SMAA. If standard distributions are used for defining the imprecise parameter values, then all sampling except weight generation are computationally very light. In the case of weight generation, if tight upper bounds are used, the we can have very high weight rejection ratios (up to 99,9%). Nevertheless, even with 99,9% weight rejection, the method is fast enough to use in an interactive decision making process with problems of reasonable size.

For obtaining sufficient accuracy for the indices, we suggest using at least 10000 simulation iterations. This gives error limits of less than 0.01 with 95% confidence (Tervonen and Lahdelma, 2007).

4 Case study and software

ELECTRE III has been used to choose the best waste incineration strategy for the Eastern Switzerland region (Rogers et al., 2000). A total of 11 alternative strategies (alternatives) were evaluated in terms of 11 criteria. ELECTRE III was run separately for 6 interest groups, each of which had different preferences. The complete study will not be presented here. The interested reader should refer to Rogers et al. (2000), Section 6.

We re-analyze the study using the SMAA-III software. The software is programmed in C++ using portable user interface libraries, and is therefore available for various operating systems. Currently Linux, Macintosh OS X, and Windows XP are supported. The software allows easy input of all models data. The interface is designed

in such a way that from each input phase, the user can move to any other one. For examples of the interface, see tables for criterion type input in Figure 3 and criterion measurement input in Figure 4.

Figure 3. Input of criteria types in the software.

Figure 4. Input of criteria measurements in the software.

We re-analyze the “baseline run” of the case-study with weight information from

the Switzerlands Federal Agency for the Environment. In this initial run, the veto thresholds were not defined. The weights were elicited by using two methods, the SRF method and the method by Rogers and Bruen (1998). According to Rogers et al. (2000), the differences in the weights obtained by using the two methods were minor. To see how small differences in the parameter values cause alteration of results, we re-analyze the problem with with five different scenarios:

(1) Original problem. Re-analysis by using the median pre-order.
(2) Imprecise weights. Original problem with imprecise weight values.
(3) Imprecise thresholds. Original problem with imprecise threshold values.
(4) Imprecise criteria measurements. Original problem with imprecise measurements for the cardinal criteria.
(5) Imprecision in all values. Original problem incorporating simultaneously all the above types of imprecision.

The goal of our re-analysis is to identify which parameters are the most sensitive, and what kind of robust conclusions can be derived from the results. We also show how the different index values should be interpreted. All these analyses are computed with 10000 Monte Carlo iterations.

4.1 Original study with median pre-order

We ran the software with exact data from the original study, obtaining a median pre-order. Rank acceptability indices are shown in Figure 5. Notice that the indices for each alternative are 100% for a single rank, and 0% for the others, therefore defining a deterministic pre-order (S4.1 in the first rank, S3.1 in the second, etc).

4.2 Imprecise weights

In the first re-analysis, we define the feasible weight space uniformly distributed and constrained to include original weights $\pm 10\%$ in all dimensions. This weight information is introduced in the software as shown in Figure 6. The choice of intervals is quite arbitrary, and in a real-world application the intervals should be approved by the DMs.

The effect of a variable number of ranks in different simulations makes it hard to interpret the results based on only the rank acceptability indices. Pair-wise winning indices are better for seeing how imprecise weights affect the ranking in comparison with exact weights. The pair-wise winning indices of this re-analysis are presented in Figure 7.

By looking at the pair-wise winning indices in Figure 7, we can see that the impre-

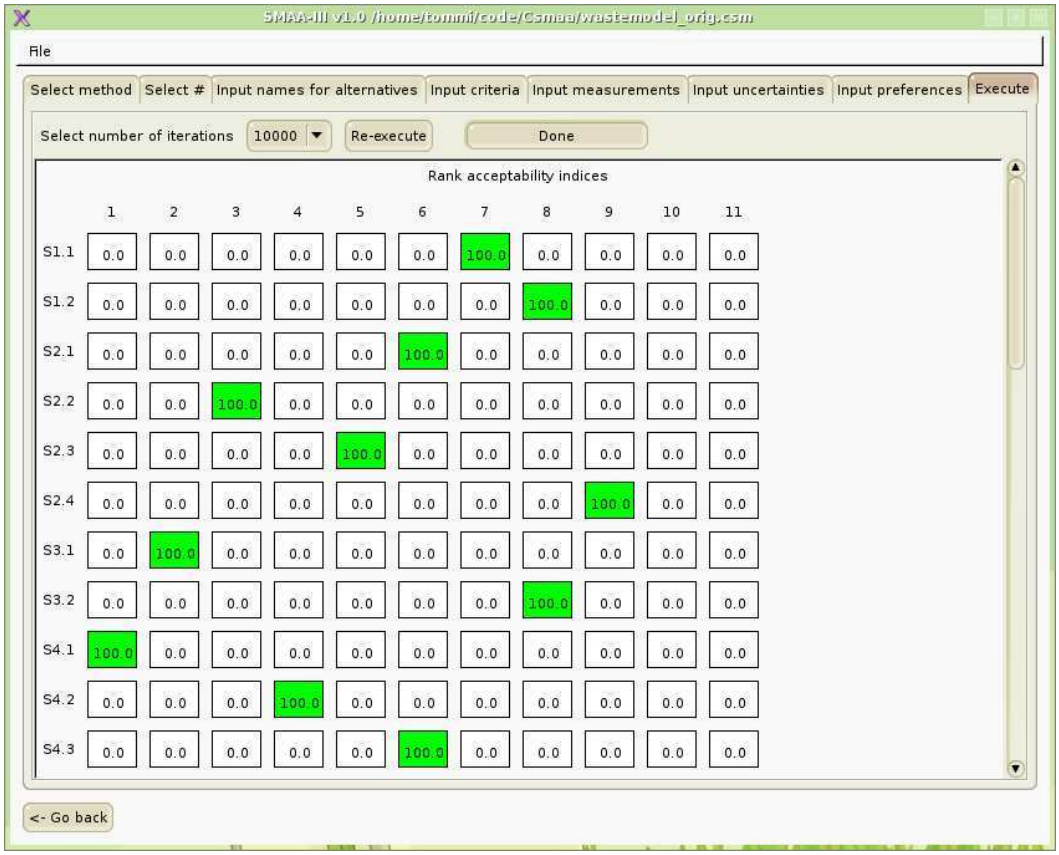


Figure 5. Rank acceptability indices of the pre-order of the original study.

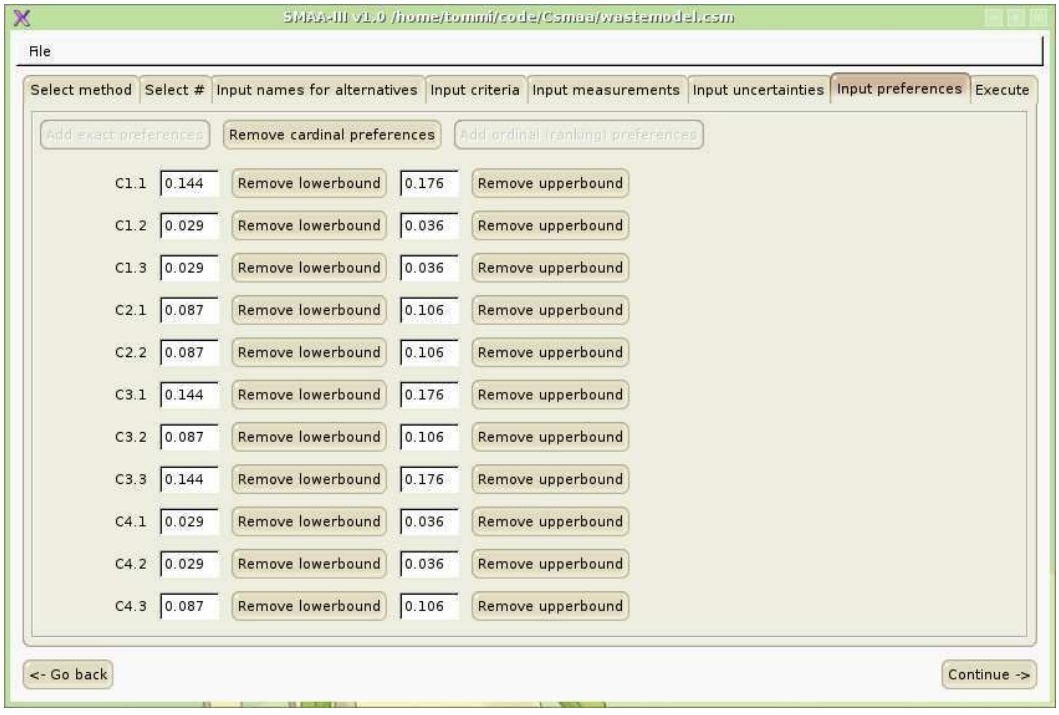


Figure 6. Input of ranges for weights in the software.

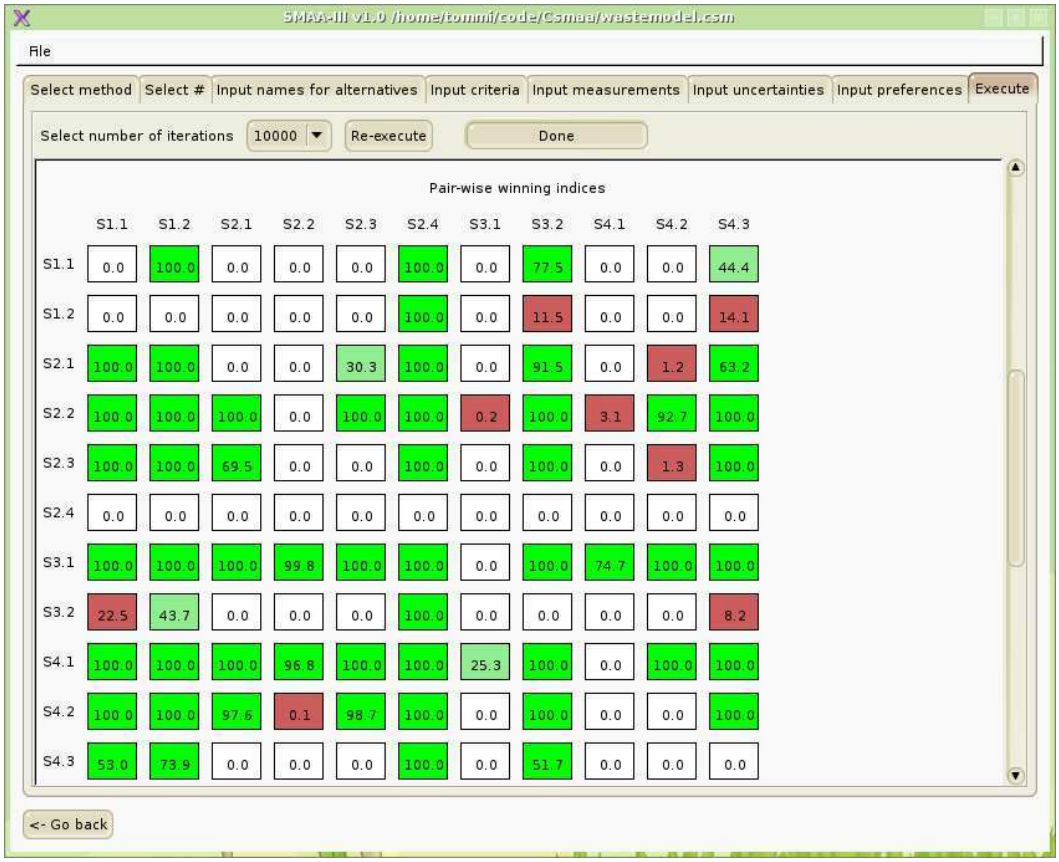


Figure 7. Pair-wise winning indices of the re-analysis with imprecise weights.

cise weights cause some alterations in the results. When the parameters are deterministic, there are only indices of 100% and 0%. Now with the imprecise values, for example, S2.1 is ranked better than S2.3 with 30% of the feasible weights values (had a worse rank in the original analysis). What is more important, is the change in ranking of alternatives that obtained good ranks in the original analysis. Although ELECTRE III is a ranking method, most of its applications, including this study, are in *selecting* an alternative to implement. The two best ranked alternatives in the original case study are S4.1 (1st rank) and S4.1 (2nd rank). But in this re-analysis incorporating imprecise weights, S3.1 obtains higher rank than S4.1 with 75% of the feasible weights. Therefore, by taking into consideration the purpose of the application (implementing the best alternative), we can say that weights are among the sensitive parameters of this model.

