

De kwaliteit van afwegingsmethoden

Een onderzoek naar kwaliteitscriteria voor het gebruik van afwegingsmethoden voor integraal Waddenzeebeheer.

December 1998

ISBN: 90-365-1639-0

Het afstudeerverslag van:

F.J. (Rik) de Roode
Universiteit Twente (UT)
Civiele Technologie en Management (CT&M)
Modelleren van integrale Civieltechnische systemen
(MICS)

Voorwoord

U heeft zojuist mijn afstudeerverslag geopend. Daar ben ik blij mee. Het geeft mij het gevoel dat het werken nut heeft gehad, het wordt immers gelezen. Om dit verslag goed te kunnen begrijpen is het eigenlijk noodzakelijk om eerst een introductie in afwegingsmethoden te lezen. Ik heb hier zelf “evaluatiemethoden, een introductie”, van het ministerie van financiën voor gebruikt, en ik moet zeggen, nu ik dat gelezen heb, begrijp ik mijn verslag ook een stuk beter.

Het verslag is opgebouwd als een betoog. Ik raad u echter aan het niet zo te lezen. Waarschijnlijk is het veel prettiger om, na het inleidende hoofdstuk I, achteraan te beginnen met lezen. In hoofdstuk V treft u een interessant stappenplan voor complexe beslissingen aan en in het einde van hoofdstuk IV vindt u een korte evaluatie van een aantal afwegingsmethoden. De hoofdstukken daarvoor werken daar naar toe. Daar vindt u de argumentatie voor de soms boude uitspraken van de laatste bladzijden. Als u deze uitspraken al gelezen heeft worden de voorgaande hoofdstukken vast een stuk interessanter.

Dit afstudeerverslag is gemaakt in opdracht van de themagroep gebruik van Rijkswaterstaat. De Universiteit Twente heeft daarbij voor inhoudelijke begeleiding en faciliteiten als bureau en computer gezorgd. Over beide ben ik zeer te spreken. Daarom een kort dankwoord aan mijn begeleiders, Rien Kolkman en Anne van der Veen: “Heb dank voor begeleiding, voor het vele leeswerk en voor het inhoudelijke commentaar”. Ook Victor de Jonge, mijn externe begeleider wil ik bedanken, we hebben elkaar maar een paar keer gezien tijdens deze opdracht, maar de positieve reacties en de onvermijdelijke knipoog motiveren mij altijd weer. Daarnaast bedankt Jeroen voor het uitgebreide commentaar op dit verslag en Nadine, voor commentaar op het verslag natuurlijk, maar vooral voor de steun, de motivatie en de liefde waarmee je me ’s morgens het bed uitwerkt om de confrontatie met een nieuwe dag aan te gaan.

20 oktober 1998

Samenvatting

Er is een groot aantal afwegingsmethoden. Om uit deze methoden een methode te selecteren die goed is, moet er een uitspraak gedaan kunnen worden over de kwaliteit van deze methoden. Dit is echter niet eenvoudig. Afwegingsmethoden zijn prescriptief, ze schrijven voor hoe een afweging gemaakt dient te worden. Het is niet mogelijk om dit voorschrift op kwaliteit te testen door het te vergelijken met metingen. Daarom wordt in dit onderzoek gebruik gemaakt van kwaliteitscriteria. Deze meten niet direct maar indirect de kwaliteit van een afwegingsmethode.

In dit onderzoek is er voor gekozen om de afwegingsmethoden te benaderen als een model van het keuzeproses. Dit heeft als voordeel dat de afwegingsmethoden beschreven kunnen worden met de modelkenmerken van (Rip 1996) en dat ze getoetst kunnen worden aan de algemene modelkwaliteitscriteria van Covello (Kolkman 1996). Voor de afwegingsmethoden is gebruik gemaakt van de inventarisatie van het (Ministerie van financiën 1992).

Met behulp van Rip zijn de methoden besproken in hoofdstuk II. Eerst zijn de afwegingsmodellen als type gekarakteriseerd: ze zijn prescriptief, maar eenmaal ingevuld kunnen ze ook gebruikt worden als representatie van de gemaakte keuze en de argumentatie daarvoor of als microcosmos waarin naar hartelust gespeeld kan worden met voorkeuren en voorspellingen om zo inzicht te krijgen in de keuzemogelijkheden. Daarnaast zijn verschillende vormen van integratie, de vooronderstellingen die aan modellen ten grondslag liggen en verschillende functies die een model zou kunnen vervullen besproken. In hoofdstuk III zijn de algemene modelkwaliteitscriteria van Covello gepresenteerd. Daarna zijn ze gespecificeerd voor afwegingsmethoden. Het resultaat was een set criteria voor de kwaliteit van afwegingsmethoden en een aantal opmerkingen over goed gebruik van afwegingsmethoden.

In hoofdstuk IV is vervolgens per criterium de kwaliteit van de afwegingsmethoden besproken een groot aantal afwegingsmethoden viel hier hopeloos door de mand. Daarna zijn de resultaten per afwegingsmethode samengevat, zodat u per afwegingsmethode kunt lezen hoe zij scoren op deze kwaliteitscriteria. Tot slot is er aan het einde van hoofdstuk 4 een tabel opgenomen waarin dit allemaal nog eens is samengevat.

In hoofdstuk V is een concept ontwerp van de besluitvormingsprocedure uitgewerkt. Uitgangspunt is hierbij de afwegingsmethode. Deze is samengesteld uit de beste onderdelen van reeds bestaande methoden. Het hoofdstuk sluit af met een beoordeling van het ontwerp vanuit de opgestelde kwaliteitscriteria.

Inhoudsopgave

Voorwoord	i
Samenvatting	ii
Inhoudsopgave	iii
<u>HOOFDSTUK I INLEIDING</u>	1
§ I.1 Projectcontext.....	1
§ I.2 De opdracht	2
§ I.3 Probleemstelling	2
§ I.4 De modelbenadering	2
§ I.5 De afwegingsmethode als model van het keuzeprocess	3
§ I.6 De onderzoeksvragen.....	6
§ I.7 De opzet van het verslag.....	6
<u>HOOFDSTUK II AFWEGINGSMETHODEN</u>	7
§ II.1 Inleiding.....	7
§II.1.1 Gebruik van afwegingsmethode	8
§II.1.2 Het model van het keuzeprocess.....	9
§ II.2 Algemene beschrijving, functie en werking van de methoden.....	10
§II.2.1 Typen modellen en hun specifieke kwaliteitscriteria.....	10
§II.2.2 Waar houdt het model op en begint de beslisser?	11
§II.2.3 Integratie	12
§ II.2.3.1 Integratie van verschillende variabelen.....	12
§ II.2.3.2 Aggregatie.....	12
§ II.2.3.3 Integratie van verschillende modellen.....	13
§II.2.4 Vooronderstellingen die aan de modellen ten grondslag liggen.	14
§ II.2.4.1 Het aantal alternatieven is beperkt.	15
§ II.2.4.2 Er is een beperkt aantal criteria.....	16
§ II.2.4.3 Het is mogelijk om uitspraken te doen over de mate waarin alternatieven aan de criteria voldoen.	17
§ II.2.4.4 Er is één beslisser.....	17
§ II.2.4.5 Beslissen is een statisch proces.....	18
§ II.2.4.6 Alles is vergelijkbaar	18
§II.2.5 De functie van gemodelleerde kennis.....	19
§ II.2.5.1 Instrumenteel gebruik van modellen.....	19
§ II.2.5.2 Conceptueel gebruik van modellen.....	19
§ II.2.5.3 Strategisch gebruik van modellen.	20
§ II.2.5.4 Het hanteerbaar maken van onzekerheid.....	20

§ II.3 Specifieke vooronderstellingen en scripts per afwegingsmethode:	21
§II.3.1 Monetaire evaluatiemethoden.....	21
§ II.3.1.1 Geld is het enige criterium of alle andere criteria zijn in geld uit te drukken.	21
§ II.3.1.2 Al het geld is evenveel waard	25
§II.3.2 Presentatiemethoden.....	25
§II.3.3 Multicriteria methoden	26
§ II.3.3.1 Weegfactoren – manieren om ze vast te stellen	26
§ II.3.3.2 Omrekenen van kwantitatieve naar kwalitatieve criteria	28
§ II.3.3.3 Standaardisatie.....	29
§ II.3.3.4 Linearisatie	32
§ II.3.3.5 Rangordening	32
§ II.3.3.6 Paarsgewijze vergelijking	33
§ II.3.3.7 Black-box berekeningsmethoden.....	33
§ II.3.3.8 Overeenkomsten tussen kostenbaten analyse en gewogen somming	33
§ II.4 Conclusies	34
<u>HOOFDSTUK III KWALITEITSCRITERIA VOOR AFWEGINGSMETHODEN</u>	37
§ III.1 Wat is modelkwaliteit?	37
§ III.2 Modelkwaliteit volgens Covello	38
§III.2.1 Interne modelkwaliteit.....	38
§III.2.2 Extern modelkwaliteit	39
§ III.3 De betekenis van modelkwaliteitscriteria voor afwegingsmethoden	40
§III.3.1 Interne criteria	40
§ III.3.1.1 Logica.....	40
§ III.3.1.2 Compleetheid	41
§ III.3.1.3 Accuratesse	43
§III.3.2 Externe criteria	44
§ III.3.2.1 Acceptatie.....	44
§ III.3.2.2 Praktische bruikbaarheid	45
§ III.3.2.3 Effectiviteit.....	46
§ III.4 Conclusies en aanbevelingen	48
<u>HOOFDSTUK IV DE TOETSING VAN DE AFWEGINGSMETHODEN</u>	51
§ IV.1 Beoordeling per criterium	51
§IV.1.1 De argumentatie	51
§IV.1.2 De realisatie van de theorie in het model	53
§IV.1.3 Vooronderstellingen en Scripts	54
§IV.1.4 De structuur van de afwegingsmethode.....	57
§IV.1.5 Alle soorten criteria moeten vergeleken kunnen worden.	58
§IV.1.6 Flexibiliteit ten aanzien van de input.	59
§IV.1.7 De betrouwbaarheid van het model.....	60
§ IV.1.7.1 Maatregel- en criteriumonzekerheid	60
§ IV.1.7.2 Effectonzekerheid	60
§ IV.1.7.3 gewichtonzekerheid	61
§ IV.1.7.4 Methodegevoeligheid.....	63
§ IV.1.7.5 Interpretatieonzekerheid.....	63
§IV.1.8 De mate van informatieverlies.	64
§IV.1.9 Het voldoen aan functies.....	65
§IV.1.10 De toevoeging van een onzekerheidsanalyse.	65

§ IV.2 Bespreking van de kwaliteit van elke methode	66
§IV.2.1 Monetaire Evaluatiemethoden	66
§ IV.2.1.1 Kostenbaten analyse	66
§ IV.2.1.2 Kosteneffectiviteits analyse	67
§IV.2.2 Presentatiemethoden	67
§ IV.2.2.1 Scorekaartmethode	67
§ IV.2.2.2 Community impact analyse	68
§ IV.2.2.3 AMOEBE.....	68
§IV.2.3 Multicriteria analyse	68
§ IV.2.3.1 Gewogen Sommering	68
§ IV.2.3.2 Verwachtingswaarde methode.....	69
§ IV.2.3.3 Concordantie analyse	70
§ IV.2.3.4 Evamix methode.....	70
§ IV.2.3.5 Regimemethode.....	70
§ IV.2.3.6 Permutatiemethode.....	71
§ IV.2.3.7 Meerdimensionale schaalanalyse.....	71
§ IV.2.3.8 Goals achievement matrix methode.....	71
§ IV.2.3.9 SAATY-methode.....	72
§ IV.3 Samenvatting en conclusies hoofdstuk IV	72
<u>HOOFDSTUK V CONCEPT ONTWERP VAN EEN AFWEGINGSMETHODE</u>	75
§ V.1 Inleiding	75
§ V.2 Het ontwerp	75
§V.2.1 Voorbereidend werk	75
§V.2.2 Criteria	76
§V.2.3 Alternatieven.....	77
§V.2.4 Impact analyse en voorbereidend werk.....	77
§V.2.5 Invullen afwegingsmodel	78
§V.2.6 Gevoeligheids- en onzekerheidsanalyse.....	79
§V.2.7 Tabel van herkomst.....	80
§ V.3 Bespreking van de methode	80
Lijst van gebruikte literatuur	83
BIJLAGEN	84
Bijlage I, de onderzochte methoden	84
Bijlage 2, de functies die afwegingsmethoden zouden kunnen vervullen	85
Bijlage 3, de concordantie analyse, de proef op de som.	86
Bijlage 4, de regimemethode, de proef op de som	88
Bijlage 5, de input van rijkswaterstaat	89
Bijlage 6, een eenvoudige determinatietabel om de meest geschikte methode te selecteren	90
Bijlage 7, een functionele decompositie van afwegingsmethoden	91

Hoofdstuk I Inleiding

§ 1.1 Projectcontext

Dit afstudeerverslag is gemaakt in opdracht van Rijkswaterstaat. Deze opdracht is als volgt geformuleerd:

Projectcontext

De Waddenzee is een uniek natuurgebied van internationale betekenis. Het is tevens een omgeving waarin mensen wonen, werken en recreëren. De overheid is verantwoordelijk voor het beheer van dit gebied. Dit betekent ondermeer het beschermen en herstellen van de natuurwaarden en het reguleren van de menselijke activiteiten die op of aan de Waddenzee plaatsvinden.