4.3 Imprecise thresholds

The second case re-analyzed is with imprecise thresholds. We define imprecise thresholds for cardinal criteria with $\pm 10\%$ imprecision. We analyze the imprecision this time by looking at the rank acceptability indices, presented in Figure 8. This figure shows a common phenomenon with rank acceptability indices, emerg-

ing when the parameters are sensitive: the amount of ranks in different simulation runs changes. This is caused by some alternatives obtaining the same rank, therefore lowering the total number of ranks. This effect does not affect the first rank acceptabilities, but is cumulative in higher ranks.

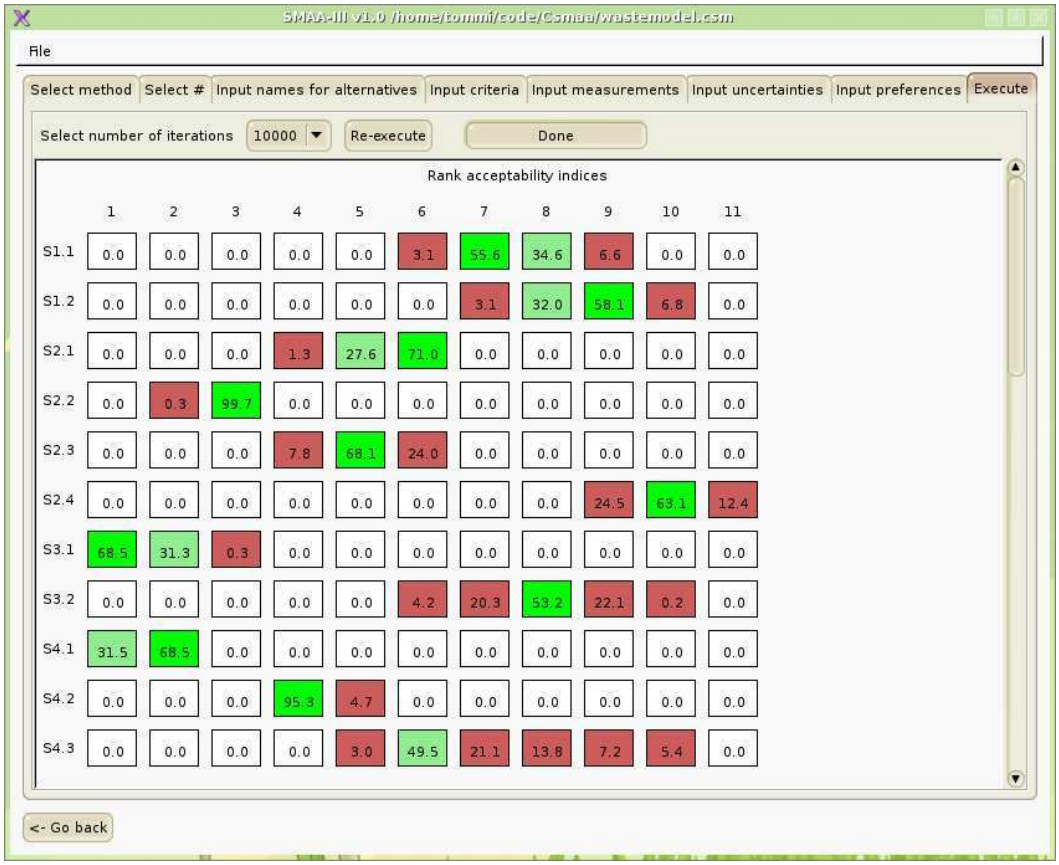


Figure 8. Rank acceptability indices of the re-analysis with imprecise thresholds.

For example, see the rank acceptabilities of alternative S2.4: 24.5% for rank 9, 63.1% for rank 10, and 12.4% for rank 11. Based on the rank acceptability indices, it would seem that the ranking of this alternative varies quite a lot. But by looking the pair-wise winning indices of the same re-analysis presented in Figure 9, more precise information is obtained. The column of alternative S2.4 tells the share of feasible parameter values for other alternatives to obtain a better rank than S2.4. All these (except for the alternatives own row, which is always 0%) are 100%, which means that alternative S2.4 is *always* ranked the last and never shares this rank with an another alternative.

Although this cumulative effect causes the rank acceptabilities to lose their intended meaning for the worse ranks, it does not hinder their help to decisions of selecting the best alternative. They also demonstrate in a comprehensive way the ranges of ranks for which alternatives can be assigned into, and whether these are overlapping with the corresponding ranges of another alternatives or not.

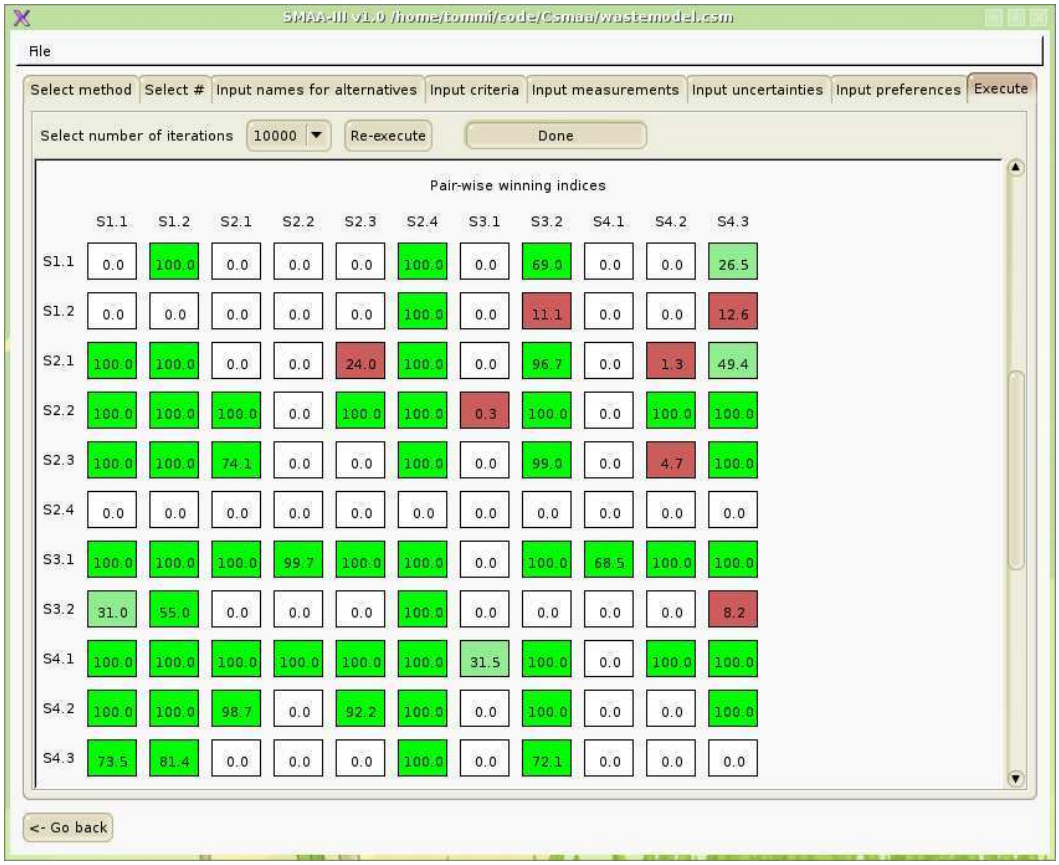


Figure 9. Pair-wise winning indices of the re-analysis with imprecise thresholds.

4.4 Imprecise criteria measurements

In the third re-analysis we add imprecision to the criteria measurements. We define $\pm 10\%$ imprecision interval for each cardinal criterion. The rank acceptability indices of this re-analysis are presented in Figure 10. As was expected, the results show quite a large amount of dispersion in the indices. In a real application, the uncertainties could have been quantified more precisely. Anyhow, one crucial observation should be made: S3.1 obtains the 1st rank with 85% of the possible criteria measurements, while S4.1 achieves the same with only 15% of the measurements.

This observation confirms what has been reported by Lahdelma and Salminen (2002): thresholds cannot be used to model imprecision. In the original study the thresholds were used for modelling imprecision in the criteria measurements, and the analysts ended up in recommending S4.1 as the most “robust” option. But by considering the criteria measurements to have uniformly distributed values in an imprecision interval, S3.1 seems to be more robust candidate for the first rank.

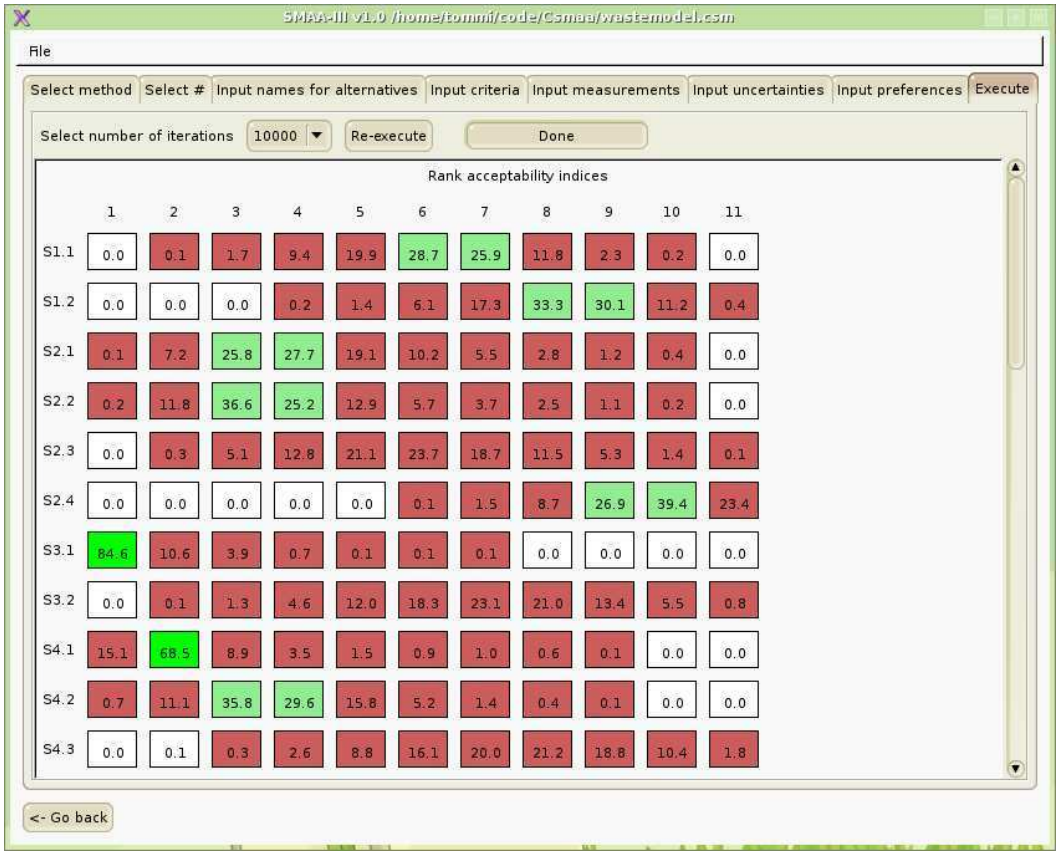


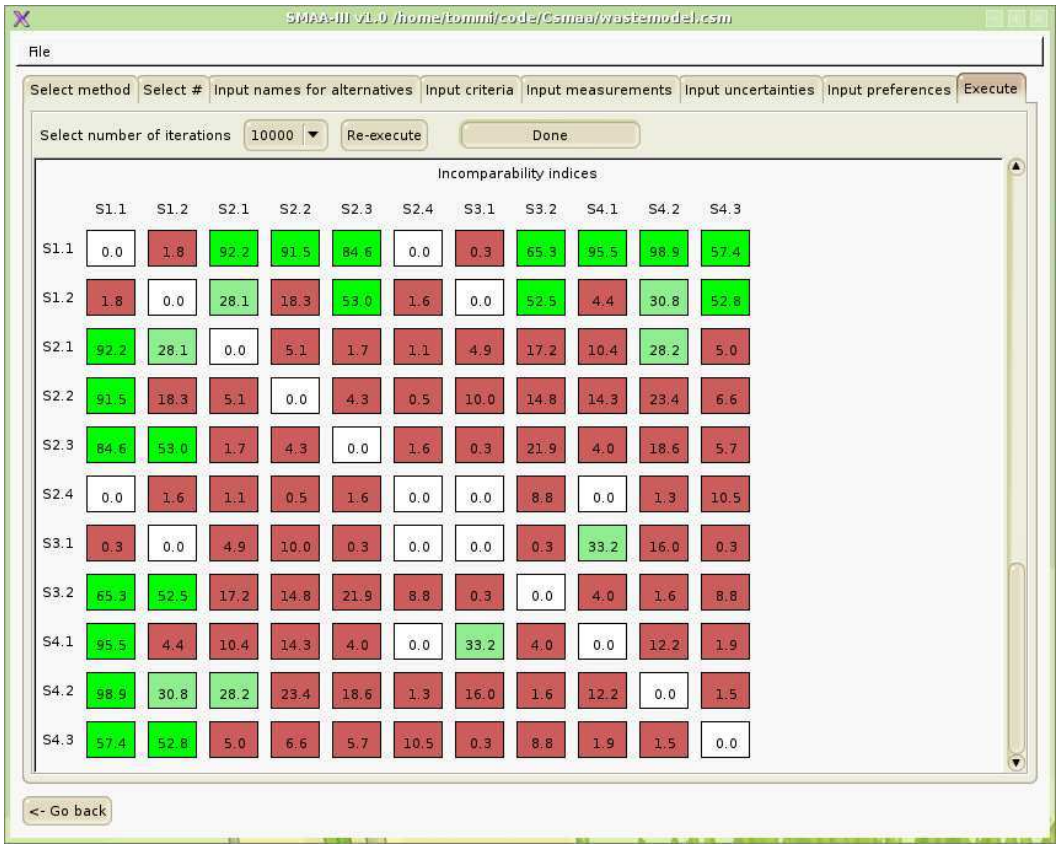
Figure 10. Rank acceptability indices of the re-analysis with imprecise criteria measurements.

4.5 All values imprecise

The last re-analysis is with all types of imprecisions applied in the previous re-analyses. Rank acceptability indices and pair-wise winning indices of this one are similar to the previous re-analyses, showing (as expected) even more dispersion in the values. Therefore, we will not present these indices here.

The incomparability indices of this re-analysis are presented in Figure 11. What should be noticed from this figure, is the high amount of incomparability. In a real decision making situation, most (if not all) parameter values are defined with imprecise values. This imprecision causes a quite large share of the pairs of alternatives to have incomparability indices of a reasonable magnitude. They can therefore not be left out of consideration is the DMs want to make the decision taking into account incomparability as well, and it brings an extra level of complexity to the decision making process. In our opinion, this is another reason why modelling incomparability is not useful in practical decision making.

Before summarizing the results of this section, we note that the authors of the case study also describe a robustness analysis of the results. But in their study, the ro-



	S1.1	S1.2	S2.1	S2.2	S2.3	S2.4	S3.1	S3.2	S4.1	S4.2	S4.3
S1.1	0.0	1.8	92.2	91.5	84.6	0.0	0.3	65.3	95.5	98.9	57.4
S1.2	1.8	0.0	28.1	18.3	53.0	1.6	0.0	52.5	4.4	30.8	52.8
S2.1	92.2	28.1	0.0	5.1	1.7	1.1	4.9	17.2	10.4	28.2	5.0
S2.2	91.5	18.3	5.1	0.0	4.3	0.5	10.0	14.8	14.3	23.4	6.6
S2.3	84.6	53.0	1.7	4.3	0.0	1.6	0.3	21.9	4.0	18.6	5.7
S2.4	0.0	1.6	1.1	0.5	1.6	0.0	0.0	8.8	0.0	1.3	10.5
S3.1	0.3	0.0	4.9	10.0	0.3	0.0	0.0	0.3	33.2	16.0	0.3
S3.2	65.3	52.5	17.2	14.8	21.9	8.8	0.3	0.0	4.0	1.6	8.8
S4.1	95.5	4.4	10.4	14.3	4.0	0.0	33.2	4.0	0.0	12.2	1.9
S4.2	98.9	30.8	28.2	23.4	18.6	1.3	16.0	1.6	12.2	0.0	1.5
S4.3	57.4	52.8	5.0	6.6	5.7	10.5	0.3	8.8	1.9	1.5	0.0

Figure 11. Incomparability indices of the re-analysis with all parameters imprecise.

bustness is analyzed through weights variations, by altering a single weight at a time. It is more a sensitivity analysis than a robustness one. As has been shown by our re-analyses, a different way should be considered to analyze robustness of ELECTRE III results; stability of all parameters should be analyzed in such a way, that the whole space of feasible parameter values is explored. Otherwise, nonlinearity of the ranking function, which ELECTRE III represents, can produce surprising results. The re-analyses of this particular case study showed that S3.1 would have been a more robust alternative than S4.1 for implementation.

5 Conclusions and avenues for future research

In this paper we introduced a new method, SMAA-III, that allows the parameters and criteria measurements of ELECTRE III to be imprecise, and to be defined with various types of constraints: no deterministic values are required. This has numerous advantages, especially in the context of MCDM with multiple DMs, because the parameters can be determined as intervals that contain the preferences of all DMs. It also allows an easily applicable robustness analysis to be performed.

We presented a software [1] implementing the proposed method, and used it to re-analyze multiple times an existing real-world case study. These re-analyses were done to study the effect of imprecision in different parameters on the results. The analyses showed that in this case all the parameters of ELECTRE III were sensitive for reasonable changes. This confirms results of the simulation study by Lahdelma and Salminen (2002): pseudo-criterion based models are unstable with respect to changes in the threshold values and indifference threshold cannot be used to model imprecision in the data.

Future research should study usefulness of SMAA-III in real-life cases. It should be studied, if the indices of SMAA-III can be interpreted in a meaningful way by analysts less accustomed to SMAA methods. In addition, new techniques for visualizing the indices are needed.

Acknowledgements

The work of Tommi Tervonen was supported by grants from Turun Yliopistosäätiö and the Finnish Cultural Foundation.

References

Belton, V., Stewart, T. J., 2002. Multiple Criteria Decision Analysis - An Integrated Approach. Kluwer Academic Publishers, Dordrecht.

Butler, J., Dia, J., Dyer, J., 1997. Simulation techniques for the sensitivity analysis of multi-criteria decision models. European Journal of Operational Research 103 (3), 531–545.

Figueira, J., Mousseau, V., Roy, B., 2005. ELECTRE methods. In: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Multiple Criteria Decsion Analysis: State of the Art Surveys. Springer Science+Business Media, Inc., New York, Ch. 4.

Figueira, J., Roy, B., 2002. Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. European Journal of Operational Research 139, 317–326.

Hokkanen, J., Salminen, P., 1997. Choosing a solid waste management system using multicriteria decision analysis. European Journal of Operational Research 98, 19–36.

Keeney, R., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Wiley & Sons, New York.

Lahdelma, R., Salminen, P., 2001. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Operations Research 49 (3), 444–454.

[1] The software presented in this paper is available from the corresponding author. It is distributed free for academic use.

Lahdelma, R., Salminen, P., 2002. Pseudo-criteria versus linear utility function in stochastic multi-criteria acceptability analysis. European Journal of Operational Research 14, 454–469.

Leskinen, P., Viitanen, J., Kangas, A., Kangas, J., 2006. Alternatives to incorporate uncertainty and risk attitude in multicriteria evaluation of forest plans. Forest Science 52 (3), 304–312.

Mousseau, V., 1995. Eliciting information concerning the relative importance of criteria. In: Pardalos, Y., Siskos, C., Zopounidis, C. (Eds.), Advances in Multicriteria analysis. Kluwer Academic Publishers, Dordrecht, pp. 17–43.

Rogers, M., Bruen, M., 1998. A new system for weighting environmental criteria for use within ELECTRE III. European Journal of Operational Research 107, 552–563.

Rogers, M., Bruen, M., Maystre, L.-Y., 2000. ELECTRE and Decision Support: Methods and Applications in Engineering and Infrastructure Investment. Kluwer Academic Publishers, Dordrecht, Massachusetts, U.S.A.

Roy, B., Présent, M., Silhol, D., 1986. A programming method for determining which Paris metro stations should be renovated. European Journal of Operational Research 24, 318–334.

Smets, P., 1991. Varieties of ignorance and the need for well-founded theories. Information Sciences 57–58, 135–144.

Tervonen, T., Figueira, J. R., 2006. A survey on stochastic multicriteria acceptability analysis methods. Journal of Multi-Criteria Decision Analysis (to appear).

Tervonen, T., Lahdelma, R., 2007. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research 178, 500–513.

Publication V

Tervonen T., Figueira J. R., 2007. A survey on stochastic multicriteria acceptability analysis methods. Journal of Multi-Criteria Decision Analysis (in press).