Bij het beleid en beheer van de Waddenzee zijn naast de overheid ook verschillende andere partijen betrokken: het bedrijfsleven, milieu-organisaties en belangengroeperingen. Elke betrokken partij heeft zijn eigen belangen, verantwoordelijkheden en kennis. Voor een goed beheer van de Waddenzee is het belangrijk deze verschillende en soms tegenstrijdige belangen tegen elkaar af te wegen op basis van de best beschikbare kennis en in overleg met alle betrokken partners. Voor dit proces wordt de term integraal beheer gebruikt. Het gaat hierbij om het beoordelen van de effecten van menselijk gebruik en het inpassen van het gebruik binnen de natuurfunctie van de Waddenzee, zodanig dat de natuur, maar ook de economie, zich duurzaam kunnen ontwikkelen.

Projectkader

Binnen het Rijkswaterstaat-kader van het Thema Gebruik (Inpassing en effecten van gebruiksfuncties in de Waddenzee) vindt ontwikkeling plaats van instrumenten die tot doel hebben: de ondersteuning van te nemen beslissingen bij gewijzigd beleid. Een van de mogelijkheden is het toepassen van wegingmethodieken. Het afstudeerproject "Wegingmethodieken in het Waddenzeebeleidsveld" richt zich op het vinden van een rationele en inzichtelijke methode om verschillende belangen in de Waddenzee tegen elkaar af te wegen. Daarbij wordt aandacht gegeven aan de verschillende niveaus (en schalen) waarop de belangen spelen, en aan het vertalen van belangen van het ene niveau naar het andere.

Afstudeerproject Universiteit Twente

Het project is uitgevoerd als afstudeeropdracht bij de afdeling Modelling van Integrale Civiel-technische Systemen (MICS) van de Universiteit Twente door student Rik de Roode. De begeleidingscommissie van deze afstudeeropdracht bestaat uit prof. dr. A. van der Veen, ir. M.J. Kolkman en prof. dr. V.N. de Jonge.

Het resultaat van het project bestaat uit:

1. Een overzicht van bestaande wegingmethodieken.
2. Een overzicht van criteria waaraan afwegingsmethodieken dienen te voldoen.
3. Een analyse van de toepasbaarheid van deze methodieken ter ondersteuning van beleidsbeslissingen in het Waddengebied alsmede de voor- en nadelen van de verschillende methoden.
4. Een conceptueel ontwerp voor een integrale afwegingsmethodiek

Een presentatie van de voorlopige resultaten van dit project heeft plaatsgevonden op de bijeenkomst van Thema Gebruik d.d. 8 juni 1998 te Leeuwarden. De presentatie van de uiteindelijke resultaten van dit afstudeeronderzoek zal plaats vinden op 3 november a.s. te Haren.

§ 1.2 De opdracht

Er bestaan goede inventarisaties van afwegingsmethoden. Veel tijd steken in het vinden van nog meer methoden lijkt mij niet erg nuttig. Er is ook al veel onderzoek gedaan op het gebied van integrale afweging. Er zijn in de loop der tijd een groot aantal methoden ontwikkeld en verder verfijnd. Het is de vraag wat de meerwaarde zou zijn van nog een methode. Daarom zal ik in dit afstudeeronderzoek niet een nieuwe inventarisatie maken van bestaande methoden, ook zal ik niet een geheel nieuwe methode ontwerpen.

De boeken en artikelen die ik gelezen heb over afwegingsmethoden beschrijven deze methoden wel goed, maar ze helpen niet of nauwelijks bij de keuze van een afwegingsmethode. Er wordt niet gezegd welke methoden goed en welke slecht zijn, eventueel afhankelijk van het doel waarvoor de methode gebruikt wordt. Veel eerder wordt er gezegd dat alle methoden hun voor- en nadelen hebben, waarna een korte beschrijving van deze voor- en nadelen volgt. Er valt echter veel meer over de kwaliteit van deze methoden te zeggen. Dat zal dan ook het hoofdbestanddeel zijn van deze opdracht: uitspraken doen over de kwaliteit van een aantal bestaande methoden. Ik zal dat doen met behulp van een “modelbenadering”.

§ 1.3 Probleemstelling

De probleemstelling van deze afstudeeropdracht wordt daarmee: welk kwaliteitsoordeel kan er uitgesproken worden over verschillende bestaande afwegingsmethoden?

§ 1.4 De modelbenadering

Het moeilijke van de beoordeling van de kwaliteit van afwegingsmethoden is dat het nauwelijks mogelijk is om het begrip kwaliteit om te zetten in een operationele norm waaraan de methode getoetst kan worden. De meest voor de hand liggende kwaliteitseis zou zijn dat de resultaten van de afwegingsmethoden “goed” zouden moeten zijn, dat wil zeggen dat als de afwegingsmethode en bepaalde keuze als de beste aangeeft, dan zou dit ook de beste keuze moeten zijn.

Het probleem is echter dat de afwegingsmethode gebruikt wordt in die situaties waarin het onmogelijk is om vast te stellen wat de beste keus is. De afwegingsmethode komt wel tot een resultaat, maar dit is niet af te zetten tegen een meting, omdat niet te meten is wat het goede resultaat is, ook niet achteraf. Daarmee is het dus ook onmogelijk om vast te stellen of de afwegingsmethode de juiste keus heeft voorgesteld.

Bij de afstudeerrichting Modelleren van Integrale Civieltechnische Systemen (MICS) van de opleiding Civiele Technologie en Management (CT&M) wordt er veel aandacht besteedt aan dit soort problemen. Zeker bij complexere modellen is het heel gewoon dat de kwaliteit niet direct getoetst kan worden aan meetgegevens. Daarom is er gezocht naar andere manieren om de kwaliteit van een model vast te stellen. Door nu een afwegingsmethode te beschouwen als een model van het keuzeprocess, is het mogelijk om de bij MICS gebruikte criteria voor modelkwaliteit toe te passen op afwegingsmethoden.

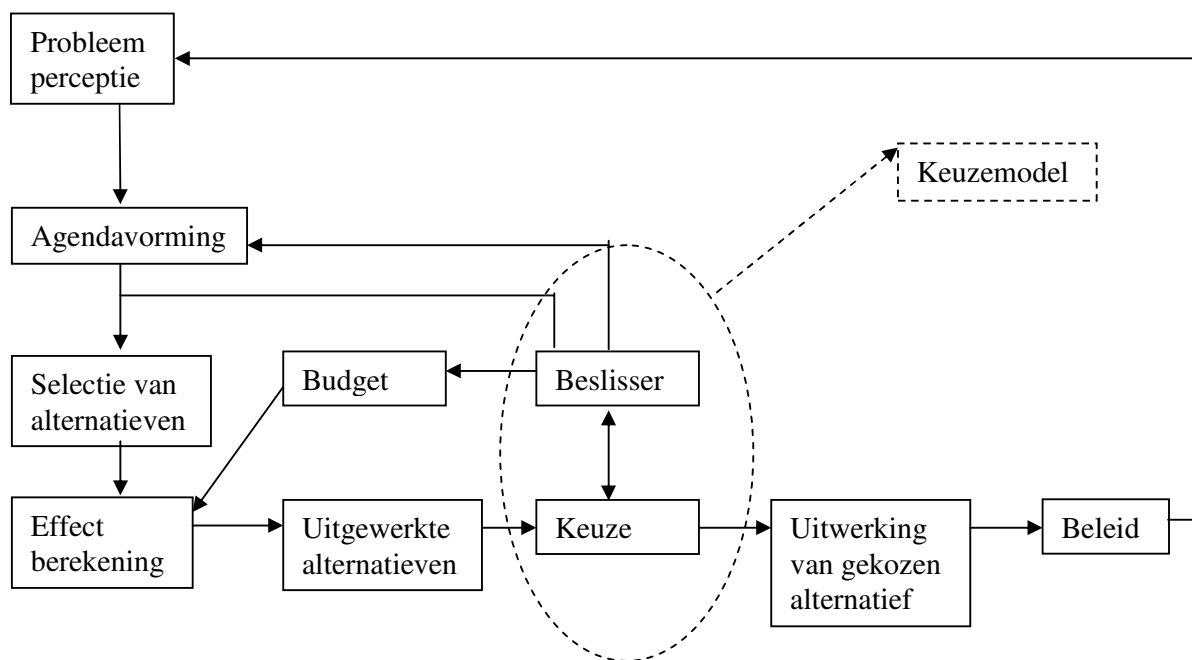
§ 1.5 De afwegingsmethode als model van het keuzeprocess

Om afwegingsmethoden te kunnen zien als een model van het keuzeprocess is het eerst nodig om dat keuzeprocess nader te bekijken. Met dat doel is het schema op de volgende pagina (Figuur 1) opgenomen. Dit is niet bedoeld als een strak passend schema van het besluitvormingsprocess, maar het geeft wel een indicatie van de positie van het keuzeprocess zoals dat verondersteld wordt in deze modelbenadering. Het is van belang om te zien dat het keuzeprocess niet op zich zelf staat.

Het gaat bij afwegingsmethoden vaak (niet noodzakelijk) pas om de laatste fase in de besluitvorming. Er wordt verondersteld dat er al alternatieven zijn gegenereerd en dat er al bekend is wat deze alternatieven waarschijnlijk zullen uitwerken. Verder wordt de afwegingsmethode meestal gevuld in interactie met de beslisser. Bij gebruik van een afwegingsmethode is de invulling van de methode en de interpretatie en het gebruik van de resultaten in feite het keuzeprocess.

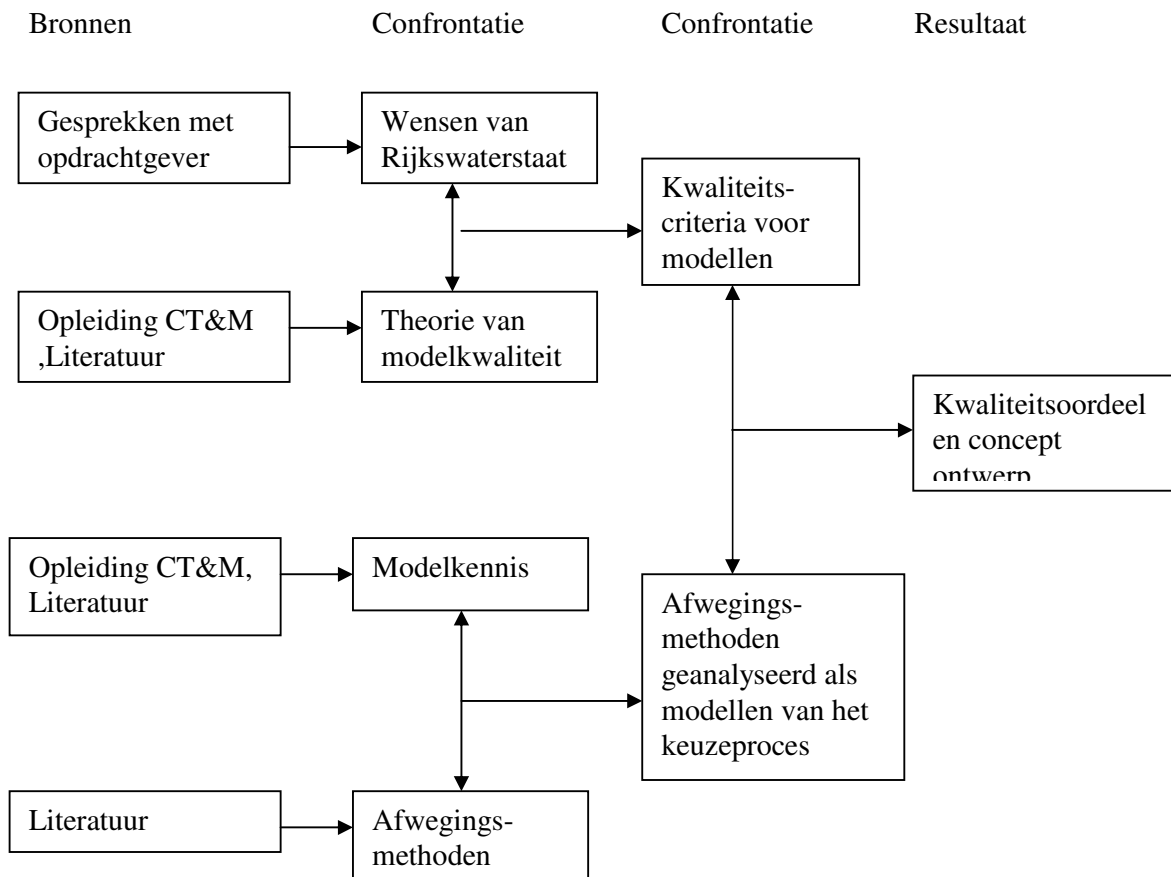
Met: “een afwegingsmethode beschouwen als een model van het keuzeprocess” wordt niet bedoeld dat de afwegingsmethode de praktijk van het beslissen zo goed mogelijk modelleert. Het gaat er veel eerder om dat de methode voorschrijft hoe de beslissing genomen zou moeten worden.

Het tweede schema op de volgende pagina beschrijft het onderzoek zoals dat in het kader van deze afstudeeropdracht is uitgevoerd. Op de pagina daarna zal dit in woorden uitgewerkt worden.



Figuur I: De positie van het model van het keuzeproces in de besluitvorming.