MCDA : 407
pp.1–14 (col.fig.: NIL)
PROD.TYPE: COM
3B2
ED: SANJAY
PAGN: S.SARAVANAN SCAN: XXXXXXXXXXX

JOURNAL OF MULTI-CRITERIA DECISION ANALYSIS
J. Multi-Crit. Decis. Anal. (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/mcda.407

A survey on stochastic multicriteria acceptability analysis methods

TOMMI TERVONEN[a,*] and JOSÉ RUI FIGUEIRA[b]
[a] *CEG-IST, Centre for Management Studies, Instituto Superior Técnico, Porto Salvo, Portugal*
[b] *Department of Information Technology, University of Turku, FIN-20014 Turku, Finland*

ABSTRACT

Stochastic multicriteria acceptability analysis (SMAA) comprises a family of multiple criteria decision-aiding (MCDA) methods for problems including incomplete, imprecise, and uncertain information. Methods of the family allow solving MCDA problems of various types. Even though the methods have been applied in the past in various real-life decision-making situations, the structure of a unified SMAA framework has not been studied. In this paper we describe the methods of the family and define a unified SMAA framework. We also point out the key points in the methodology for future research. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: stochastic multicriteria acceptability analysis (SMAA); multiple criteria decision aiding (MCDA); simulation

1. INTRODUCTION

Stochastic multicriteria acceptability analysis (SMAA) is a recently developed family of multiple criteria decision-aiding (MCDA) methods. Different SMAA methods can be used to handle the three main MCDA problem statements (Figueira *et al.* (2005)): choosing, ranking, and sorting. The methodology considers these problem statements in a wider sense. For example, instead of resulting in a ranking, the SMAA-2 method provides probabilities for alternatives to obtain certain ranks. The methodology is based on an inverse analysis of the space of feasible parameter values. It allows ignorance on criteria measurements and preferences. One of the advantages of SMAA over most other MCDA methodologies is that it can be used without any preference information.

We refer to ignorance divided into three subcategories: incompleteness, imprecision, and uncertainty (Smets, 1991). Incomplete information means that a value is missing. Imprecise information means that we have a value for the variable but not with the required precision. Uncertainty, instead, is a form of ignorance appearing when the observer is taken into account. It means that the observer gives complete and precise information, but is unreliable itself. For information and references on approaches dealing with ignorance, see Stewart (2005).

In this survey, we describe the methods and extensions of the SMAA family and provide recommendations on which method to use in different MCDA contexts. We find the key points of the methodology by defining an SMAA framework. We provide a list of published SMAA applications with the description of particularities of each one in order to give historical insight into the practices involved in the application of the methodology.

The rest of this paper is organized as follows: Section 2 describes the origins of the methodology. Section 3 contains a description of SMAA and SMAA-2, the methods that form the basis for the whole family. Extensions are presented in Section 4. The simulation technique used in the SMAA computations is described in Section 5. In Section 6, we present a table of applications. The SMAA framework is defined and discussed in Section 7. We end this paper with conclusions in Section 8.

2. ORIGINS OF SMAA

There are numerous MCDA methods that apply different approaches to tackle the difficulties encountered in real-life decision-making problems. One of the oldest, and the most successful ones, is

*Correspondence to: CEG-IST, Centre for Management Studies, Instituto Superior Técnico, Porto Salvo, Portugal. E-mail: figueira@ist.utl.pt

Copyright © 2007 John Wiley & Sons, Ltd.

UNCORRECTED PROOF

the utility function-based approach, in which alternatives are evaluated based on utility scores derived using a function operating on criteria values. The utility function-based approach has been researched intensively and applied in various models (see Figueira *et al*. (2005)). Although the approach has a history of relevant applications, it has become apparent that the exact parameter values required by earlier methods of the approach were not sufficient in all decision-making situations. In some, the decision makers (DMs) did not want to reveal their preference model, and therefore exact parameter values could not be obtained; in others, the alternatives had uncertain or imprecise values for criteria measurements. Therefore, new advances seem necessary to preserve the usefulness of the approach.

One way to overcome the weaknesses of the utility function-based approach is through an inverse method: instead of asking parameter values and giving an answer to the problem in question, the values resulting in different outcomes are described. The inverse SMAA method includes computing multidimensional integrals over feasible parameter spaces in order to support DMs with descriptive measures. The method solves various problems encountered in the traditional approach by allowing to use parameters with ignorance on the values. For example, usually different weight elicitation techniques produce different values; therefore, deterministic weights are harder to justify than, for example, weight intervals.

Before SMAA there were other inverse MCDA methods. Two of the most important ones for the development of SMAA are the comparative hypervolume criterion and the overall compromise criterion.

2.1. Comparative hypervolume criterion

The first precursor of the SMAA methodology was introduced by Charnetski (1973) and Charnetski and Soland (1978), through their comparative hypervolume criterion. This method is based on computing, for each alternative, the volume of the multi-dimensional weight space that makes each alternative the most preferred one. It can handle preference information in form of linear constraints for weights but is restricted to deterministic criteria measurements and an additive utility function. Rietveld (1980) and Rietveld and Ouwersloot (1992) presented similar methods for problems with ordinal criteria and ordinal preference information.

2.2. Overall compromise criterion

The overall compromise criterion by Bana e Costa (1986) is another precursor of the SMAA methodology. The method consists of calculating the amount of conflict between the preferences of different DMs to define a joint probability density function for the weight space. Although in theory it is very useful, in practice the method is rather limited, as it can handle only three criteria.

3. SMAA AND SMAA-2

The discrete decision-making problem considered refers to a set of m alternatives $X = x_1, \ldots, x_i, \ldots, x_m\}$, which are evaluated on the basis of n criteria $\{g_1, \ldots, g_j, \ldots, g_n\}$. The evaluation of action x_i on criterion g_j is denoted $g_j(x_i)$. The model considers multiple DMs, each having a preference structure representable through an individual weight vector w and a real-valued utility function $u(x_i, w)$ that has a commonly accepted shape. The most commonly used utility function is the linear one:

$$u(x_i, w) = \sum_{j=1}^{n} w_j g_j(x_i) \tag{1}$$

The weights will be assumed non-negative and normalized. Therefore, the feasible weight space will be

$$W = \left\{ w \in R^n : w \geqslant 0 \text{ and } \sum_{j=1}^{n} w_j = 1 \right\}$$

The SMAA methods are developed for situations where criteria values and/or weights or other model parameters are not precisely known. Uncertain or imprecise criteria values are represented by stochastic variables ξ_{ij} corresponding to the deterministic evaluations $g_j(x_i)$ with density function $f_\chi(\xi)$ in the space $\chi \subseteq R^{m \times n}$. Similarly, the DMs unknown or partially known preferences are represented by a weight distribution with joint density function $f_W(w)$ in the feasible weight space W. Total lack of preference information on weights is represented by the uniform weight distribution in W:

$$f_W(w) = 1/\mathrm{vol}(W)$$

As for the utility function-based approaches, one

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

should note that the weights are defined as scale factors: the weights rescale the values of partial utility functions in such a way that the full swing in the scaled function indicates the importance of the criterion (see Belton and Stewart, 2002, Section 5.4).

3.1. SMAA

The fundamental idea of SMAA is to provide decision support through descriptive measures calculated as multidimensional integrals over stochastic parameter spaces. The original SMAA (Lahdelma *et al*., 1998) introduced three measures: the acceptability index, the central weight vector, and the confidence factor. For this purpose, the set of favourable weights $W_i(\xi)$ is defined as follows:

$$W_i(\xi) = \{w \in W : u(\xi_i, w) \geqslant u(\xi_k, w), \forall k = 1, \ldots, m\} \tag{2}$$

Any weight $w \in W_i(\xi)$ makes the overall utility of x_i greater than or equal to the utility of all other alternatives.

The descriptive measures of SMAA are computed through Monte Carlo simulation. This means that they might contain errors, but the error margins are so small that usually they do not have to be taken into account (when the number of Monte Carlo iterations is large enough, see Section 5).

3.1.1. Acceptability index. The acceptability index of an alternative describes the share of different valuations making an alternative the most preferred one. It is computed as a multidimensional integral over the criteria distributions and the favourable weight space as

$$a_i = \int_{\xi \in \chi} f_\chi(\xi) \int_{w \in W_i(\xi)} f_W(w)\, \mathrm{d}w\, \mathrm{d}\xi \tag{3}$$

Acceptability indices can be used to classify alternatives into stochastically efficient ($a_i \gg 0$) or inefficient ones (a_i near zero, for example, < 0.05). A zero acceptability index means that an alternative is never considered the best with the assumed preference model. For stochastically efficient alternatives, the index measures the strength of the efficiency considering simultaneously the ignorance on the criteria measurements and the DMs' preferences.

Scaling the criteria affects the acceptability indices. Therefore, scaling must not performed arbitrarily when trying to classify the alternatives on the basis of acceptability indices (Lahdelma and Salminen, 2001). For example, if the minimum and maximum criterion values are chosen as the corresponding scaling points, the possible introduction of a new alternative might change these values and, therefore, also the acceptability indices to a large extent (Bana e Costa, 1988).

3.1.2. Central weight vector. The central weight vector w_i^c is defined as the expected centre of gravity of the favourable weight space. It is computed as a multidimensional integral over the criteria and weight distributions as

$$w_i^c = \frac{1}{a_i} \int_{\xi \in \chi} f_\chi(\xi) \int_{w \in W_i(\xi)} f_W(w) w\, \mathrm{d}w\, \mathrm{d}\xi \tag{4}$$

The central weight vector describes the preferences of a typical DM supporting this alternative with the assumed preference model. By presenting the central weight vectors to the DMs, an inverse approach for decision support can be applied: instead of eliciting preferences and building a solution to the problem, the DMs can learn what kind of preferences lead into which actions, without providing any preference information.

3.1.3. Confidence factor. The confidence factor p_i^c is defined as the probability for an alternative to be the preferred one with the preferences expressed by its central weight vector. It is computed as a multidimensional integral over the criteria distributions as follows:

$$p_i^c = \int_{\substack{\xi \in \chi : u(\xi_i, w_i^c) \geqslant u(\xi_k, w_i^c) \\ \forall k = 1, \ldots, m}} f_\chi(\xi)\, \mathrm{d}\xi$$

The confidence factors measure whether the criteria measurements are accurate enough to discern the efficient alternatives. If the problem formulation is to choose an alternative to realize, the ones with low confidence factors should not be chosen. If they are deemed as attractive ones, more accurate criteria data should be collected in order to make a reliable decision.

3.2. SMAA-2

The acceptability index of the original SMAA method was not designed to rank the alternatives but to classify them to those which should be taken into future consideration and to those which should not. SMAA-2 (Lahdelma and Salminen, 2001) extends SMAA by taking into account all ranks and provides five new descriptive measures:

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

the rank acceptability index, three best rank-type measures, and the holistic acceptability index. These measures provide DMs with more insight into the decision-making problem. For that, a ranking function is defined as follows:

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

where $\rho(\mathrm{true}) = 1$ and $\rho(\mathrm{false}) = 0$. Note that $\mathrm{rank}(i,\xi,w) \in \{1,\ldots,m\}$. Let us also define the sets of favourable rank weights $W_i^r(\xi)$ as follows:

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

3.2.1. Rank acceptability index. The rank acceptability index is defined similarly to the acceptability index in (3), extending it to take into account the acceptability for a certain rank. The rank acceptability index b_i^r describes the share of parameter values granting alternative x_i rank r. It is computed as a multidimensional integral over the criteria distributions and the favourable rank weights as follows:

$$b_i^r = \int_{\xi\in\chi} f_\chi(\xi) \int_{w\in W_i^r(\xi)} f_W(w)\,\mathrm{d}w\,\mathrm{d}\xi$$

The most acceptable (best) alternatives are those with high acceptabilities for the best (smallest) ranks. Evidently, the rank acceptability indices are within the range $[0,1]$, where 0 indicates that the alternative will never obtain a given rank and 1 indicates that it will obtain the given rank always with any choice of weights. The first rank acceptability index b_i^1 is equal to the acceptability index a_i.

3.2.2. k-best rank indices. The rank acceptability indices are key indicators for the performance of alternatives. When the number of alternatives is large, it is sometimes appropriate to aggregate them in the early phase of the decision-making process to *k-best ranks* (*kbr*) *acceptabilities* as

$$a_i^k = \sum_{r=1}^{k} b_i^r$$

The kbr acceptabilities can be used in an iterative process in which weaker alternatives are eliminated until a small group of alternatives reach sufficient acceptabilities.

The central weight vectors can also be extended in a similar way to define the *central kbr weight vector* w_i^k as

$$w_i^k = \frac{1}{a_i} \int_{\xi\in\chi} f_\chi(\xi) \sum_{r=1}^{k} \int_{w\in W_i^r(\xi)} f_W(w) w\,\mathrm{d}w\,\mathrm{d}\xi$$

The kbr weight vector describes the preferences of a typical DM judging an alternative to be among the k best ones. Also the confidence factors can be extended similarly to define the kbr confidence factor p_i^k as

$$p_i^k = \int_{\xi\in\chi:\ \mathrm{rank}(i,\xi,w_i^k)\leqslant k} f_\chi(\xi)\,\mathrm{d}\xi$$

3.2.3. Holistic acceptability index. The problem of comparing alternatives in terms of their rank acceptabilities can be seen as a 'second-order' MCDA problem (Lahdelma and Salminen, 2001). The DMs attitude towards risk defines the required magnitude of confidence factors and acceptability indices. The rank acceptability indices can be aggregated into holistic acceptability indices a_i^{h} as

$$a_i^h = \sum_r \alpha^r b_i^r$$

where the α^r are called metaweights. There are numerous possible ways of choosing the metaweights (see Lahdelma and Salminen, 2001), the only constraints being that they should be non-negative, normalized, and non-increasing when the rank increases. However, using the holistic acceptability indices in decision-making has its limitations. This 'second-order' decision-making problem imposes an additional level of complexity to the indicators and adds assumptions that the DMs might not realize.

In our opinion the holistic acceptability indices should only be used when there is no analyst available or when SMAA is used as an automated decision-making tool. In these cases it should be questioned if SMAA was an appropriate method to apply in the first place. The most appropriate use of the holistic acceptability indices could be in problems with a large amount of alternatives, to filter out alternatives that do not deserve attention from the DMs. However, in this type of problems, the kbr acceptability indices might be more adequate.

3.3. Preference information

In most decision-making problems it is possible to elicit some preference information from the DMs.

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

This information can possibly be imprecise and uncertain. Although SMAA allows preference information to be represented with an arbitrary density function, it is usually easier to elicit the preferences as constraints for the weight space. Then, the density function is defined with a uniform distribution in the restricted weight space W' as

$$f_{W'}(w) = \begin{cases} 1/\mathrm{vol}(W') & \text{if } w \in W' \\ 0 & \text{if } w \in W \backslash W' \end{cases}$$

In particular, SMAA-2 introduces the following types of constraints:

1. Intervals for weights ($w_j \in [w_j^{\min}, w_j^{\max}]$).
2. Intervals for weight ratios (trade-offs) ($w_j/w_k \in [w_{jk}^{\min}, w_{jk}^{\max}]$).
3. Linear inequality constraints for weights ($Aw \leqslant c$).
4. Non-linear inequality constraints for weights ($f(w) \leqslant 0$).
5. Partial or complete ranking of the weights ($w_j > w_k$).

When there are multiple DMs, the constraints have to be aggregated before applying. Possible non-interactive aggregation techniques include forming union or intersection or averaging weight space density functions of different DMs. There also exists a technique based on belief functions for eliciting and aggregating the preference information, see Tervonen *et al.* (2004b,c).

4. EXTENSIONS

In this section we will describe the most important SMAA extensions for ordinal criteria measurements, dependent criteria, cross-confidence factors, and those based on the outranking approach. There is also a variant of SMAA based on data envelopment analysis (SMAA-D) and another one applying prospect theory (SMAA-P), see Lahdelma and Salminen (2006b, 2003).

4.1. Ordinal criteria (SMAA-O)

SMAA-O (Lahdelma *et al.*, 2003) extends SMAA to consider ordinal criteria measurements, meaning that the DMs have ranked the alternatives according to each (ordinal) criterion. In SMAA-O, the ordinal information is mapped to cardinal without forcing any specific mapping. This means that nothing is assumed about the weights of criteria ranks in the piecewise linear mapping.

The possibility of using ordinal measurements has its advantages. Usually the experts defining criteria measurements can rank alternatives with respect to each criterion faster than if they use cardinal measurements. Therefore, if ordinal measurements provide sufficient accuracy for the decision-making problem in question, savings can be obtained.

Ordinal criteria are measured by assigning for each alternative a rank-level number $r_j = 1, \ldots, j^{\max}$, where 1 is the best and $j^{\max}$ the worst rank-level. Alternatives considered equally good are placed on the same rank level and rank levels are numbered consecutively. On an ordinal scale, the scale intervals do not contain any information and should therefore be treated as such without imposing any extra assumptions. However, some mapping can be assumed to underlie the ordinal information. In SMAA-O, all mappings that are consistent with the ordinal information are simulated numerically during Monte Carlo iterations. This means generating random cardinal values for the corresponding ordinal criteria measurements in a way that preserves the ordinal rank information. Figure 1 illustrates a sample mapping generated in this way.

The SMAA methods can be used with any kind of utility function jointly accepted by the DMs, but if we have an additive utility function, the shape of the function will be considered unknown. In this case, the DMs partial utility functions are simulated in the same way as the ordinal to

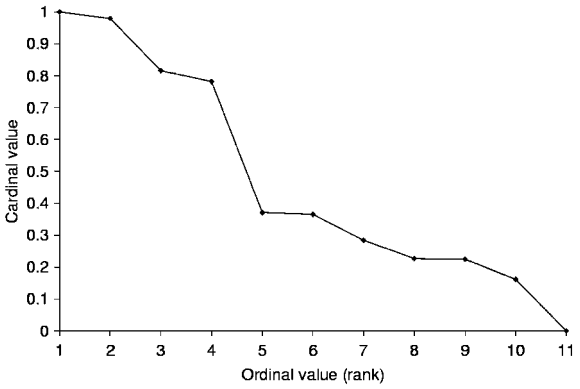


Figure 1. A sample ordinal to cardinal mapping of SMAA-O (Lahdelma *et al.*, 2003).

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

cardinal mappings. However, simulation is not necessary for ordinal criteria because the simulated cardinal values can be interpreted directly as partial values on a linear scale. Therefore, if the DMs accept an additive utility function, it is not necessary for the DMs to agree on a common shape of the partial utility functions for the ordinal criteria.

SMAA-O has been combined with the so-called SWOT methodology in the work of Kangas *et al.* (2003b). For an alternative technique for applying ordinal criteria in simulation-based approaches, see Leskinen *et al.* (2004).

4.2. Handling-dependent criteria

In many real-life applications of SMAA, the criteria measurements as well as their uncertainties are dependent. If we do not consider them as such, the results will contain bias (Lahdelma *et al.*, 2006b). SMAA allows using external sampling as a source for criteria measurements. This technique implicitly takes into account the dependencies. Another technique reported in the literature (Lahdelma *et al.*, 2006a) is to model the criteria with a multivariate Gaussian distribution.

Although it may be possible to determine the correlation of the variables 'by hand', in practice in most applications it is too time consuming or even impossible. The multivariate Gaussian model is more suitable in applications, in which there is a simulation model or real-life process producing values for the criteria measurements (see e.g. Tervonen *et al.*, 2007b).

4.3. Cross-confidence factors

SMAA has been developed for problems with ignorance on both the preferences and the criteria measurements. When the information is very imprecise, problems emerge because a large set of alternatives might seem acceptable as indicated by the acceptability indices. In this kind of situations, it would be desirable to obtain more precise information on the DMs preferences and on the criteria measurements, but it is not always possible due to time or money constraints.

One technique to improve discrimination over a large set of efficient alternatives is to use *cross-confidence factors* (Lahdelma and Salminen, 2006a). These descriptive measures are confidence factors computed for each alternative using each other's central weight vectors. The cross-confidence factor for alternative x_i with respect to alternative x_k is computed as

$$p_i^k = \int_{\xi\in\chi: w_k^c \in W_i^1(\xi)} f_\chi(\xi)\,\mathrm{d}\xi$$

defined when the target alternative is efficient (and therefore has a central weight vector defined). The cross-confidence factor is the probability for an alternative to obtain the first rank (assuming ignorance on the criteria measurements) when the central weight vector of the target alternative is chosen.

The cross-confidence factors provide additional information about why the discrimination of alternatives is weak: an alternative that obtains a high-cross-confidence factor with respect to another is similar, and because of that, poorly discriminated. To identify such alternatives, the model defines reference sets, which are ordered stochastic sets of pairs (a, p_i^k):

$$\{(i(k,r), p_{i(k,r)}^k) | r = 1, \ldots, m(k)\}$$

where $m(k) \leqslant m$ determines the number of elements in the reference set and the index function $i(k,r)$ orders the elements by their cross-confidence factors into descending order. This ordering facilitates quick identification of the most poorly discriminated alternatives. Reference sets can be visualized as column charts, as shown in Figure 2. The reference sets of this figure present a case in which alternative x_1 is very similar to x_2, and when using the central weight vector of x_2, x_1 has a higher probability to obtain the first rank.

4.4. Reference point approach (Ref-SMAA)

Although the SMAA methods can be used without any information on the weights, it is preferable to try to elicit some information from the DMs. Rather than using weights, a more straightforward technique for representing preferences is through reference points. With reference points, the DMs specify desirable or preferable values for each criterion, instead of specifying trade-offs between criteria. Reference points model satisfying behaviour, instead of trying to find optimal solutions, and can thus be more suitable in some decision-making contexts.

The Ref-SMAA method (Lahdelma *et al.*, 2005) (also called SMAA-A) allows to use reference points with multiple DMs by providing descriptive information about the sets of reference points that favour each alternative. An identical method

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

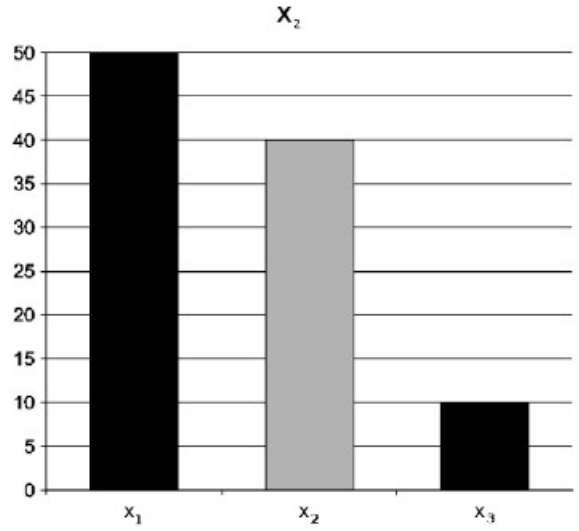


Figure 2. Sample cross-confidence factors of alternative x_2 in a 3-alternative problem.