Het startpunt in deze beschrijving is de probleemperceptie. Iemand, een groep mensen of de beslisser zelf ondervindt een probleem. Een probleem wordt daarbij gedefinieerd als: “Het verschil tussen een gewenste en de bestaande of verwachte situatie.” Dit probleem moet vervolgens op de politieke agenda komen, wil het behandeld worden. De beslisser beïnvloedt waarschijnlijk deze agenda. Komt het probleem op de agenda, dan kan er naar oplossingen gezocht worden. Van de mogelijke oplossingen moeten de effecten worden berekend. Hoe goed deze effectberekening is, is mede afhankelijk van het budget dat de onderzoekers krijgen. Het resultaat is een verzameling uitgewerkte alternatieven. Hieruit moet een keuze worden gemaakt. Dit proces, waarin de beslisser uit mogelijke alternatieven er één kiest om mee verder te gaan is het keuzeproces. Daarna kan het alternatief verder uitgewerkt worden en uiteindelijk worden uitgevoerd in beleid. De beslisser heeft direct en indirect ook invloed op andere delen van het proces, zoals te zien is in de figuur.



Figuur II: Het onderzoeksmodel

Door afwegingsmethoden te zien als modellen van het keuzeprocés van de beslisser, kunnen afwegingsmethoden, met behulp van modelkwaliteitscriteria beoordeeld worden. Met behulp van deze modelkwaliteitscriteria kan vervolgens een concept ontwerp van een afwegingsmethode voor rijkswaterstaat opgesteld worden. De modelkwaliteitscriteria worden opgesteld door algemene modelkwaliteitscriteria te specificeren met behulp van het eisen- en wensenpakket van Rijkswaterstaat. Dat pakket van eisen en wensen wil ik inventariseren met behulp van interviews met de opdrachtgever. Algemene modelkwaliteitscriteria haal ik uit literatuur van CT&M. Met behulp van literatuur en internet wil ik een inventarisatie maken van de verschillende afwegingsmethoden die ontwikkeld zijn. Deze afwegingsmethoden zal ik vervolgens interpreteren en beschrijven als een model van het keuzeprocés, zodat ze beoordeeld kunnen worden met modelkwaliteitscriteria.

§ 1.6 De onderzoeksvragen

Welk kwaliteitsoordeel kan er uitgesproken worden over verschillende bestaande afwegingsmethoden?

1. Welke afwegingsmethoden zijn er?
2. Hoe modelleren deze afwegingsmethoden het keuzeproces?
 - 2.1. Welke aspecten van modellen kunnen onderscheiden worden?
 - 2.2. Hoe worden afwegingsmethoden vervolgens op deze aspecten beschreven?
 - 2.3. Wat zijn de overeenkomsten en verschillen tussen de verschillende methoden?
 - 2.4. Welke functies kunnen afwegingsmethoden vervullen?
3. Hoe moet modelkwaliteit beoordeeld worden?
4. Hoe kan dit toegepast worden op afwegingsmethoden?
 - 4.1. Wat verandert er voor deze specifieke groep modellen?
 - 4.2. Wat is de voor afwegingsmethoden bruikbare set criteria?
5. Wat is de modelkwaliteit van de afwegingsmethoden?
 - 5.1. Hoe scoren de modelonderdelen op de opgestelde kwaliteitscriteria?
 - 5.2. Welke methoden vallen af en welke zijn geschikt?
6. Hoe ziet een model waarin de beste eigenschappen van verschillende methoden geïntegreerd zijn er uit?

§ 1.7 De opzet van het verslag

In hoofdstuk II zullen de geselecteerde afwegingsmethoden eerst beschreven worden als een model, met behulp van (Rip, 1996). Dit geeft al een zeer uitgebreide beschrijving van de afwegingsmethoden als model van het keuzeproces. Daarmee zullen de onderzoeksvragen 1 en 2 beantwoord zijn. In hoofdstuk III zullen de modelkwaliteitscriteria van Covello (Kolkman 1996) besproken worden. Daarna worden deze uitgewerkt voor afwegingsmethoden. Vervolgens kunnen er operationele criteria waarop afwegingsmethoden beoordeeld kunnen worden uit worden gedestilleerd. Daarnaast volgen er nog een aantal richtlijnen uit die van belang zijn voor het gebruik van afwegingsmethoden. Daarmee worden de onderzoeksvragen 3 en 4 beantwoord. In hoofdstuk IV wordt vervolgens aan de hand van de operationele criteria de afwegingsmethoden beoordeeld. Het resultaat is een overzichtstabel met daarin alle afwegingsmethoden en hun beoordeling op de kwaliteitscriteria (zie pag. 72). In hoofdstuk V zal tenslotte een methode samengesteld worden uit onderdelen van de geëvalueerde methoden, waarbij een groot aantal van de in de voorgaande hoofdstukken besproken elementen aan bod zullen komen. Daarmee is ook onderzoeksvraag 6 beantwoord.

Hoofdstuk II Afwegingsmethoden

§ II.1 Inleiding

Afwegings- of evaluatiemethoden¹ worden gebruikt om op een rationele manier tot een beslissing te komen. Met rationeel bedoel ik hier dat de manier waarop de beslissing genomen wordt onder dezelfde omstandigheden onafhankelijk van de beslisser altijd tot dezelfde resultaten leidt. Rationeel afwegen heeft tot gevolg dat verantwoording kan worden afgelegd en suggereert een objectieve afweging. De afwegingsmethode zelf leidt als het goed is tot rationele resultaten. Maar verschillende methoden kunnen nog steeds tot verschillende resultaten leiden. De keuze voor de afwegingsmethode zelf zal dus ook verantwoord moeten worden. Om de kwaliteit van afwegingsmethode vast te kunnen stellen zal ik in dit onderzoek gebruik maken van algemene modelkwaliteitscriteria.

Afwegingsmethoden zijn voor te stellen als een model van het keuzeproces van de beslisser. De afwegingsmethode is dan een manier om dit keuzeproces te modelleren en te structureren. Door de afwegingsmethode te beschrijven met behulp van modelkenmerken uit (Rip, 1996) ontstaat een breed inzicht in de kenmerken en mogelijkheden van afwegingsmethoden. Bovendien biedt dit aangrijpingspunten voor modelkwaliteitscriteria. Deze zullen betrokken worden uit de algemene modelkwaliteitscriteria van Covello (Kolkman 1996)

Evaluatiemethoden kunnen ingedeeld worden in continue en discrete methoden. Continue methoden optimaliseren, ze zoeken zelf binnen opgelegde randvoorwaarden de beste mogelijkheid uit. Discrete methoden kiezen uit een beperkt aantal voorgestelde opties. Dit onderzoek richt zich op evaluatiemethoden die geschikt zijn voor de afweging van menselijk gebruik en natuur in de Waddenzee. Vanwege de grootte, de complexiteit en de geringe controle van de mens over dit watersysteem, is het niet mogelijk om continue optimalisatiemethoden te gebruiken ter evaluatie. De systeemkennis die daarvoor nodig zou zijn is eenvoudig niet beschikbaar. Daarom beperk ik mij tot discrete afwegingsmethoden.

In (Ministerie van financiën, 1992) is een inventarisatie gemaakt van bestaande discrete afwegingsmethoden. Deze inventarisatie vormt de feitenbasis van mijn verslag. Daarnaast heb ik gebruik gemaakt van het onderzoek van Ivan Pouwels naar evaluatiemethoden (Pouwels 1996) en van materiaal van de faculteit bestuurskunde over de SAATY-methode², die ook in (Ministerie van Financiën 1996) kort genoemd wordt. Het is niet mijn bedoeling de beschrijvingen van de methoden door het ministerie van financiën en het onderzoek van Ivan Pouwels hier te herhalen. Beiden geven een duidelijke presentatie van de inhoud van de verschillende afwegingsmethoden.

¹ Beide termen zullen in dit onderzoek door elkaar gebruikt worden.

² Zie ook bijlage ****

§II.1.1 Gebruik van afwegingsmethode

Om het nu volgende verslag goed te kunnen begrijpen moet de lezer in ieder geval een rudimentair inzicht in het gebruik en de functie van afwegingsmethoden hebben. Voor de technische details verwijs ik naar (Ministerie van financiën 1992, en Pouwels 1996). Erg interessant in verband met de positie van de afwegingsmethode in het besluitvormingsproces is de analyse van (Nieuwkamer 1995, hoofdstuk 2), de achtergronden die hij beschrijft komen overeen met de vooronderstellingen die ik impliciet voor dit verslag heb gemaakt. Met name de eerste vier paragrafen van hoofdstuk 2 zijn interessant om afwegingsmethoden in het besluitvormingsproces te kunnen plaatsen. In hoofdstuk II.7 worden nog een aantal goede opmerkingen over afwegingsmethoden gemaakt, die overigens inhoudelijk ook in dit verslag aan bod zullen komen.

Hier wil ik alleen nog een aantal woorden herdefiniëren. Dit omdat ik ervaren heb dat een aantal termen uit afwegingsmethoden bij Rijkswaterstaat een heel andere betekenis hebben.

Een afwegingsmethode maakt altijd gebruik van alternatieven, criteria en scores. Zoals in (Nieuwkamer 1995) te lezen is wordt er verondersteld dat, voordat de afwegingsmethode gebruikt wordt, er een besluitvormingstraject is afgelegd, waarbij in ieder geval een probleemstelling is geformuleerd. Deze wordt gevolgd door een doelstelling en een set criteria waaraan de doelstelling getoetst kan worden. Criteria zijn dus grootheden waarmee de mate van doelbereiking, uitgedrukt in goed gedefinieerde eenheden, getoetst kan worden. Daarna zijn er een aantal alternatieve oplossingen uitgewerkt, de alternatieven. Hierbij mag het nul-alternatief (voortzetten van de huidige situatie) nooit ontbreken, alleen al vanwege het referentiekader dat dit biedt. Van deze oplossingen worden voor alle criteria de verwachte resultaten voorspeld. Deze voorspellingen kunnen gebaseerd zijn op modellen, expert-opinion, enquêtes of gewoon natte vinger werk. Deze verwachte resultaten noem ik scores. Voor elk alternatief moet er dus een score ‘berekend’ worden op elk criterium.

Deze criteria, alternatieven en scores vormen vervolgens de input van de afwegingsmethode. Deze volgorde hoeft overigens niet strikt gehandhaafd te worden, het is zeer goed mogelijk, zelfs aanbevelenswaardig, om iteratief te werk te gaan. Bij het maken van een voorspelling kan bijvoorbeeld blijken dat de eerder gestelde criteria niet operationeel genoeg zijn en daarom niet goed bruikbaar voor een voorspelling.³ Dan is het verstandig om deze criteria wel operationeel te maken en met deze nieuwe criteria de voorspelling af te maken.

Technisch gezien lijken de afwegingsmethoden erg veel op elkaar. Om een beeld te geven van de opbouw van afwegingsmethoden volgt nu een opsomming van de onderdelen

³ Denk bijvoorbeeld aan het criterium kwaliteit van de natuur. Deze is niet direct meetbaar. Wordt de kwaliteit echter benaderd met behulp van meetbare parameters als: overlevingskansen van zeehonden in hun eerste levensjaar, concentratie organotin in het water of schattingen van experts over de kwaliteit van de natuur, dan kan er wel gemeten of voorspeld worden.

waaruit de meeste afwegingsmethoden bestaan. De volgende onderdelen komen in de methoden voor, tussen haakjes een indicatie van het voorkomen van een onderdeel:

- Criteria (altijd)
- Alternatieven (Altijd)
- Scores van alternatieven op criteria (altijd)
- Omrekening van scores naar meer bruikbare scores (soms)
- Standaardisatie van de scores (soms)
- Indeling van scores per actor (soms)
- Vaststellen van weegfactoren (vaak)
- Berekening van weegfactoren (soms)
- Berekening van een voorkeursvolgorde in de alternatieven (vaak)
- Presentatie van de scores (soms)
- Presentatie van de berekening (soms)
- Presentatie van de voorkeursvolgorde (vaak)

Bij een aantal methoden zijn onderdelen identiek. Sommige methoden zijn niet meer dan een verbetering van een andere methode. Daarom kunnen onderdelen vaak tegelijkertijd besproken worden voor meerdere methoden. De overeenkomst tussen verschillende methoden kan ook gebruikt worden om de beste onderdelen van verschillende methoden samen te nemen om daarmee een nieuwe methode samen te stellen. Dit gebeurt in hoofdstuk V. Tot slot moeten nog twee afkortingen genoemd worden die in dit verslag veelvuldig zullen worden gebruikt. KBA voor KostenBaten Analyse(s) en MCA voor MultiCriteria Analyse(s).

§II.1.2 Het model van het keuzeprocess

Om een afwegingsmethode als model te beschrijven is het nuttig eerst naar de verschillende kenmerken van modellen te kijken. (Rip, 1996) onderscheidt een aantal modeleigenschappen die mijns inziens zeker beschreven moeten worden om een compleet beeld van afwegingsmethoden als model te kunnen geven. Deze eigenschappen zijn achtereenvolgens: de aard gekoppeld aan de functie van het model, de reikwijdte van het model, de integratie in het model, de ‘scripts’ of de ‘geography of responsibilities’ en de functie en bruikbaarheid van de gemodelleerde kennis.⁴ Dit vormt paragraaf II.2. In paragraaf II.3 wordt verder ingegaan op specifieke ‘scripts’ per afwegingsmethode. In paragraaf II.4 volgt een samenvatting en conclusies.

⁴ De betekenis van de termen zullen in de betreffende paragrafen worden toegelicht.

§ II.2 Algemene beschrijving, functie en werking van de methoden.