(although with a simpler simulation model) has been presented by Durbach (2006).

Achievement functions are used to overcome some weaknesses of traditional goal programming and are used in Ref-SMAA to characterize non-dominated solutions. An achievement function is a function $s_{\check{x}} : X \to R$, where $\check{x} \in R^k$ is an arbitrary reference point. The achievement function of Ref-SMAA can be selected in various ways, as for example:

$$s_{\check{x}}(x_i) = \min_{i=1,\ldots,k} [w_i(x_i - \check{x}_i)] + \mu \sum_{i=1}^{k} w_i(x_i - \check{x}_i)$$

where μ is a sufficiently small scalar and w is a fixed positive scaling vector. Usually, w_i is set to be equal to the inverse of the difference between the best and the worst value for each criterion.

Ref-SMAA operates on the basis of a set of favourable reference points for each alternative x_i, defined as

$$\check{X}_i(\xi) = \{\check{x} \in \check{X} | s_{\check{x}}(\xi_i) \geqslant s_{\check{x}}(\xi_j),\ j = 1, \ldots, m\}$$

Any reference point $\check{x} \in \check{X}_i(\xi)$ makes the overall preference of x_i greater than or equal to the preference of any other alternative. The feasible reference point space $\check{X}_i$ can be defined according to needs, for example, as a convex combination of the reference points of all DMs. Similar to the acceptability index (3), Ref-SMAA defines the reference acceptability index r_i, computed as a multidimensional integral over the criteria value distributions and the favourable reference point space as

$$r_i = \int_{\xi \in \chi} f_\chi(\xi) \int_{\check{x}_i \in \check{X}_i(\xi)} (\xi) f(\check{x})\, \mathrm{d}\check{x}\, \mathrm{d}\xi$$

The *central reference point* $\check{x}_i$ is defined as the expected centre of gravity of the set of favourable reference points, computed as a multidimensional integral of the reference point vector $\check{x}$ over the criteria value distributions and the favourable reference point space as

$$\check{x}_i = \frac{1}{r_i} \int_{\xi \in \chi} f_\chi(\xi) \int_{\check{x}_i \in \check{X}_i(\xi)} f(\check{x}) \check{x}\, \mathrm{d}\check{x}\, \mathrm{d}\xi$$

All the descriptive measures of Ref-SMAA are related to reference points; therefore, the measures as well as the original alternatives all belong to the criterion space. For some DMS, this type of model might be easier to understand, as no artificial concepts such as weights are used.

4.5. Outranking-based SMAA approaches

SMAA has been extended for using instead of utility function (1) an outranking-based aggregation procedure to rank alternatives. This and other approaches described in this section are based on using ELECTRE-type pseudo-criteria. The pseudo-criteria are defined by using thresholds that are denoted as follows:

- $q_j(g_j(\cdot))$ is the *indifference threshold* for criterion g_j;
- $p_j(g_j(\cdot))$ is the *preference threshold* for criterion g_j, and, finally;
- $v_j(g_j(\cdot))$ is the *veto threshold* for criterion g_j.

By using these thresholds, a *concordance index* is defined. It is computed by considering individually for each criterion g_j the support it provides for the assertion of the outranking aS_jb, 'alternative a is at least as good as alternative b'. The partial concordance index is a fuzzy index computed as follows, for all $j = 1, \ldots, n$:

$$c_j(a,b) = \begin{cases} 1 & \text{if } g_j(a) \geqslant g_j(b) - q_j(g_j(b)) \\ 0 & \text{if } g_j(a) < g_j(b) - p_j(g_j(b)) \\ \dfrac{g_j(a) + p_j(g_j(b)) - g_j(b)}{p_j(g_j(b)) - q_j(g_j(b))} & \text{otherwise} \end{cases}$$

After computing the partial concordance indices, a comprehensive concordance index is calculated as

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

follows:

$$c(a,b) = \sum_{j \in J} w_j c_j(a,b)$$

If veto thresholds are used, a *discordance index* can also be defined. For more information on pseudo-criteria-based models, see Roy and Bouyssou (1993).

4.5.1. Outranking aggregation procedure (SMAA-3). SMAA-3 (Hokkanen *et al.*, 1998) method is a variant of the original SMAA that applies, instead of the utility function, ELECTRE-type pseudo-criteria and maxi–min choice procedure. According to this procedure, an alternative becomes the preferred one (not necessary unique) if the following set of constraints hold:

$$\min_{l=1,\ldots,m, l\neq i} c(x_i, x_l) \geqslant \min_{l=1,\ldots,m, l\neq k} c(x_k, x_l)$$

$$k = 1, \ldots, m, \quad k \neq i$$

On the basis of this the favourable weights of an alternative, (2) are redefined as

$$W_i = \left\{ w \in W : \min_{l=1,\ldots,m, l\neq i} \sum_{j=1}^{n} w_j c_j(x_i, x_l) \right.$$

$$\left. \geqslant \min_{l=1,\ldots,m, l\neq k} \sum_{j=1}^{n} w_j c_j(x_k, x_l) \right.$$

$$\left. k = 1, \ldots, k, k \neq i \right\}$$

The rest of the analysis is performed as in SMAA, with the exception that the criteria measurements are considered to be deterministic (no integration over χ is performed); therefore, no confidence factors are computed. It should be noted that now the central weight vector can lie outside the space of favourable weights of an alternative, because this preference model is non-linear. In this kind of (easily detectable) situations, a favourable weight vector is chosen with a minimal distance to the central weight vector.

In the literature there exists simulation tests of SMAA against SMAA-3. In these tests the results of SMAA-3 were found to be quite unstable with respect to the indifference threshold (Lahdelma and Salminen, 2002). Therefore, when SMAA-3 is applied in practice, great care should be put into choosing the thresholds.

There also exists a variant of SMAA that applies the complete ELECTRE III procedure to produce a ranking. For more details on it, see Tervonen *et al.* (2004a).

4.5.2. SMAA-TRI. All the SMAA variants described until here are for ranking or choosing problem statements. ELECTRE TRI (Yu, 1992) is a method for sorting problem statements, and SMAA-TRI (Tervonen *et al.*, 2007a) extends it to allow ignorance on the parameter values.

ELECTRE TRI uses concordance and discordance indices for sorting the alternatives into predefined and ordered categories. Let us denote the categories in ascending preference order $C_1, \ldots, C_h, \ldots, C_k$ (C_1 is the 'worst' category). These categories are defined by upper and lower profiles, which consist of measurements for all criteria. In the assignment procedure alternatives are iteratively compared with the profiles. The profiles are denoted $p_1, \ldots, p_h, \ldots, p_{k-1}$. p_h is the upper limit of category C_h and the lower limit of category C_{h+1}. The profiles have to be strictly ordered, that is, they have to satisfy

$$p_1 \Delta p_2 \Delta \cdots \Delta p_{k-2} \Delta p_{k-1} \tag{5}$$

where Δ is the dominance relation ($p_1 \Delta p_2$ means that p_2 dominates p_1). This dominance relation needs to be interpreted in a wide sense, because domination depends not only on the values of components of the two profiles but also on the threshold values. We will not describe the assignment procedure here, which requires an additional technical parameter, the lambda cutting level, to be defined. The interested reader should refer to Tervonen *et al.* (2007a).

SMAA-TRI is developed for parameter stability analysis of ELECTRE TRI and consists of analysing finite spaces of arbitrarily distributed parameter values in order to describe for each alternative the share of parameter values that assign it to different categories. It analyses the stability of weights, profiles, and the cutting level.

The input for ELECTRE TRI in SMAA-TRI is the following:

1. Uncertain or imprecise profiles are represented by stochastic variables ϕ_{hj} with joint density function $f_\Phi(\phi)$ in the space $\Phi \subseteq R^{(k-1)\times n}$. The joint density function must be such that all possible profile combinations satisfy (5). Usually the category profiles are defined to be independently distributed, and in this case the distributions must not overlap. For example, if the profile values for a criterion are Gaussian distributed, the distributions must have tails

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

truncated as shown by the vertical lines in Figure 3.

2. The lambda cutting level is represented as a stochastic variable Λ with density function $f_L \times (\Lambda)$ defined within the valid range [0.5, 1].
3. The weights and criteria measurements are represented as in SMAA-2.
4. The data and other parameters of ELECTRE TRI are represented by the set $T = \{M, q, p, v\}$. These components are considered to have deterministic values.

SMAA-TRI produces category acceptability indices for all pairs of alternatives and categories. The category acceptability index π_i^h describes the share of possible parameter values that have an alternative x_i assigned to category C_h. Let us define a *categorization function* that evaluates the category index h to which an alternative x_i is assigned by ELECTRE TRI:

$$h = K(i, \Lambda, \phi, w, T)$$

and a category membership function

$$m_i^h(\lambda, \phi, w, T) = \begin{cases} 1 & \text{if } K(i, \Lambda, \phi, w, T) = h \\ 0 & \text{otherwise} \end{cases}$$

which is applied in computing the category acceptability index numerically as a multi-dimensional integral over the finite parameter spaces as

$$\pi_i^h = \int_{0.5}^{1} f_L(\Lambda) \int_{\Phi} f_\Phi(\phi) \int_{W} f_W(w) m_i^h \times (\Lambda, \phi, w, T)\, \mathrm{d}w\, \mathrm{d}\phi\, \mathrm{d}\Lambda$$

The category acceptability index measures the stability of the assignment, and it can be interpreted as a fuzzy measure or a probability for membership in the category. If the parameters are stable, the category acceptability indices for each alternative should be 1 for one category and 0 for the others. In this case the assignments are said to be robust with respect to the imprecise parameters.

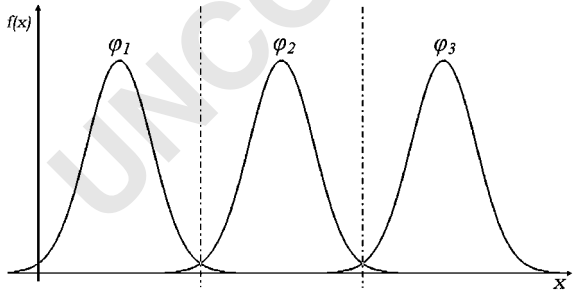


Figure 3. Probability distribution functions for three Gaussian distributed profile values (for a single criterion). The vertical lines show where the tails of the distributions must be truncated.

5. SIMULATION

The various distributions applied in the integrals of SMAA vary according to the application and can be arbitrarily complex. Usually the integrals have high dimensionality as well. Numerical integration techniques based on discretizing the distributions with respect to each dimension are infeasible, because the required effort depends exponentially on the number of dimensions. Therefore, instead of trying to obtain exact values for the integrals, Monte Carlo simulation is applied to obtain sufficiently accurate approximations. In this section we address the simulation technique, accuracy of the computations, and the complexity issues. For a full description of the algorithms, we refer to Tervonen and Lahdelma (2007).

5.1. Simulation technique

Monte Carlo simulation is applied in computation of the integrals. For all the acceptability index-type measures, a similar technique is applied: in each iteration, measurements for the parameters (criteria measurements, weights, etc.) are drawn from their corresponding joint distributions, and a ranking or a classification is built based on these values. After this, counters for the corresponding ranks or classes with respect to the alternatives are increased. After a number of iterations, the indices are obtained by dividing the counters with the number of iterations. The central weights are computed in a similar fashion, so that in each iteration, when an alternative obtains the first rank, the weight vector is added to its 'summed weight vector'. This vector is divided componentwise in the end by the number of iterations to obtain the central weight vector.

Weight generation is an important part of the simulation technique. If there is no preference information available, the n uniform distributed weights are generated as follows: first $n-1$ independent random numbers are generated from the uniform distribution within the range [0, 1],

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

UNCORRECTED PROOF

and sorted into ascending order $(q_1, \ldots, q_{n-1})$. After that, 0 and 1 are inserted as the first (q_0) and last (q_n) numbers, respectively. The weights are then obtained as intervals between consecutive numbers $(w_j = q_j - q_{j-1})$ (Tervonen and Lahdelma, 2007).

If there exists preference information, the weight generation technique must be altered. In the case of complete ordinal preference information, the weights can simply be sorted according to the ranking. Lower bounds for weights can be handled by using a simple transformation technique, because the lower-bounded feasible weight space is homomorphic with the original one. The lower-bounded weights are defined by generating the random numbers from interval $[0, 1 - s]$, where s is the sum of all lower bounds, and adding to them the corresponding lower bounds.

Upper bounds for weights cannot be handled with a similar technique, but instead a simple rejection technique is applied, in which the weight vectors not satisfying the upper bounds are rejected. The tip of the simplex cut off by the upper bounds has relatively small area compared with the one of lower bounds. Therefore, the increase in computational complexity due to upper bounds is relatively low. In addition, lower bounds might even render some of the upper bounds redundant. Consider for example a 3-criteria problem with lower bounds of 0.2 for all weights. The maximum value that any weight can obtain is $1 - 0.2 - 0.2 = 0.6$; therefore, all upper bounds higher than 0.6 are redundant. The amount of weights rejected due to upper bounds can be estimated in the following way: if we consider all weights to have a common upper bound $w^{\max}$, the probability for the largest of the generated weights to exceed the upper bound is

$$P[\max\{w_j\} > w^{\max})] = n(1 - w^{\max})^{n-1} - \binom{n}{2}(1 - 2w^{\max})^{n-1} + \cdots + (-1)^{k-1}\binom{n}{k}(1 - kw^{\max})^{n-1} \cdots$$

where the series continues as long as $1 - kw^{\max} > 0$ (David, 1970).

There exists an algorithm for generating random tuples in simple polytopes (Leydold and Hörmann, 1998). It can possibly be applied for generating upper-bounded weights in a more efficient manner. Future research should explore this possibility.

5.2. Accuracy of computations

Accuracy of computations can be calculated by considering the Monte Carlo simulations as point estimators for the descriptive measures. To achieve accuracy of A with 95% confidence for the rank acceptability indices, we need the following number of Monte Carlo iterations K (Milton and Arnold, 1995):

$$K = \frac{1.96^2}{4A^2}$$

For example, to achieve 95% confidence on error limits of ± 0.01 for the rank acceptability indices, we need to execute 9604 Monte Carlo iterations. The accuracy of confidence factors depends on the accuracy of central weight vectors in a complicated manner (Tervonen and Lahdelma, 2007), but if we disregard this source of error, the same equation for accuracy applies. The accuracy of the central weight vectors depends on the acceptability indices, and the required amount of iterations is calculated as follows:

$$K = \frac{1.96^2}{a_i 4A^2}$$

It should be noted that the accuracy of the computations does not depend on the dimensionality of the problem but only on the number of iterations.

5.3. Complexity issues

The required number of Monte Carlo iterations in typical SMAA applications is fairly high; therefore, for having practical applicability the complexity of SMAA computations should not be too high with respect to the number of criteria and alternatives. The complexity of SMAA-2 and SMAA-O has been analysed by Tervonen and Lahdelma (2007). The complexity of computing the acceptability indices and central weight vectors with independent criteria measurements and cardinal criteria is $O(K \cdot (n\log(n) + m \cdot n + m\log(m)))$. The complexity of computing the confidence factors is $O(K \cdot m^2 \cdot n)$. In these formulas K is the number of Monte Carlo iterations, m the number of alternatives, and n the number of criteria.

The use of ordinal criteria adds to the complexity with a factor of $\log(m)$. In practice this has very little effect (Tervonen and Lahdelma, 2007). What has a larger impact to the running times is the

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

handling of preference information. The formulas above assume that there are no constraints on the weights, which in practice is usually not the case. As described in Section 5.1, lower bounds for weights do not affect the complexity of the weight generation, but upper bounds might have a great impact on it.

6. APPLICATIONS

SMAA was originally developed in conjunction with a real-life decision-making problem and has been since applied in a variety of real-life cases. We list the published SMAA applications with their particularities in Table I.

7. THE FRAMEWORK

We define now a SMAA framework to decide a method to choose on a specific decision-making context. The first question to ask is whether we are dealing with a ranking or sorting problem. If we are dealing with a sorting one, the only method of the SMAA family we can use is SMAA-TRI. With ranking problems, we have to choose the type of preference model we have: whether it is based on weights or on reference points. If we have a weight-based model, we have to choose the type of aggregation procedure: utility function or outranking method. With the reference point approach we use Ref-SMAA. With all this information, we can choose whether to apply SMAA-2, SMAA-3, or Ref-SMAA for the ranking problem. Depending on the method to apply, we obtain as output different descriptive measures that can be used to derive 'second-order' aggregate measures. Choice of the method is presented as a decision-tree in Figure 4.

Other way to choose the method for a ranking problem is to question what kind of information is not available. Are the DMs willing to provide a shape for the utility function? If not, SMAA-2 cannot be applied. Same type of questions can be posed with respect to other parameters of the methods in order to find out which method would be the most suitable.

In the context of this framework, we should notice that all other methods than Ref-SMAA, which is based on reference points, can be used with arbitrary weight information. This means that we can apply them with no preference information at all, as well as with mixed information of ordinal and cardinal types. In practice, the most useful ones are (partial) ordinal information and cardinal weight constraints. Complex weight constraints might be hard for the DMs to under-

Table I. SMAA applications

Type, particularities	Applied method	Publication
Harbour citing Development of SMAA, no preference information available	SMAA	Hokkanen *et al.* (1999)
Waste treatment facility citing Only ordinal criteria	SMAA-2	Lahdelma *et al.* (2002)
Planning of a general plan Development of SMAA-3	SMAA-3	Hokkanen *et al.* (1998)
Ranking candidates for cleaning polluted soil Weight bounds applied	SMAA-2	Hokkanen *et al.* (2000)
Landfill reparation method choosing Ordinal and cardinal criteria, ordinal preferences	SMAA-2	Lahdelma *et al.* (2001)
Ecosystem management planning Both cardinal and ordinal criteria	SMAA-2	Kangas *et al.* (2003a)
Forest planning Comparison against multicriteria approval method	SMAA-2	Kangas and Kangas (2003)
Socioecological landspace planning Only ordinal criteria	SMAA-2	Kangas *et al.* (2005)
Elevator planning Dependent criteria	SMAA-2	Tervonen *et al.* (2008)

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

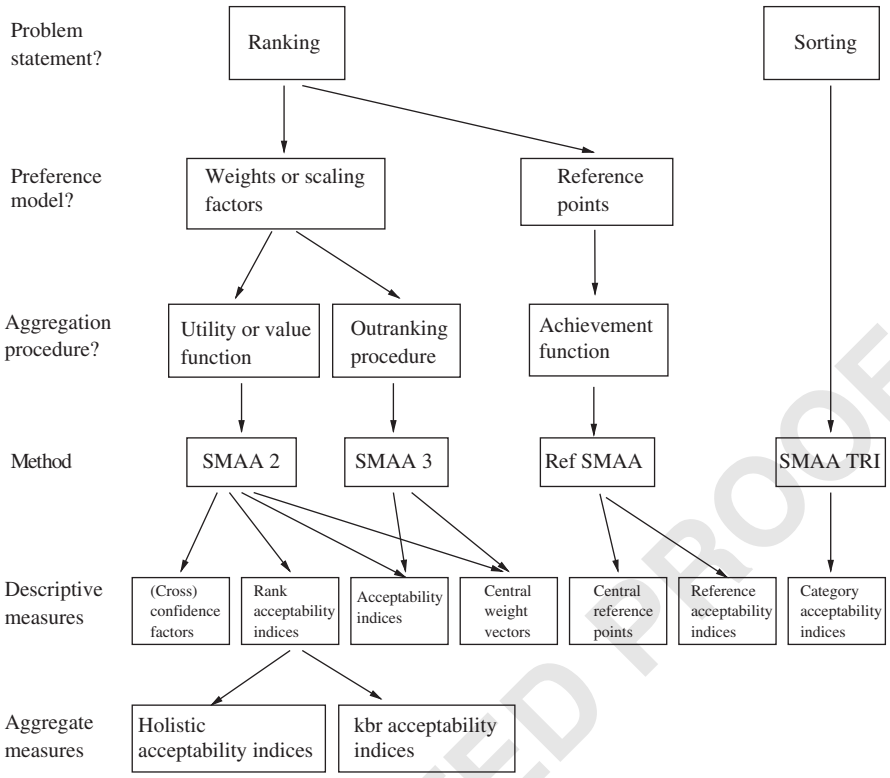


Figure 4. Decision-tree to choose the SMAA variant.

stand; therefore, by using more complex distributions, the possibility for the information to contain uncertainty increases. If the DMs have problems understanding the underlying preference model, the achievement function-based approach (Ref-SMAA) might be more suitable.

The shortcoming of the utility-function-based approach (SMAA-2) is that the scaling has large effect on the results, and the meaning of the weights is based on the scale. Therefore, if the shape of the utility function is hard to define, it might be more suitable to use SMAA-3 instead.

Arbitrarily distributed imprecise or uncertain criteria can be applied in all methods of the family except SMAA-3, which requires criteria measurements to have imprecision defined through thresholds. It should be noted that SMAA-O is not a stand-alone method, but rather a computational technique to handle ordinal criteria measurements. The possibility of using external sampling and the following generalization to use SMAA with external methods can be considered a great advantage. For example, the approach applied in SMAA-TRI can probably be applied to other methods as well, to use them with ignorance on the parameter values in order to analyse the stability of the results.

One of the unsolved questions in SMAA is how to obtain aggregate measures based on the different acceptability indices. The holistic acceptability indices can be used for this purpose, but they require meta-weights to be defined. This is an artificial concept with no connection to a cognitive decision-making process and; therefore, their use is hard to justify. It might be that to obtain more easily interpretable measures, we need to make more complex models. This would mean adding more parameters or using a more complex preference model. On the other hand, the complexity introduced in this way brings new sources of ignorance. More research should be put on this subject.

8. CONCLUSIONS

Stochastic multicriteria acceptability analysis (SMAA) is a recent methodology providing a

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

general framework that has extensions to handle various specificities in MCDA problems. In this paper, we presented the two basic methods, SMAA and SMAA-2, and the most important extensions of the methodology. The SMAA framework derived from these methods allows the decision analyst to choose the specific model to apply depending on the characteristics of the problem.