Discrete afwegings- of evaluatiemethoden worden in (Pouwels 1995 en Ministerie van Financiën 1992) opgedeeld in drie soorten methoden, de monetaire methoden, de presentatiemethoden en de multicriteria methoden. Alle methoden gaan ervan uit dat er een keuze gemaakt moet worden uit een aantal alternatieven. Deze keuze moet gemaakt worden door de alternatieven te toetsen aan criteria. Monetaire en multicriteria methoden zoeken daarbij naar een voorkeursvolgorde in de alternatieven. Presentatiemethoden geven alleen meer inzicht in de te maken keuze, een neveneffect dat ook de monetaire en de meeste multicriteria methoden hebben.

Evaluatiemethoden kunnen op twee manieren als model beschouwd worden. In het eerste geval wordt de methode beschouwd als een model dat voorschrijft hoe een beslissing genomen dient te worden. In het tweede geval is gemodelleerd hoe en waarom een beslissing genomen is. Voor de keuze van een afwegingsmethode is de eerste vorm primair van belang, bij het gebruik van afwegingsmethoden de tweede natuurlijk ook.

§II.2.1 Typen modellen en hun specifieke kwaliteitscriteria

(Rip. 1996) onderscheidt drie archetypische modellen, namelijk: het model als representatie, het model als microkosmos en het model als voorbeeld. Deze drie modelsoorten zijn uitersten, de meeste modellen zullen in meer of mindere mate kenmerken hebben van deze drie modellen. Het representatiemodel richt zich op een zo goed mogelijke representatie van de gemodelleerde werkelijkheid. Natuurlijk vinden er vereenvoudigingen plaats, maar de relevante aspecten moeten zo goed mogelijk de werkelijkheid representeren. Een model als microkosmos is bedoeld als speeltuin. De concrete vormgeving van het model is erg belangrijk, omdat door een goede vormgeving naar hartelust met het model geëxperimenteerd kan worden. De relevantie van de modeluitkomsten zijn sterk afhankelijk van de mate van representativiteit. Tot slot, het model als voorbeeld is vooral bedoeld als leidraad. Te denken valt aan tekenmodellen, rolmodellen en normen en waarden (modellen van ethisch juist en onjuist handelen).

Nu zijn bijna alle afwegingsmethoden vooral voorbeeldmodellen, die in feite aangeven hoe een afweging gemaakt zou moeten worden. Daarnaast veronderstellen ze impliciet een bepaalde visie op de manier waarop het hele keuzeproces plaats zou moeten vinden. Het belangrijkste kwaliteitscriterium volgens Rip (Rip 1996) is de relevantie van het model voor de beoogde actie en de compleetheid van het model ten aanzien van de beoogde actie.

Echter, als het afwegingsmodel gebruikt en ingevuld is, is er sprake van een representatie- en een speeltuinfunctie. Het ingevulde afwegingsmodel geeft weer welke criteria de beslisser belangrijk vindt en soms ook hoe belangrijk hij die vindt. Daarnaast wordt in de meeste afwegingsmethoden ook nog voorspeld of gemodelleerd in welke volgorde de beslisser de verschillende beleidsopties waardeert. Het model representeert dus eigenlijk de voorkeuren

en het tot stand komen van de voorkeuren van de beslisser, of in ieder geval hoe deze voorkeuren tot stand hadden moeten komen.

Daarnaast biedt de ingevulde afwegingsmethode de mogelijkheid aan beleidsmedewerkers om nieuwe beleidsalternatieven te ontwerpen die de beslisser naar verwachting nog beter zal waarderen. Bovendien kan de beslisser ook zelf, door het manipuleren van scores, weegfactoren en andere modelinput, meer leren over zijn eigen preferenties en voorkeursvolgorde. (Hoe sterk is mijn voorkeur, wat als alternatief B toch duurder uitvalt dan geraamd?)

Omdat de ingevulde methodes dus een representatieve en een speeltuinfunctie hebben, moeten de afwegingsmethoden ook gecontroleerd worden op de kwaliteitseisen voor deze modelsoorten. Voor representatiemodellen is de structuur van het model van groot belang. De meeste afwegingsmodellen gaan er bijvoorbeeld van uit dat de voorkeur van de beslisser afhankelijk is van criteria en van het gewicht dat aan die criteria wordt toegekend. Maar een tweelagen structuur waarbij criteria en gewichten per belangengroep worden gemodelleerd, waarna deze belangengroepen en de pressie die zij kunnen uitoefenen op de beslisser de belangrijkste input voor de voorkeursvolgorde zijn, zou misschien wel voor een veel betere representatie kunnen zorgen. Een tweede belangrijk punt is het aggregatieniveau⁵ van het model en hoe modelonderdelen van verschillende aggregatieniveaus met elkaar communiceren.

De mikrokosmos-modellen moeten vooral afgerekend worden op de vraag of er interessante spelen gespeeld kunnen worden in de speeltuin. Daarnaast is de mate van representatie hier ook van belang, tenminste, zodra de gebruiker de kennis opgedaan door het spelen ook daadwerkelijk wil gaan gebruiken.

Samenvattend kan gesteld worden dat afwegingsmethoden deel hebben aan alledrie de modelsoorten, zei het op verschillende niveaus. Welke modelsoort de nadruk gaat krijgen is sterk afhankelijk op welk niveau de afwegingsmethode beoordeeld moet worden en dit zal weer grotendeels afhankelijk zijn van de functies die de methode moet vervullen.

§II.2.2 Waar houdt het model op en begint de beslisser?

Rip onderscheidt in (Rip 1996) nog een vierde dimensie, namelijk de reikwijdte van het model. Het gaat hier om de vraag: wat wordt er allemaal gemodelleerd? Als het onderscheid gemaakt wordt tussen modelbouwers, opdrachtgevers, modelgebruikers en de rest van de werkelijkheid, dan wordt normaal gesproken alleen een deel van de rest van de werkelijkheid gemodelleerd. Omdat gebruikers van modellen, zeker als ze deze modellen gebruiken om de werkelijkheid te manipuleren, ook invloed hebben op de werkelijkheid, kan het interessant zijn om deze gebruiker mee te modelleren. De reikwijdte van het model neemt daarmee enorm toe. Tegelijkertijd wordt hiermee de gebruiker buitenspel gezet. Het model voorspelt wat hij gaat

⁵ Aggregatie is het samennemen van details in één overkoepelend begrip. Een model met een hoger aggregatieniveau is minder gedetailleerd en heeft waarschijnlijk een groter bereik. Modeloutput van een lager aggregatieniveau moet eerst geaggregeerd worden voordat het gebruikt kan worden in een model met een hoger aggregatieniveau.

beslissen. Het model heeft zijn input niet meer nodig, de beslisser heeft geen invloed meer op de modelresultaten. Iets dergelijks gebeurt ook in afwegingsmethoden. In feite wordt de beslisser gemodelleerd. Toch is dit niet helemaal hetzelfde. Als de beslisser gemodelleerd is wordt hij niet buitenspel gezet, tenminste dat hoeft niet, het is eerder zo dat de gebruiker van het model niet meer de beslisser is maar een beleidsvoorbereider, of dat de beslisser met behulp van het door het model verkregen inzicht zelf zijn beslissingen neemt. Het model structureert dan alleen de beslissing. De beslisser zou pas echt buitenspel gezet worden op het moment dat ook de reactie van de beslisser op het model gemodelleerd zou zijn.

Vastgesteld kan worden dat afwegingsmethoden het besluitvormingsproces van de beslisser modelleert en tegelijk beïnvloedt.

§II.2.3 Integratie

“Modellen vergen, juist als ze complexe werkelijkheden modelleren, integratie van verschillende soorten inputs.”, aldus (Rip 1996). Integratie speelt op verschillende niveaus. Van belang zijn de integratie van verschillende variabelen, integratie van microgegevens naar een hoger niveau, dit wordt aggregatie genoemd, en de integratie van verschillende modellen. In deze paragraaf worden ze alle drie behandeld.

§ II.2.3.1 Integratie van verschillende variabelen.

Kenmerk van afwegingsmethoden is dat op de één of andere manier appels met peren vergeleken moeten worden. Als dat niet het geval was, zou er geen afwegingsmethode nodig zijn, de vergelijking tussen 6 en 12 appels in snel gemaakt. Elke afwegingsmethode moet een manier zoeken om verschillende criteria te integreren.

Er zijn grofweg drie manieren waarop dat kan. De integratie van de verschillende criteria kan impliciet blijven. Dit gebeurt bij presentatiemethoden. Alle relevante gegevens over de alternatieven worden gepresenteerd, de beslisser moet dan zelf de verschillende criteria integreren. Bij deze presentatiemethoden wordt de eigenlijke afweging dus ook niet expliciet gemaakt, deze vindt ergens in het hoofd van de beslisser plaats. Een tweede manier waarop de verschillende criteria geïntegreerd kunnen worden is die welke in monetaire evaluatiemethoden wordt gebruikt. Hier worden de criteria gereduceerd tot één criterium: de mate van welvaart, uitgedrukt in geld. De derde methode is de expliciete integratie van de verschillende criteria met behulp van standaardisatie en de toekenning van gewichten. Methoden die op deze manier met verschillende criteria omgaan worden multicriteria-analysmethoden of gewoon multicriteria methoden genoemd.

§ II.2.3.2 Aggregatie

De afwegingsmethode moet met kennis gevuld worden. Multicriteria methoden hebben bijvoorbeeld criteria, alternatieven en scores van alternatieven op criteria nodig. Omdat de meeste multicriteria methoden maar een beperkt aantal criteria aankunnen, al was het alleen maar om het overzicht te bewaren, zullen bij grote aantallen criteria de criteria ook samengevoegd moeten worden tot een kleiner aantal criteria met een hoger abstractie niveau. Bij de SAATY-methode wordt deze aggregatie expliciet gemaakt, bij alle andere methoden wordt verondersteld dat het juiste aggregatieniveau is bereikt. Om alle criteria en het juiste aggregatieniveau waarop de criteria weergegeven worden te vinden, kan gebruik gemaakt worden van top-down decompositie en van bottom-up integratie. In het eerste geval wordt het einddoel (bijvoorbeeld: een betere Waddenzee) opgesplitst in subdoelen (een mooiere Waddenzee, een schonere Waddenzee, een winstgevender Waddenzee enz.), in feite zijn dit criteria waaraan het behalen van het einddoel afgemeten wordt. Deze criteria kunnen weer opgesplitst worden in meer operationele, meetbare criteria (directe en indirecte werkgelegenheid ten gevolge van activiteiten in de Waddenzee, bruto productie ten gevolge van activiteiten in de Waddenzee, enz.) Bij een bottom-up benadering worden eerst allerlei criteria gegenereerd, daarna worden deze criteria gegroepeerd tot abstractere criteria op een hoger aggregatieniveau. Natuurlijk kunnen beide methoden ook gecombineerd worden. Ordening van criteria is nuttig, omdat zo ook meer inzicht wordt verkregen in ontbrekende criteria en in dubbeltellingen.

§ II.2.3.3 Integratie van verschillende modellen

Tot slot moet de integratie tussen verschillende deelmodellen nog besproken worden. Ik beschouw hierbij de afwegingsmethode als een onderdeel van het totale besluitvormingsproces. De afwegingsmethode krijgt dus ook input van verschillende andere onderdelen van de besluitvormingsprocedure en levert zelf ook weer output. De waarde van de afwegingsmethode is daarmee ook afhankelijk van de kwaliteit van de andere stappen in de besluitvormingsprocedure. De selectie van criteria bijvoorbeeld, die bij de meeste afwegingsmethoden wordt voorondersteld, moet wel compleet zijn, en bij voorkeur zonder dubbeltellingen. Een slechte selectie van criteria zorgt dus ook voor een slechte afweging. Maar ook de afstemming tussen de selectie en de afwegingsmethode is van belang. Een op zich goede selectie van criteria die slecht omrekenbaar zijn in geld, combineert slecht met een monetaire kostenbaten analyse.

Hoewel beide onderdelen op zich goed zijn, is de integratie van lage kwaliteit. Voor criteria, maar ook voor bijvoorbeeld weegfactoren, is de schaal waarin de criteria of weegfactoren weergegeven worden van belang. (Ministerie van Financiën, 1992) onderscheidt de volgende schalen:

1. Nominaal: woorden.
2. Binaire Schaal: kwalitatief, ja of nee, goed of fout.
3. Ordinaal: rangorde, bijzondere vorm: consumentengids: ++, +, 0, -, --
4. Intervalschaal: relatieve schaal, geen absoluut referentiepunt, voorbeeld: 5 °C
5. Ratioschaal: absolute schaal, absoluut referentiepunt: SI, monetair, percentage.

Vaak wordt onderscheid gemaakt tussen kwantitatieve en kwalitatieve waarden. De eerste drie schalen zijn kwalitatief, de laatste twee kwantitatief. Het is van groot belang dat de schaal waarin scores op criteria of weegfactoren aan de afwegingsmethoden worden aangeleverd overeenstemmen met de schaal die de afwegingsmethode vraagt. Schalen kunnen zonder problemen een schaal lager (lager cijfer) gemaakt worden, er gaat dan echter wel informatie verloren. Schalen kunnen ook opgewaardeerd worden, maar daarvoor moet informatie toegevoegd worden. De eerste schaal is in de meeste methoden nutteloos (uitzondering zijn de presentatiemethoden.) Een criterium als kleur (Nominaal: geel, blauw, rood) is niet onderscheidend, er kan geen voorkeursvolgorde uit gedestilleerd worden, tenzij er meer kennis is over de betekenis van kleuren. Met die kennis kan kleur omgezet worden in een ordinale schaal: rood is erg mooi, dus ++, grijs is saai, dus: -.