The SMAA methodology is applicable in a broad range of decision-making contexts. Nevertheless, there exists unsolved questions, the most important being whether we can develop aggregate measures that would help further in the decision-making process. The holistic acceptability index is such, but its applicability in practice is questionable. Therefore, future research on the methodology should address this area. Other crucial need is a user-friendly and computationally efficient software implementing the methodology. There is currently an open-source implementation of the basic methodology available (by one of the authors, downloadable from http://monet.fe.uc.pt/thesessoftware/), but it lacks a graphical user interface. As the principles of SMAA are quite simple although the equations for computing the descriptive measures look complicated, we believe that a software with a graphical user interface would allow the methodology to be applied in everyday decision-aiding problems by users less adapted to the techniques of numerical computation.

ACKNOWLEDGEMENTS

The work of Tommi Tervonen was supported by grants from Turun Yliopistosäätiö and the Finnish Cultural Foundation. The work of José Rui Figueira was partially supported by a research grant from CEG-IST.

REFERENCES

Bana e Costa CA. 1986. A multicriteria decision aid methodology to deal with conflicting situations on the weights. *European Journal of Operational Research* **26**: 22–34.

Bana e Costa CA. 1988. A methodology for sensitivity analysis in three-criteria problems: a case study in municipal management. *European Journal of Operational Research* **33**: 159–173.

Belton V, Stewart TJ. 2002. *Multiple Criteria Decision Analysis—An Integrated Approach*. Kluwer Academic Publishers: Dordrecht.

Charnetski J. 1973. The multiple attribute problem with partial information: the expected value and comparative hypervolume methods. *Ph.D. Thesis*, University of Texas at Austin.

Charnetski J, Soland R. 1978. Multiple-attribute decision making with partial information: the comparative hypervolume criterion. *Naval Research Logistics Quarterly* **25**: 279–288.

David HA. 1970. *Order Statistics*. Wiley: New York.

Durbach I. 2006. A simulation-based test of stochastic multicriteria acceptability analysis using achievement functions. *European Journal of Operational Research* **170**(3): 923–934.

Figueira J, Greco S, Ehrgott M (eds). 2005. *Multiple Criteria Decision Analysis*: *State of the Art Surveys*. Springer Science+Business Media, Inc.: New York.

Hokkanen J, Lahdelma R, Miettinen K, Salminen P. 1998. Determining the implementation order of a general plan by using a multicriteria method. *Journal of Multi-Criteria Decision Analysis* **7**(5): 273–284.

Hokkanen J, Lahdelma R, Salminen P. 1999. A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor. *Socio-Economic Planning Sciences* **33**: 1–23.

Hokkanen J, Lahdelma R, Salminen P. 2000. Multicriteria decision support in a technology competition for cleaning polluted soil in Helsinki. *Journal of Environmental Management* **60**: 339–348.

Kangas J, Hokkanen J, Kangas A, Lahdelma R, Salminen P. 2003a. Applying stochastic multicriteria acceptability analysis to forest ecosystem management with both cardinal and ordinal criteria. *Forest Science* **49**(6): 928–937.

Kangas J, Kangas A. 2003. Multicriteria approval and SMAA-O method in natural resources decision analysis with both ordinal and cardinal criteria. *Journal of Multi-Criteria Decision Analysis* **12**: 3–15.

Kangas J, Kurttila M, Kajanus M, Kangas A. 2003b. Evaluating the management strategies of a forestland estate—the S-O-S approach. *Journal of Environmental Management* **69**: 349–358.

Kangas J, Store R, Kangas A. 2005. Socioecological landscape planning approach and multicriteria acceptability analysis in multiple-purpose forest management. *Forest Policy and Economics* **7**: 603–614.

Lahdelma R, Hokkanen J, Salminen P. 1998. SMAA—stochastic multiobjective acceptability analysis. *European Journal of Operational Research* **106**: 137–143.

Lahdelma R, Makkonen S, Salminen P. 2006a. Multivariate Gaussian criteria in SMAA. *European Journal of Operational Research* **170**(3): 957–970.

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

Lahdelma R, Makkonen S, Salminen P. 2006b. Two ways to handle dependent uncertainties in multi-criteria decision problems. *Omega*, to appear.

Lahdelma R, Miettinen K, Salminen P. 2003. Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA). *European Journal of Operational Research* **147**: 117–127.

Lahdelma R, Miettinen K, Salminen P. 2005. Reference point approach for multiple decision makers. *European Journal of Operational Research* **164**(3): 785–791.

Lahdelma R, Salminen P. 2001. SMAA-2: stochastic multicriteria acceptability analysis for group decision making. *Operations Research* **49**(3): 444–454.

Lahdelma R, Salminen P. 2002. Pseudo-criteria versus linear utility function in stochastic multi-criteria acceptability analysis. *European Journal of Operational Research* **14**: 454–469.

Lahdelma R, Salminen P. 2003. Loss aversion in SMAA. *Technical Report 269*, School of Business and Economics, University of Jyväskylä.

Lahdelma R, Salminen P. 2006a. Classifying efficient alternatives in SMAA using cross confidence factors. *European Journal of Operational Research* **170**(1): 228–240.

Lahdelma R, Salminen P. 2006b. Stochastic multicriteria acceptability analysis using the data envelopment model. *European Journal of Operational Research* **170**(1): 241–252.

Lahdelma R, Salminen P, Hokkanen J. 2002. Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. *European Journal of Operational Research* **142**: 345–356.

Lahdelma R, Salminen P, Simonen A, Hokkanen J. 2001. Choosing a reparation method for a landfill using the SMAA-O multicriteria method. *Multiple Criteria Decision Making in the New Millenium*. Lecture Notes in Economics and Mathematical Systems, vol. 507. Springer: Berlin; 380–389.

Leskinen P, Kangas AS, Kangas J. 2004. Rank-based modelling of preferences in multi-criteria decision making. *European Journal of Operational Research* **158**: 721–733.

Leydold J, Hörmann W. 1998. A sweep-plane algorithm for generating random tuples in simple polytopes. *Mathematics of Computation* **67**(224): 1617–1635.

Milton JS, Arnold JC. 1995. *Introduction to Probability and Statistics* (3rd edn). Probability and Statistics. McGraw-Hill International editions: New York.

Rietveld P. 1980. *Multiple Objective Decision Methods and Regional Planning*. North-Holland: Amsterdam.

Rietveld P, Ouwersloot H. 1992. Ordinal data in multicriteria decision making, a stochastic dominance approach to siting nuclear power plants. *European Journal of Operational Research* **56**: 249–262.

Roy B, Bouyssou D. 1993. *Aide Multicritère à la Décision*: *Méthodes et Cas*. Economica: Paris.

Smets P. 1991. Varieties of ignorance and the need for well-founded theories. *Information Sciences* **57**–**58**: 135–144.

Stewart T. 2005. Dealing with uncertainties in MCDA. In *Multiple Criteria Decision Analysis*: *State of the Art Surveys*, Figueira J, Greco S, Ehrgott M (eds). Springer Science + Business Media, Inc.: New York; Chapter 1.

Tervonen T, Figueira J, Lahdelma R, Salminen P. 2004a. An inverse approach for ELECTRE III. *Research Report 20*/*2004*, The Institute of Systems Engineering and Computers (INESC-Coimbra), Coimbra, Portugal. Available at http://www.inescc.pt

Tervonen T, Figueira J, Lahdelma R, Salminen P. 2004b. Modelling MCDA group preferences for public human resource management: evaluating the quality of education at the Department of Information Technology, the University of Turku (Finland). *Research Report 22*/*2004*, The Institute of Systems Engineering and Computers (INESC-Coimbra), Coimbra, Portugal, Available at http://www.inescc.pt

Tervonen T, Lahdelma R, Dias JA, Figueira J, Salminen P. 2007. SMAA-TRI: a parameter stability analysis method for ELECTRE-TRI. In *Environmental Security in Harbors and Coastal Areas*, Kiker GA, Linkov I (eds). Springer: Berlin, 217–231.

Tervonen T, Hakonen H, Lahdelma R. 2007b. Elevator planning with Stochastic Multicriteria Acceptability Analysis. *Omega*, to appear.

Tervonen T, Lahdelma R. 2007. Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research* **178**(2): 500–513.

Tervonen T, Lahdelma R, Salminen P. 2004c. A method for elicitating and combining group preferences for Stochastic Multicriteria Acceptability Analysis. *Technical Report 638*, TUCS—Turku Center for Computer Science. Available at http://www.tucs.fi

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.

UNCORRECTED PROOF

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA – SER. AI

ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA

Osa
Tom.

365. Wang, Qi. Synthesis of Dinucleating Azacrown Ligands and Studies on Their Metal Ion Complexes as Cleaving Agents of RNA Phosphodiester Bonds. 101 p. 2007. EUR 15,00.

366. Lehtinen, Niina. Multiwavelength Studies of Solar Eruptions. 105 p. 2007. EUR 15,00.

367. Peurla, Markus. Magnetic Flux Pinning in $YBa_2Cu_3O_{6+x}$ Films Prepared by Pulsed Laser Deposition from Pure and $BaZrO_3$ Doped Nanocrystalline Targets. 106 p. 2007. EUR 15,00.

368. Lindfors, Elina. Modelling the Emission Mechanisms in Relativistic Jets. 82 p. 2007. EUR 13,00.

369. Kokko, Leena. Eu(III)-chelate Dyed Nanoparticles: Radiant Label Technology for Homogeneous Assays. 120 p. 2007. EUR 15,00.

370. Ropo, Matti. *Ab Initio* Simulation of Substitutionally Disordered Materials. 100 p. 2007. EUR 15,00.

371. Holopainen, Janne. Investigations of Structure in Dark Matter Haloes. 70 p. 2007. EUR 13,00.

372. Lehtinen, Janne. Improvements in the Assessment of Bacterial Viability and Killing. 116 p. 2007. EUR 15,00.

373. Viinikanoja, Antti. Self-Assembled Electroactive Mono and Multilayers: Preparation, Structure and Properties. 139 p. 2007. EUR 15,00.

374. Karonen, Maarit. Plant Proanthocyanidins. Characterization and Quantification by Degradation Methods, HPLC-DAD/ESI-MS and NMR. 139 p. 2007. EUR 15,00.

375. Tenho, Mikko. Significance of Crystallographic Properties in Pharmaceutical Compacts. 108 p. 2007. EUR 15,00.

376. Tervonen, Tommi. New directions in Stochastic Multicriteria Acceptability Analysis. 152 p. 2007. EUR 16,00.

Myynti/Distribution: Turun yliopiston kirjasto
Turku University Library
FI-20014 TURUN YLIOPISTO, FINLAND
puh/tel +358 (0)2 333 6170, fax +358 (0)2 333 5050
http://kirjasto.utu.fi/julkaisupalvelut/, julkmyynti@utu.fi

tai/or: Turun yliopiston kirjakauppa
Turku University Bookstore
FI-20014 TURUN YLIOPISTO, FINLAND
puh/tel +358 (0)2 333 6666, fax +358 (0)2 333 6660
http://kirjakauppa.utu.fi, tykk@utu.fi

ISBN 978-951-29-3405-8
ISSN 0082-7002
Painosalama Oy - Turku, Finland 2007