De integratie van verschillende modellen (bijvoorbeeld WadBOS en de afwegingsmethode) hoeft niet persé expliciet gemaakt te worden. Gegevens van WadBOS kunnen direct in een afwegingsmodel ingevoerd worden, als bruto toegevoegde waarde een output van WadBOS is en tegelijkertijd een input van een afwegingsmethode, omdat bruto toegevoegde waarde één van de criteria is waarop afgewogen wordt, dan is er sprake van een expliciete integratie. Maar het is ook goed mogelijk om de output van WadBOS te laten interpreteren door een gebruiker, en de interpretatie van deze gebruiker te gebruiken als input voor de afwegingsmethode. Deze impliciete integratie van modellen heeft als voordeel dat de gebruiker inzicht houdt in de betrouwbaarheid van de afweging, het is nog mogelijk om de uitkomsten van WadBOS te interpreteren voordat ze verwerkt worden door de afwegingsmethode, waarna ze onherkenbaar zijn. Het voordeel van een expliciete integratie is weer dat het tijd scheelt en dat de gebruiker niet over uitgebreide systeemkennis hoeft te beschikken, waarmee de modeluitkomsten geïnterpreteerd kunnen worden.

§II.2.4 Vooronderstellingen die aan de modellen ten grondslag liggen.

Een afwegingsmethode staat niet op zichzelf. De ontwerpers van de methode hebben een bepaalde voorstelling gehad van het hele afwegingsproces en hebben daar een methode bij ontworpen. Elke methode heeft daarom impliciet een aantal vooronderstellingen in zich. Vooronderstellingen ten aanzien van de gebruiker van de methode, vooronderstellingen ten aanzien van het totale afwegingsproces en vooronderstellingen over hoe de methode gebruikt zou moeten worden. Het is van belang deze vooronderstellingen te kennen, omdat ze de kwaliteit van een methode kunnen beïnvloeden. Het model kan namelijk zo ontwikkeld zijn dat het de gebruiker dwingt van dezelfde vooronderstellingen uit te gaan. Een eendimensionaal waterstromingsmodel is niet goed bruikbaar voor waterstroming in de Waddenzee, de vooronderstelling van een stroming in voornamelijk één richting wordt dwingend opgelegd aan de gebruiker, een tweede dimensie wordt eenvoudig niet gemodelleerd. Een dergelijke opgelegde beperking noemt Rip een “script”.

Elk model heeft een impliciete “geography of responsibilities”, een verdeling van de taken tussen de gebruiker en de modelleur. Bepaalde zaken in het model zullen wel beïnvloedt

kunnen worden door de gebruiker, andere zullen vastgelegd zijn door de modelleur⁶. Verregaande beïnvloeding van het model door de gebruiker is goed, omdat de gebruiker daarmee de ruimte krijgt het model aan zijn eisen of aan de eisen van de huidige situatie aan te passen. Beperking van de gebruiker is ook goed, omdat dit voorkomt dat de betrouwbaarheid van het model aangetast wordt door ondeskundige verbouwingen door de gebruiker. Om het juiste midden te kunnen vinden moet de modelleur goed op de hoogte zijn van de deskundigheid en van de gewenste aanpassingsruimte van de gebruiker. De vooronderstellingen in het model en de daaruit voortvloeiende “scripts” zouden belangrijke criteria moeten zijn bij de selectie van modellen.

Er zijn een aantal vooronderstellingen die alle of bijna alle afwegingsmethoden maken. Ik noem de volgende:

1. Er is een beperkt aantal alternatieven. (selectie van alternatieven)
2. Er is een beperkt aantal criteria. (selectie van criteria)
3. Het is mogelijk om uitspraken te doen over de mate waarin alternatieven aan de criteria voldoen. (kwantitatieve en kwalitatieve scores)
4. Er is één beslisser.
5. Beslissen is een statisch proces.

Ik bespreek nu elke vooronderstelling apart.

§ II.2.4.1 Het aantal alternatieven is beperkt.

Alle afwegingsmethoden gaan ervan uit dat er een beperkt aantal alternatieven mogelijk is. Een beperkt aantal alternatieven omdat het hier gaat om discrete afwegingsmethoden, zodat elk alternatief apart moet worden berekend. Hier vormen de reken- en onderzoekstijd een beperkende factor. Daarnaast worden overzichtsmethoden verwarrend als er teveel alternatieven zijn, zodat ze aan hun doel voorbij schieten. Deze methoden hebben dus extra baat bij een beperkt aantal alternatieven. Tot slot zijn er nog de permutatiemethode en de methoden die gebruik maken van paarsgewijze vergelijking per criterium. Bij deze methoden neemt het rekenwerk en de tijd die de gebruiker aan de methode kwijt is, meer dan evenredig toe met het aantal alternatieven.⁷

De stelling dat er een beperkt aantal alternatieve beleidsmogelijkheden is in de Waddenzee is een enorme vereenvoudiging van de werkelijkheid. Het is echter wel een heel bruikbare vereenvoudiging. Juist door het beeld van de werkelijkheid zo te vereenvoudigen wordt deze hanteerbaar. Een bijna oneindig aantal alternatieven is nauwelijks beschrijfbaar, laat staan berekenbaar. Een beperkt aantal alternatieven kan wel beschreven en berekend worden.

Deze beperking hoeft niet tot slechtere resultaten te leiden. Een heel duidelijk voorbeeld is te vinden in het schaken. Een schaakcomputer berekent alle mogelijkheden en beoordeelt deze

⁶ Vooropgesteld dat dit twee verschillende personen zijn

⁷ Overigens is dit alleen echt relevant bij handmatige berekening, de huidige generatie pc's vormen deze berekeningen waarschijnlijk geen enkele belemmering.

elk op hun kwaliteit. Het aantal mogelijkheden neemt enorm toe per extra zet die vooruit gerekend wordt. Een mens heeft niet de geheugen- en rekencapaciteit van een computer. Daarom worden niet alle alternatieven doorgerekend, maar alleen de meest veelbelovende. De selectie van alternatieven vindt plaats op basis van intuïtieve herkenning van patronen. Daardoor hoeft slechts een beperkt aantal alternatieven doorgerekend te worden. Deze intuïtie is niet narekenbaar, maar het werkt wel. De kwaliteit van deze intuïtieve voorselectie verschilt per persoon en er kan op getraind worden. Op een soortgelijke manier kan een goede voorselectie van alternatieven tot een aanvaardbare keuze leiden. De afwegingsmethoden veronderstellen dus dat er al een goede voorselectie van alternatieven is gemaakt. Zij doen geen uitspraken over hoe de kwaliteit van die voorselectie gecontroleerd of geforceerd kan worden.

§ II.2.4.2 Er is een beperkt aantal criteria.

Zoals eerder opgemerkt kunnen de afwegingsmethoden met een beperkt aantal criteria omgaan. Hoe beperkt dit aantal is hangt af per methode. Kostenbaten en kostenefficiëntie methoden kennen slechts één criterium: geld. Presentatiemethoden moeten wel met een beperkt aantal criteria werken, omdat anders het overzicht verloren gaat. Overzicht is ook belangrijk om goed weegfactoren toe te kunnen kennen. De SAATY-methode maakt gebruik van paarsgewijze vergelijking van alternatieven om de weegfactoren vast te stellen. Daardoor is overzicht niet meer belangrijk. Maar dit heeft tot gevolg dat het aantal vergelijkingen dat gemaakt moet worden gelijk is aan de faculteit van het aantal criteria gedeeld door twee ($n=c!/2$). Met een toename van het aantal criteria vindt dus een enorme toename van het aantal vergelijkingen die de gebruiker moet maken om de weegfactoren vast te stellen. Deze beperking van het aantal criteria leidt tot de problematiek die ik al in § II.2.3.2, aggregatie heb besproken.

Hier wil ik nog de vraag behandelen of het überhaupt noodzakelijk is om met criteria te werken. Alle afwegingsmethoden gaan ervan uit dat de criteria op de een of andere manier bepalen welke alternatieven beter zijn. Echter een besluit wordt in de praktijk meestal niet genomen omdat het zo goed scoort op criteria, het wordt meestal genomen op basis van argumenten of op basis van intuïtie. Een intuïtief besluit is daarbij gedefinieerd als een besluit dat niet op basis van argumenten genomen wordt. Argumenten en intuïtie zijn dus volgens deze definitie complementair. In tegenstelling tot intuïtie zijn argumenten echter overdraagbaar.

Bij het nemen van besluiten waarover verantwoording moet worden afgelegd, dat wil zeggen, de kwaliteit van het besluit moet getoond worden aan anderen, is communicatie onmisbaar, tenzij de anderen vertrouwen op de intuïtie van de beslisser. De kwaliteit van de besluitvorming bij dit soort besluiten zal dus meestal aangetoond moeten worden met behulp van argumenten. Een keuzeproces is een speciale vorm van een besluitvormingsproces, namelijk één waarbij er gekozen moet worden uit een beperkt aantal mogelijke besluiten. Een model dat het keuzeproces modelleert, kan beperkt worden tot het argumentatieve deel van het besluitvormingsproces, als het model gebruikt wordt ter verantwoording van de

besluitvorming. Voor de voorspelling van de uitkomst van het keuzeproces is het model alleen dan geschikt, als er daadwerkelijk alleen op basis van argumenten besloten wordt, een situatie waarvan ik vermoed dat die weinig voor zal komen.

Argumenten zijn wel makkelijk om te zetten in criteria. Als iemand beweert dat olieboringen verboden moeten worden, omdat ze het mooie uitzicht over de Waddenzee bederven, dan is schoonheid blijkbaar een criterium. Visserij verbieden omdat dit zeegras aantast (argument) geeft als criterium de aantasting van zeegras, of positiever geformuleerd, de levenskansen van zeegras in de Waddenzee. Zo is uit elk argument een criterium te destilleren waar alternatieven op kunnen scoren. De vooronderstelling dat alle besluitvorming op basis van criteria plaatsvindt is niet altijd in overeenkomst met de werkelijkheid, maar als het gaat om het afleggen van verantwoording over een genomen keuze, is de vooronderstelling wel geschikt.

§ II.2.4.3 Het is mogelijk om uitspraken te doen over de mate waarin alternatieven aan de criteria voldoen.

De afwegingsmethoden zoals die beschreven zijn in (Pouwels 1995 en Ministerie van Financiën 1992), beschrijven geen van allen hoe de informatie over de scores van alternatieven op criteria verkregen zou moeten worden. Wel wordt verondersteld dat het mogelijk is om uitspraken te doen over de scores van alternatieven op criteria. Toch zullen deze scores lang niet in alle gevallen te kwantificeren zijn. Een aantal afwegingsmethoden is speciaal ontwikkeld voor deze gevallen en gaan uit van kwalitatieve scores. De scores zijn van grote invloed op de uiteindelijke prioriteitsstelling. Hoe beter de scores de werkelijkheid benaderen, hoe beter het model aan de representatiefunctie zal voldoen. Een betere benadering van de werkelijkheid zal in de meeste gevallen ook meer onderzoek en dus geld en tijd vergen. Daar zal in de keuze voor de afwegingsmethode al rekening mee gehouden moeten worden.

§ II.2.4.4 Er is één beslisser.

Bijna alle multicriteria methoden gaan ervan uit dat er slechts één beslisser is, of in ieder geval een groep beslisser met exact dezelfde mening. Multicriteria methoden zijn in principe modellen waar feiten en meningen gecombineerd worden om tot een uitkomst te komen. De scores van alternatieven op criteria kunnen gemeten worden, in dat geval kunnen ze als feiten gezien worden.⁸ Scores kunnen echter ook de mening van de beslisser weergeven. Gewichten geven altijd de mening van de beslisser weer.⁹ Omdat de multicriteria methoden maar één gewichtenset toelaten modelleren ze dus ook maar één beslisser. Eventueel zou ook een

⁸ Omdat de keuze van criteria en van alternatieven nog steeds gebaseerd zijn op een mening over welke alternatieven kansrijk en welke criteria belangrijk zijn, is een objectieve, puur op feiten gebaseerde afweging met toch al niet mogelijk

⁹ Dat wil zeggen, zouden de mening van de beslisser zo goed mogelijk weer moeten geven, of dat lukt is natuurlijk nog de vraag.

compromis over de gewichten gemodelleerd kunnen worden. De presentatiemethoden kunnen onafhankelijk van een beslisser gemaakt worden, op het moment dat alleen objectieve scores op criteria weergegeven worden. Zelfs dan is de keuze voor criteria en alternatieven subjectief en in principe afhankelijk van welke criteria en alternatieven de beslisser belangrijk respectievelijk kansrijk vindt. De beslisser heeft dus bij alle methoden invloed op het model, zij het dat de grootte van deze invloed varieert per afwegingsmethode. Omdat bijna alle methoden één beslisser vooronderstellen, kunnen ze in een situatie met meerdere beslissers aanzienlijk van de werkelijkheid afwijken. Bij meerdere beslissers zal het model dus op de één of andere manier omgevormd worden zodat het door meerdere beslissers gehanteerd kan worden. In feite gaat het hier weer om een vorm van integratie, waarbij het de vraag is op welk moment de verschillende meningen van de beslissers geïntegreerd gaan worden. Dit valt echter buiten het bereik van dit onderzoek.

§ II.2.4.5 Beslissen is een statisch proces

Geen van de afwegingsmethoden houdt rekening met verandering in voorkeur, verandering van informatie, de invloed van discussies en onderhandelingen of de mogelijkheid van een compromis. Het zijn allen statische modellen, en ze modelleren een nog vast te stellen maximalisatie voor een vage nog vast te stellen eenheid als “het land”, “het bedrijf” of “de doelgroep”. Uitzonderingen hierop zijn de presentatiemethoden, omdat deze alleen informatie leveren en dus in feite geen uitspraken doen over het keuzeprocess zelf. De veronderstelling die de modellen in feite maken is dat criteria, scores en weegfactoren vast te leggen zijn. Eenmaal gemeten zijn zij waar en blijven ze constant. Zeker bij langere besluitvormingsprocedures is dit niet het geval. Juist omdat voorkeuren, informatie, maar ook een doelgroep, kunnen veranderen, is een gemodelleerde voorkeur maar voor een beperkte tijd geldig. Het is de vraag of een ingevuld model bruikbaar is voor nieuwe alternatieven of voor een nieuw probleem, of zelfs voor hetzelfde probleem een paar maanden later. Bij meerdere beslissers zal de geldigheidsduur van een model nog korter zijn, terwijl de mogelijkheid van een compromis nadrukkelijk aanwezig is. De veronderstelling dat er dan iets te berekenen valt aan een maximalisatie voor de doelgroep wordt dan nog twijfelachtiger.

§ II.2.4.6 Alles is vergelijkbaar

Een vooronderstelling die zowel de multicriteria methode als de kostenbaten analyse maken is dat alle criteria met elkaar vergelijkbaar en uitwisselbaar zijn. Hoewel volgens (van der Veen, 199?) multicriteria methoden juist ontwikkeld zijn voor die situaties waarin criteria onvergelijkbaar zijn, blijken de ontworpen methoden juist wel uit te gaan van vergelijkbaarheid. Voor afweging tussen gebruik en natuur in de Waddenzee betekent dit dat het gebruik van een multicriteria methode of een kostenbaten analyse betekent dat gekozen wordt voor een zachte vorm van duurzaamheid.

Er kan onderscheid gemaakt worden tussen zachte en harde duurzaamheid. Bij harde duurzaamheid mag het natuurlijk systeem helemaal niet veranderen. Er zou dan wel

economische activiteit in de Waddenzee plaats mogen vinden, maar de natuur mag helemaal niet aangetast worden. Bij een minder harde vorm van duurzaamheid mag economische activiteit wel natuur aantasten, op voorwaarde dat dit elders gecompenseerd wordt. Bij zachte duurzaamheid mag natuur aangetast worden, mits dit gecompenseerd wordt door een groei van kapitaalgoederen elders, het zij van natuurlijk kapitaal, het zij van economisch kapitaal. In het laatste geval wordt natuur dus ook gezien als een kapitaalgoed waarmee producten voor menselijk gebruik gefabriceerd worden. Door alle criteria vergelijkbaar en inwisselbaar te veronderstellen wordt impliciet gekozen voor zachte duurzaamheid. Voor een verdere uitdieping van de achtergronden van een dergelijke keuze verwijs ik naar (De Roode 1997). In ieder geval zal vooraf expliciet gekozen moeten worden voor een bepaalde duurzaamheidsbenadering. Daarna kan de afwegingsmethode daar aan worden aangepast.

§II.2.5 De functie van gemodelleerde kennis.

(Rip 1996) onderscheidt vier functies waarvoor wetenschappelijke kennis gebruikt kan worden. Deze functies zijn: instrumenteel gebruik, conceptueel gebruik, strategisch gebruik en hanteerbaar maken van onzekerheid. Het onderscheid is kunstmatig, het is niet zo dat wetenschappelijke kennis maar op één manier gebruikt kan worden, maar de indeling is wel nuttig, omdat de bruikbaarheid van een model mede afhangt van welke functies het zou moeten vervullen.

§ II.2.5.1 Instrumenteel gebruik van modellen

Instrumenteel gebruik van modellen wordt vaak als wenselijk beschouwd. Je moet het model gebruiken als middel om bepaalde kennis te genereren. Een meteorologisch model moet je gebruiken om het weer te voorspellen (verwachten), een zeegrasmodel gebruik je om kennis te vergaren over waar zeegras de beste kans maakt om terug te keren in de Waddenzee. Dit is een heel duidelijke manier om modellen te gebruiken. Er is een bepaald duidelijk gedefinieerd probleem, er is een model (of het wordt gemaakt) dat kennis kan genereren om dat probleem op te lossen, dat model wordt ook met dat doel gebruikt. Voor afwegingsmodellen betekent dit dat het model een prioriteitsvolgorde in de aangeboden alternatieven moet opleveren, of dat het meer inzicht in de voorkeur van de beslisser moet geven.

§ II.2.5.2 Conceptueel gebruik van modellen

Conceptueel gebruik van modellen leidt niet tot de oplossing van een probleem maar geeft meer inzicht in de problematiek zelf. Door een bepaald probleem te modelleren ontstaat er meer inzicht in het probleem. Tegelijkertijd worden er concepten en termen ontwikkeld waarmee het probleem communicateerbaarder wordt. Daardoor kunnen ook anderen meer inzicht krijgen in problematiek die ze zelf ervaren. Een duidelijk voorbeeld is de doorsijpeling

van psychologische modellen en termen naar de ‘gewone burger’, die nu ook spreekt van frustratie, schizofrenie of allerlei fobieën, waarmee bepaalde problemen die zij ondergaan of in hun omgeving ervaren benoemd en verklaard kunnen worden. Ook afwegingsmethoden kunnen op deze manier gebruikt worden. Het opstellen van criteria zou bijvoorbeeld kunnen leiden tot meer inzicht in de gaten in de benodigde kennis, daarnaast zou ook meer inzicht verkregen kunnen worden in welke criteria überhaupt van belang zijn. Door een afwegingsmethode op te stellen zou uiteindelijk meer inzicht verkregen kunnen worden in de hele problematiek van de afweging van menselijke belangen en natuur in de Waddenzee.

§ II.2.5.3 Strategisch gebruik van modellen.

De derde manier om modellen te gebruiken is het strategische gebruik. Het model wordt dan gebruikt om de eigen positie te versterken. Uitkomsten die positief ervaren worden krijgen de nadruk, uitkomsten die negatief zijn, worden gediskwalificeerd of gebagatelliseerd. Strategisch gebruik van afwegingsmodellen is zeer goed mogelijk. Juist vanwege het enorme potentieel om de uitkomsten te manipuleren via de keuze van criteria, alternatieven en weegfactoren is het mogelijk om het afwegingsmodel achteraf te gebruiken om een eerder gemaakte keuze te ondersteunen. Zeker als de uitkomsten exact en inzichtelijk weergegeven kunnen worden, met de suggestie van rationaliteit en objectiviteit.

§ II.2.5.4 Het hanteerbaar maken van onzekerheid.

De laatste functie is het hanteerbaar maken van onzekerheid. Voor een deel kunnen modellen onzekerheid wegnemen. Door bestaande kennis over een systeem in een model te gieten komt deze kennis beschikbaar voor de modelgebruiker. Daarnaast kan de onzekerheid die inherent is aan een systeem ook weergegeven worden in een model. Een voorbeeld van inherente onzekerheid is de statistische spreiding van bijvoorbeeld golfhoogtes. Juist met statistische technieken kan deze onzekerheid gemodelleerd worden. Tot slot kan meer inzicht in onzekerheden verkregen worden juist door het gebruik van modellen. Een aantal, maar lang niet alle onzekerheden zijn te reduceren met wetenschap. Juist de onzekerheden die niet reduceerbaar zijn, zijn in dit verband interessant. Door met behulp van modellen de gevolgen van onzekere gebeurtenissen door te rekenen kunnen strategieën uitgetoetst worden zonder dat eventuele desastreuze gevolgen direct in de werkelijkheid plaats vinden. Daarmee worden de onzekerheden niet gereduceerd, ze worden wel hanteerbaar, mocht een bepaalde onzekere gebeurtenis plaats vinden, dan is de beste reactie strategie al bedacht. Voor afwegingsmethoden geldt dat ze onzekerheid wegnemen doordat belangrijke criteria geformuleerd worden. Hierdoor wordt duidelijk welke zaken in ieder geval belangrijk gevonden worden. Door de mogelijkheid van extra niet geformuleerde criteria te veronderstellen en door voor de scores van alternatieven op deze criteria een bereik af te bakenen kan nagegaan worden wat het mogelijke effect van vergeten criteria op de uitkomst van het afwegingsmodel zou kunnen zijn. Zonder de onzekerheid over het bestaan van deze

criteria te kunnen wegnemen kan wel onderzocht worden wat het effect van dergelijke criteria op de uitkomst zou kunnen zijn.

Een analyse van de onzekerheden en een schatting van de grootte van deze onzekerheden is een verplicht onderdeel voor een representatief model. Zonder deze inschatting is de informatie die het model levert in feite onbruikbaar, de betrouwbaarheid van de informatie is onbekend. Bij evaluatiemethoden wordt er nauwelijks rekening gehouden met onzekerheid in de uitkomsten. Terwijl hier toch alle redenen toe is, alleen al gezien het feit dat verschillende methoden tot verschillende voorkeursvolgorden kunnen leiden bij een gelijke input. Toevoegen van een onzekerheidsanalyse aan bestaande modellen is dan ook een must.

§ II.3 Specifieke vooronderstellingen en scripts per afwegingsmethode:

§II.3.1 Monetaire evaluatiemethoden

Bij een klassieke kostenbaten analyse worden de monetaire kosten en baten als enige criterium voor de afweging gebruikt. De methode is oorspronkelijk gebaseerd op ervaring van het bedrijfsleven met investeringsselectie. Deze bedrijfseconomische kostenbaten analyses hielden alleen rekening met de kosten en de baten van een investering voor het bedrijf. Later werden hier de kosten en baten van derden aan toegevoegd, zo ontstaat de maatschappelijke of sociaal-economische kostenbaten analyse. “Het doel van een maatschappelijke kostenbaten analyse is één of meer projecten of beleidsmaatregelen gericht op de oplossing van een gegeven probleem, te toetsen op sociaal-economische rentabiliteit, of verschillende alternatieven te rangschikken op basis van een rentabiliteitscriterium. Op deze wijze wordt beoogd tot een zo efficiënt mogelijke allocatie van productiemiddelen te komen.”(Ministerie van Financiën 1992, pag. 24)

Ik wil twee vooronderstellingen noemen die in de kostenbaten analyse ingebakken zitten. Ten eerste wordt verondersteld dat geld het enige criterium is of dat alle criteria in geld uit te drukken zijn. Ten tweede wordt de vooronderstelling gemaakt dat al het geld evenveel waard is.

§ II.3.1.1 Geld is het enige criterium of alle andere criteria zijn in geld uit te drukken.

Dat geld (winstverwachting) het enige criterium zou zijn gaat al niet op bij bedrijven. Bedrijven streven niet alleen naar winstmaximalisatie maar vaak ook naar overlevingskansen op lange termijn, goede arbeidsomstandigheden voor het personeel¹⁰, of bijvoorbeeld het leveren van kwaliteit. Toch is voor bedrijven de rentabiliteit van investeringen een belangrijk criterium. Voor de overheid speelt dit criterium ook mee, maar meer op de achtergrond. De

¹⁰ Zonder dat dit direct hoeft te leiden tot een verhoogde arbeidsproductiviteit.

rentabiliteit van beleidsopties kan zeker niet als het enige criterium gezien worden. Werkgelegenheid is bijvoorbeeld óók belangrijk. De veronderstelling die bij maatschappelijke kostenbaten analyses gemaakt wordt is dat deze criteria in voldoende mate om te rekenen zijn in geld. Een goed voorbeeld voor hoe een dergelijke omrekening van verschillende criteria in de Waddenzee plaats zou moeten vinden is (De Groot, 1994). Ik wil hier drie kanttekeningen bij plaatsen. Ten eerste heb ik het idee dat vaak het criterium zelf niet wordt uitgedrukt in geld, maar alleen het economische aspect van het criterium. Ten tweede zijn een aantal criteria, vooral ethische, niet om te zetten in geld. Ten derde is een omzetting in geld vaak niet democratisch.

Om de drie kanttekeningen uiteen te kunnen zetten zal ik eerst een aantal inleidende opmerkingen maken. Monetaire kostenbaten analyses maken gebruik van het begrip markt. Op de markt bepalen vraag en aanbod de meest efficiënte prijs. Dit principe werkt alleen (efficiënt) bij een markt met veel kleine vragers en aanbieders. Maar als het werkt is het waarschijnlijk de meest efficiënte manier om prijs, aanbod en vraag te bepalen. Als het niet werkt (een marktfout), bijvoorbeeld doordat een aanbieder of een groep aanbieders (monopolie, oligopolie, kartels) teveel marktmacht hebben waardoor ze een voor hen gunstig prijspeil kunnen afdwingen, of bijvoorbeeld doordat er niet een bestaande of niet aanwijsbare groep aanbieders of vragers is, dan is er alle reden voor overheidsingrijpen. Kostenbaten analyses kunnen een dergelijk overheidsingrijpen ondersteunen. Om zo efficiënt mogelijk te zijn probeert men daarbij zo veel mogelijk aan het marktprincipe vast te houden. Daarom wordt van zoveel mogelijk scores op criteria de prijzen bepaald. Moet bij een alternatief een nieuw baggerschip aangeschaft worden, dan bepaalt de markt de prijs voor het schip. Voor een aantal criteria is er echter geen markt of wordt de markt ongewenst beïnvloed door de eerder genoemde marktfouten. Er zijn een aantal methoden om dan toch een prijs te berekenen. Al deze methoden proberen zo goed mogelijk een markt te simuleren. Een kostenbaten analyse is dus altijd gericht op efficiëntie, want markten dienen de efficiëntie. Dit betekent dat zelfs KBA die gericht zijn op winstmaximalisatie (in tegenstelling tot efficiëntie-maximalisatie) nog steeds zijn gebaseerd op een eerdere efficiëntie-maximalisatie. Ik zeg dit, omdat zelfs als de prijsberekeningsmethoden de hypothetische markt perfect zouden simuleren, deze nog alleen de meest efficiënte oplossing berekent. Dit hoeft niet noodzakelijk de beste oplossing te zijn, zeker niet als ook andere waarden dan efficiëntie meespelen.

(Pearce, 1990) noemt de 'hedonic price approach', 'contingent valuation', 'travel cost approach', en 'willingness to pay vs willingness to accept'. Bij de 'hedonic price approach' probeert men de invloed van verschillende invloedsfactoren op de (gemeten) waarde van een gebied of object (grondprijs, waarde van een huis). Als de bijdrage van alle factoren aan de prijs bekend is kan ook voor andere situaties berekend worden wat de verwachte waarde is. In feite wordt hier dus een minimodel van de voorkeuren van 'de markt' voor verschillende factoren gebouwd. De 'contingent valuation method' baseert zich op interviews met vragen naar de 'willingness to pay' of de 'willingness to accept'. De 'travel cost approaches' proberen de waarde van een gebied af te leiden van de hoeveelheid geld die men bereid is te betalen voor de reis naar een gebied toe (dit is inclusief het geld dat verdiend had kunnen worden als men niet gereisd maar gewerkt had.) Hoe meer men bereid is te betalen, hoe meer het gebied waard is

Tot slot wordt de ‘willingness to pay’ en de ‘willingness to accept’ besproken. Het antwoord op de vraag naar hoeveel men bereid is te betalen voor een ‘niet-markt-goed’ blijkt vaak een factor 2 tot 4 kleiner dan het antwoord op de vraag hoeveel men minimaal wenst te ontvangen voordat men bereid is het gebruik van een ‘niet-markt-goed’ op te geven. Daarnaast blijken de verschillende methoden om een semi-marktprijs te bepalen soms ook tot een factor 4 van elkaar af te wijken! Deze laatste twee opmerkingen ter relativisering van de gedachte dat KBA veel betrouwbaardere rekenmethodes zouden zijn dan MCA.

Alleen het economisch aspect van een criterium wordt gemodelleerd

Wat meten de in de vorige perikoop geschetste methoden nu eigenlijk? Natuurlijk wordt er ook hier weer gestreefd naar efficiëntie. Daarnaast wordt alleen het economische aspect in de berekeningen meegenomen. Bij de ‘travel cost approach’ wordt alleen gevraagd naar de hoeveelheid **geld** die de reis naar een locatie kost. Maar als de reiskosten worden berekend, zelfs inclusief de hoeveelheid geld die er anders verdiend zou kunnen hebben in de tijd dat er nu gereisd wordt, dan worden de andere factoren niet meegenomen. De reis kan veel meer kosten dan alleen geld, bijvoorbeeld de zere rug van het autorijden of de verveling van de lange treinreis. En ook als mensen niet gaan kunnen ze de aanwezigheid van een gebied of park toch waarderen (bestaanswaarde), bijvoorbeeld omdat anderen er zo van genieten, omdat de luidruchtige burens er zo vaak heen gaan of gewoon, omdat het idee dat iets dergelijks bestaat ze aanstaat (intrinsieke waarde). Bovendien wordt het consumentensurplus niet meegenomen. Als ik naast het Rutbeek¹¹ woon kost een bezoek aan het Rutbeek mij niets. Dit betekent niet dat het Rutbeek mij niets waard is. Een onderzoek naar mijn ‘willingness to pay’ voor het Rutbeek zou wel eens uit kunnen wijzen dat ik f200,- per jaar over heb voor het voortbestaan van het Rutbeek. Samenvattend kan een ‘travel cost approach’ niet een compleet beeld geven van wat mensen over hebben voor het bestaan van een gebied.

Ook bij willingness to pay en willingness to accept methoden wordt ook alleen naar geld gevraagd. Intrinsieke waarde en consumentensurplus worden in principe wel meegenomen, voor zover men bereid is hiervoor te betalen. Maar de bereidheid om tijd te investeren of om morele normen of waarden te overschrijden wordt niet gemeten. En deze zijn lang niet altijd in geld uit te drukken¹², terwijl ze bij mensen wél meespelen in hun afweging van belangen. Samenvattend baseert een kostenbaten analyse zich op efficiëntiemaximalisatie en de modellering beperkt zich tot het financiële aspect van de criteria.

Niet alle criteria zijn om te rekenen in de eenheid geld

Geld is een ruilmiddel. Het heeft in zichzelf geen betekenis, het krijgt betekenis door wat je er voor kunt krijgen. Sommige dingen wil men echter hoe dan ook niet ruilen. Het

¹¹ Het Rutbeek is een recreatiegebied vlak buiten Enschede.

¹² Zie ook de volgende paragraaf.

leven, je kinderen of morele overtuigingen, die wil je voor geen geld kwijt. “You can’t buy my love” zongen de Beatles, en ik geloof dat ze daarmee dit punt wel ongeveer weergeven. Sommige dingen willen mensen echt niet ruilen, hoeveel geld je er ook voor biedt. Op dat moment gaat de vooronderstelling dat alles ruilbaar is en dus een prijs heeft niet meer op. Omrekenen in geld is dan zinloos of onmogelijk. In feite ontstaat dit probleem uit de vooronderstelling dat alle criteria uitwisselbaar zijn, zie ook § II.2.4.6.

Een goed voorbeeld van de gevolgen hiervan is een probleem dat bijvoorbeeld speelt bij het berekenen van de gevolgen van een dijkdoorbraak. Hoeveel is een mensenleven waard? Dijkverhogen kost geld. De dijk niet verhogen kost mogelijk mensenlevens. Bij het probabilistisch ontwerpen van een dijk wordt de overstromingskans vermenigvuldigd met de kosten van een inundatie. Deze kosten bestaan uit waterschade en mensenlevens. Een dijkverhoging levert een lagere kans op overstroming op, waardoor statistisch gezien het financiële risico kleiner wordt. De verhoging zelf kost echter geld. Als deze kosten lager zijn dan de verandering in het financiële risico, dan heeft de verhoging van de dijk zin. In dit soort berekeningen is het van belang hoeveel een mensenleven waard is. Wordt dit bijvoorbeeld op een ton gesteld dan zou een dijk daardoor bijvoorbeeld acht centimeter hoger ontworpen worden (in verhouding tot het niet meenemen van mensenlevens in de berekening.) Is een mensenleven een miljoen waard, dan wordt het ontwerp vijftien centimeter hoger. Hoewel het toekennen van een waarde aan een mensenleven raar aandoet, misschien zelfs ethisch onverantwoord is¹³, kan je de waarde van dat mensenleven niet zomaar weg laten. In dat geval wordt de dijk alleen maar lager. In de praktijk wordt de waarde van een mensenleven dan gewoon op nihil gesteld, wat ethisch gezien nog veel minder verantwoord is. Om dit soort problemen te omzeilen is onderzoek gedaan naar het risico dat mensen zelf acceptabel vinden (Smit 1995).

Een persoonlijk risico van 1 op 10^7 per jaar blijkt men dan gemiddeld wel acceptabel te vinden. Door dijken rond gebieden waar mensen persoonlijk risico lopen ten gevolge van een dijkdoorbraak¹⁴ zo hoog te maken dat aan deze risicovooraarde wordt voldaan is het hele probleem van de waardering van mensenlevens omzeild. Nu is er nog veel te zeggen rond deze problematiek, ook van economische aard, maar daar wil ik nu niet op ingaan. Het gaat mij er hier om te illustreren dat niet alle afwegingen economisch opgelost kunnen en hoeven te worden. Blijkbaar zijn er waarden die niet geruild kunnen worden tegen andere waarden, maar die gewoon gehaald moeten worden. Deze zijn derhalve niet economisch te waarderen en kunnen dus niet rechtstreeks in een KBA opgenomen worden. Zij zouden wel als harde randvoorwaarden voor een besluit gesteld kunnen worden.

Omrekenen naar geld is niet democratisch

Democratie gaat uit van het principe van: “one man one vote.” Door technieken als “willingness to pay” en “willingness to accept” te gebruiken, krijgt een vermogende industrieel

¹³ Als een mensenleven een miljoen waard is, dan is 2 miljoen een redelijk bedrag voor een moord.

¹⁴ In de meeste gebieden kunnen mensen gewoon op zolder gaan zitten, die loopt niet onder.

meer invloed dan een arme student. Maar ook de bioloog die vanwege zijn liefde voor de natuur een slecht betaald onderzoek doet naar een natuurgebied zou minder over hebben voor deze natuur dan de rijke edelman die er zo graag mag jagen en daarom wel een behoorlijke bom duiten overheeft voor het behoud van het gebied. Deze technieken mogen daarom alleen gebruikt worden als marktwerking op zijn plaats is. Op het moment dat het gaat om zaken waar democratisch gekozen moet worden, zijn de technieken niet toepasbaar.

§ II.3.1.2 Al het geld is evenveel waard

Ten tweede is er de vooronderstelling dat alle geld evenveel waard is. Nominaal is dat natuurlijk ook zo. Maar de ervaring is anders. Honderd gulden loonsverhoging is meer voor Jan Modaal dan voor de directeur. Voor de provincie Groningen is een miljoen gulden bruto door de Waddenzee toegevoegde waarde in Groningen belangrijker dan een miljoen gulden bruto door de Waddenzee toegevoegde waarde in heel Nederland.¹⁵ Eigenlijk mogen dus alleen kosten en baten voor één partij bij elkaar opgeteld worden.

Dezelfde overwegingen gaan in mindere mate op bij kostenefficiëntie analyses (KEA). Als de kosten of de baten bij elke variant identiek zijn, dan hoeven deze niet in geld omgezet te worden. Als bijvoorbeeld drie verschillende boortoren-ontwerpen alledrie tot dezelfde gasproductie leiden, dan hoeven alleen de maatschappelijke kosten in beeld gebracht te worden. Voor deze maatschappelijke kosten gelden nog steeds de bovengenoemde opmerkingen. In de meeste gevallen echter zullen nog de kosten nog de baten bij alle alternatieven gelijk zijn. In dat geval zijn de problemen bij KEA gelijk aan die bij KBA.

§II.3.2 Presentatiemethoden.

Presentatiemethoden gaan er van uit dat de beslisser zelf in staat is om een afweging tussen de verschillende alternatieven te maken. Deze afweging zelf hoeft niet expliciet gemaakt te worden. Alles wat de beslisser nodig heeft is een presentatie van alle relevante informatie. Volgens de scorekaart methode bestaat deze relevante informatie uit de verzameling criteria, de verzameling alternatieven en de verzameling score van deze alternatieven op de criteria. Het maakt verder niet uit in welke schaal deze scores aangeleverd worden, de beslisser is in staat om verschillende schalen in zijn weging mee te nemen.

De community impact analyse voegt daaraan toe dat er per belangengroep informatie beschikbaar moet zijn. Er moet dus per belangengroep vastgesteld worden wat volgens hen de scores van de alternatieven op de door de groep opgestelde criteria zullen zijn. Daarnaast moet ook weergegeven worden wat mogelijke maatregelen zijn om negatieve effecten te verminderen of te compenseren en hoe positieve effecten versterkt kunnen worden. Eventueel moeten dan ook weer de effecten van deze mogelijke maatregelen doorgerekend worden. Het onderscheiden van verschillende groepen is bij de community impact analyse van groot belang.

¹⁵Dat dit besef al eeuwen oud is, blijkt wel uit Marcus 12:41-44

Dit is ook de enige methode die daar rekening mee houdt. Ook het onderzoeken van mogelijk compenserende of versterkende maatregelen is een toevoeging die bij andere methoden niet terug te vinden zijn. In feite geeft een community impact analyse een veel breder beeld dan de meeste andere afwegingsmethoden. In principe moet het wel mogelijk zijn om deze gegevens ook in andere afwegingsmethoden te verwerken. Gebruik van de community impact methode betekent dat er veel meer informatie over de verschillende belangengroepen gezocht èn dat er een goede indeling in of keuze uit belangengroeperingen gemaakt moet worden.

Een speciale positie binnen de presentatiemethoden heeft de Amoebe.¹⁶ Bij de Amoebe moeten de scores wel kwantitatief aangeleverd worden en er wordt ook nog eens gestandaardiseerd. Het is daarom ook noodzakelijk om heldere kwantitatieve doelen te stellen. Tot slot wordt er wel weer verondersteld dat de beslisser genoeg heeft aan een presentatie van de scores zonder een expliciete integratie. De beslisser wordt daarbij geholpen door de grafische representatie. Dit heeft weer als nadeel dat elk criterium (visueel gezien) even zwaar meetelt.

§II.3.3 Multicriteria methoden

§ II.3.3.1 Weegfactoren – manieren om ze vast te stellen

Eén kenmerk van multicriteria methoden is het gebruik van weegfactoren. Alle multicriteria methoden maken op de een of andere manier gebruik van weegfactoren, een maat voor de verhouding van het belang van verschillende criteria tot elkaar. Zoals eerder vermeld is de toedeling van weegfactoren een zuiver subjectief gebeuren. De meeste multicriteria methoden veronderstellen dat er weegfactoren zijn en gaan hier verder niet op in. De gebruiker moet deze zelf maar invullen. Dat is lang niet makkelijk, zeker bij grotere aantallen criteria is het moeilijk om evenwichtig weegfactoren aan criteria toe te delen. (Ministerie van Financiën 1992) noemt een aantal methoden die wel in gaan op de “berekening” van weegfactoren, namelijk de verwachtingswaarde methode, de meerdimensionale schaalanalyse en in een speciale paragraaf (paragraaf 6.3.2 Het bepalen en berekenen van gewichten, vanaf pag. 126) directe methoden, de SAATY-methode, de extreme gewichtenmethode, de random-gewichten methode en hypothetische gewichtensets. Ik bespreek ze elk apart.

De eerste en eenvoudigste optie is het direct toekennen van gewichten. Deze gewichten kunnen op twee manieren verkregen worden, via “stated-” en via “revealed preference”. Bij “stated preference” wordt de beslisser (of de betrokken partijen) gevraagd welke gewichten zij toekennen. Hiervoor kan de beslisser bijvoorbeeld gevraagd worden om honderd punten te verdelen over de criteria. Verondersteld wordt dus dat de beslisser in staat is zijn afweging in een voorgevormde modelstructuur te passen. Het is bijvoorbeeld de vraag of een beslisser kan zeggen of hij milieu 2, 3 of 4 keer zo belangrijk vindt dan economie. Bij “revealed preference”

¹⁶ Een door Rijkswaterstaat ontwikkelde presentatiemethode waarin de kwaliteit van het ecosysteem van de Waddenzee grafisch gerepresenteerd wordt. Zie ook: (De Bruin, J., de Graaf, P.J.F., 1991)

kan op basis van eerder genomen beslissingen berekend worden hoe groot de criteria blijkbaar waren. Het model wordt dan als het ware gecalibreerd (vakjargon, betekent: geijkt). Dit laatste veronderstelt dat de voorgaande beslissingen consistent genomen zijn en dat weegfactoren constant blijven in de tijd.

Saaty, de ontwerper van de SAATY-methode gaat ervan uit dat het toekennen van weegfactoren eenvoudiger wordt als er minder criteria zijn. De eenvoudigste manier van toekennen is het vergelijken van twee criteria. Daarom heeft hij een systeem ontworpen waarbij op basis van paarsgewijze vergelijking weegfactoren berekend kunnen worden. De beslisser moet dan van elke mogelijke combinatie van twee weegfactoren aan geven hoe die in verhouding tot elkaar staan. Vervolgens kan met de SAATY-methode berekend worden hoe groot de weegfactoren dan moeten zijn, én hoe consistent de paarsgewijze vergelijking is ingevuld. Ter illustratie: stel er zijn drie criteria, namelijk kosten, natuurbehoud en werkgelegenheid. Stel de verhouding tussen kosten en natuurbehoud is 1:2, de verhouding tussen kosten en werkgelegenheid is 1:3 en de verhouding tussen natuurbehoud en werkgelegenheid is 2:3. Het zal duidelijk zijn dat de weegfactoren daarmee 1 voor kosten, 2 voor natuurbehoud en 3 voor werkgelegenheid worden. Maar als de gebruiker bij de derde paarsgewijze vergelijking 3:4 of 1:2 invult, dan is het antwoord van de gebruiker niet consistent. Met behulp van SAATY kunnen er dan toch weegfactoren berekend worden, daarnaast wordt ook de consistentie van de antwoorden berekend. Bij grotere aantallen paarsgewijs vergelijken zullen de antwoorden bijna altijd inconsistent worden, omdat een mens bij dat soort aantallen het overzicht niet kan bewaren. De mate van inconsistentie is dan een maat voor de kwaliteit van de weegfactoren, het geeft een indicatie van hoe goed de gebruiker aanvoelt wat de weegfactoren zouden moeten zijn. Bij grotere aantallen criteria neemt het aantal paarsgewijze vergelijkingen kwadratisch toe. Want als n het aantal criteria is, dan is het aantal paarsgewijze vergelijkingen gelijk aan $n*(n-1)/2$. Daarom is het bij grotere aantallen criteria handig om gebruik te maken van een criteriaboom. Hoofdcriteria worden dan onderverdeeld in subcriteria. Via paarsgewijze vergelijking wordt het aandeel van subcriteria in de hoofdcriteria (sub-weegfactoren), en het aandeel van de hoofdcriteria in het eindresultaat (weegfactoren) berekend. Het aantal paarsgewijze vergelijkingen neemt daarmee af. De structurering van de criteria neemt juist weer toe. Samenvattend is de SAATY-methode langzamer en bewerkelijker, maar tegelijkertijd wordt het bepalen van weegfactoren eenvoudiger.

De verwachtingswaarde methode gaat ervan uit dat de beslisser in staat is aan te geven welke weegfactoren hij belangrijker vindt. Als er een rangorde in de gewichten aangebracht is, legt de verwachtingswaarde methode de gewichten over een hypothetische schaal. Deze schaal is niet lineair en gaat ervan uit dat de belangrijkste gewichten relatief grotere afstanden tot elkaar hebben dan de minder belangrijke gewichten. Het resultaat is een set gewichten waarvan een beperkt aantal gewichten een relatief hoge score heeft en een groter aantal gewichten een relatief lage score.

Een vergelijkbare methode is de meerdimensionale schaalanalyse. Ook hier moet de beslisser eerst en volgorde in de gewichten aanbrenge. Daarna worden, met behulp van een meerdimensionale schaal analyse, kwantitatieve gewichten berekend.

De extreme gewichtenmethode heeft ook alleen een rangordening in de gewichten nodig. Vervolgens worden de extreme waarden van de gewichten berekend. Bij drie gewichten zal het grootste gewicht bijvoorbeeld maximaal een waarde van 1 hebben (100%) en minimaal van 1/3 (Dan zijn de andere twee gewichten net niet groter). Gewicht twee varieert van 0 tot, gewicht drie van 0 tot 1/3. Door nu maximale waarden voor de gewichten te kiezen kunnen de extreme punten in de resultaten berekend worden. Dit zou een redelijke indicatie van de eindrangschikking moeten geven. De extreme gewichtenmethode veronderstelt dat met een dergelijke brede range van waarden voor de weegfactoren toch onderscheid tussen de verschillende alternatieven gemaakt kan worden.

De Random-gewichten methode heeft een soortgelijke uitwerking als de extreme gewichtenmethode, al worden nu niet de extreme waarden van de gewichten als uitgangspunt genomen. In plaats daarvan worden een groot aantal keer "at random" waarden voor de gewichtenset getrokken (Waarbij de totale waarde gelijk moet blijven aan 1). De resultaten van de berekende prioriteitsvolgorde kunnen nu in een frequentietabel gezet worden. Een dergelijke methode van bewerken wordt een Monte-Carlo methode genoemd. Deze methode is in principe óók geschikt voor de evaluatie van de onzekerheid in het gehele model. Voorwaarde is dan wel dat er dan intervallen beschikbaar zijn voor de scores op criteria.

Bij hypothetische gewichtensets worden een aantal mogelijke en karakteristieke gewichtensets gegenereerd. Door de alternatieven aan deze gewichtensets bloot te stellen komt de robuustheid van alternatieven ten aanzien van verschillende hypothetische invalshoeken naar voren. Er zou bijvoorbeeld voor een economische, een milieuvriendelijke en een sociale invalshoek gekozen kunnen worden, waarbij de economische, de ecologische respectievelijk de sociale criteria meer nadruk krijgen. Als alternatieven bij verschillende gewichtensets goed blijven scoren, zijn ze blijkbaar robuust en dus acceptabel voor een grote groep mensen. Het invullen van gewichtensets voor anderen lijkt mij nog moeilijker dan voor de beslisser zelf. Anderzijds hoeft de hypothetische gewichtenset niet persé andere groeperingen te representeren, het kan ook gebruikt worden om de robuustheid van een alternatief te testen. In zijn algemeenheid is de keuze voor gewichten een subjectief gebeuren. De voorkeur van de beslisser telt. Als dit een probleem is, dan is het ook mogelijk om weegfactoren in te vullen met behulp van een enquête onder actoren. Dan wordt niet meer de subjectieve voorkeur van de beslisser weergegeven maar de subjectieve voorkeur van een (representatief) deel van de bevolking.

§ II.3.3.2 Omrekenen van kwantitatieve naar kwalitatieve criteria

Hoewel het technisch gezien mogelijk is om kwalitatieve criteria om te zetten in kwantitatieve criteria, is het toch niet verstandig om dit te doen, volgens (Ministerie van Financiën 1996). Toch zijn er twee methoden die dat wel degelijk doen. Dit zijn de meerdimensionale schaalanalyse en de verwachtingswaarde methode. Omdat informatie nooit preciezer gemaakt kan worden door een wiskundige bewerking alleen, kan er ook hier de kwaliteit van de informatie niet verhoogd worden. Hoewel er dus kwantitatieve informatie wordt gegenereerd, is de precisie van de informatie nog steeds kwalitatief. Dat de

betrouwbaarheid van de informatie niet toeneemt blijkt ook wel uit een eenvoudige voorbeeldsom. Deze wordt gemaakt voor de meerdimensionale schaalanalyse in §IV.1.1. Voor de verwachtingswaarde methode zou een analoge proef gedaan kunnen worden, met hetzelfde resultaat: er wordt een veel grotere mate van precisie gesuggereerd dan er in werkelijkheid voor handen is.

Als het noodzakelijk is om kwalitatieve kennis om te zetten in kwantitatieve, dan kan dat maar op één manier: informatie toevoegen. Gedacht moet dan worden aan extra onderzoek, expert-opinion of bijvoorbeeld schattingen van de beslisser in het geval van weefactoren.

§ II.3.3.3 Standaardisatie

Veel multicriteria methoden maken gebruik van standaardisatie. De kwantitatieve criteria kunnen niet zomaar met elkaar vergeleken worden. Daarom worden ze eerst omgezet (via standaardisatie) naar grootheden van dezelfde orde. Formeel worden de criteria omgezet in dimensieloze grootheden. (Van der Veen, 199?) Daarmee worden ze dan ook formeel vergelijkbaar. In de praktijk echter blijken de grootheden nog even onvergelijkbaar, de onvergelijkbaarheid wordt alleen voor de beslisser verdoezeld. Het is echter heel eenvoudig om terug te rekenen hoe de beslisser de verhoudingen tussen twee “onvergelijkbare” criteria inschat. Voor het model maakt het niet uit of deze onvergelijkbaarheid verdoezeld wordt of niet. Het model berekent gewoon een verhouding tussen de criteria, “onvergelijkbaar” of niet. Voor de kwaliteit van de uitkomsten zou het beter zijn om niet te verdoezelen. Vreemd genoeg wordt er bij de meeste methoden nauwelijks aandacht gegeven aan het verband tussen standaardisatie en het vaststellen van weefactoren. In (Ministerie van Financien 1992) wordt wel “...het toekennen van gewichten onafhankelijk van de oorspronkelijke scores afgeraden”. Dit is natuurlijk vreemd in het licht van het doel van standaardisatie zoals genoemd in (Van der Veen, 199?), namelijk het terugbrengen tot dimensieloze grootheden voordat vergeleken kan worden. Toch is het wel een heel nuttige opmerking. Het volgende voorbeeld laat dit zien. Stel dat eerst wordt gestandaardiseerd en daarna pas weefactoren worden vastgesteld onafhankelijk van de oorspronkelijke waarden. De volgende twee tabellen zijn dan na standaardisatie identiek:

	Alt 1	Alt 2	Alt 1 na stand.	Alt 2 na stand.	Weefactor
Milieu	50	80	5	8	3
Economie	7.000.000	2.000.000	7	2	2
Totaal ¹⁷			29	28	

Tabel 1 De eerste voorbeeldtabel. Milieu wordt gemeten in oppervlakte gereserveerd voor natuurgebied (ha) economie in bruto toegevoegde waarde gegenereerd in het gebied.

	Alt 1	Alt 2	Alt 1 na stand.	Alt 2 na stand.	Weefactor

¹⁷ Totaal wordt berekend via gewogen somming, weefactor₁ * score op criterium₁ + weefactor₂ * score op criterium₂

Milieu	500	800	5	8	3
Economie	7.000	2.000	7	2	2
Totaal			29	28	

Tabel 2 De tweede voorbeeldtabel. Milieu wordt gemeten in oppervlakte gereserveerd voor natuurgebied (ha) economie in bruto toegevoegde waarde gegenereerd in het gebied.

In tabel 1 is twintig hectare natuurgebied drie miljoen gulden waard. In tabel 2 is tweehonderd hectare nog maar drieduizend gulden waard. Dit scheelt een factor tienduizend. De beslisser die de weegfactoren vaststelt heeft dus geen idee wat hij feitelijk aan het afwegen is, zolang hij niet weet wat er gemeten wordt en hoe dit wordt gestandaardiseerd. Toch veronderstelt het afwegingsmodel dat de beslisser een zinnige uitspraak gedaan heeft over het belang dat hij aan de verschillende criteria hecht. De uitspraak dat de beslisser moet weten wat hij beslist is dus correct, de veronderstelling dat de criteria onvergelykbaar zijn en daarom eerst door standaardisatie naar dimensieloze variabelen teruggerekend moeten worden vervalt daarmee. Hier zijn naast de praktische¹⁸ ook theoretische argumenten voor aan te voeren. Ten eerste lost het standaardiseren de onvergelykbaarheid helemaal niet op, deze wordt alleen verdoezeld. Ten tweede zou er helemaal geen sprake moeten zijn van onvergelykbaarheid, de criteria zijn prima te vergelijken. Wat vergeleken wordt is de score op een criterium en de waarde die aan dat criterium gehecht wordt. Deze wordt vergeleken met een score op een ander criterium en de waarde die daar aan gehecht wordt. Dit soort vergelijkingen worden continu door mensen gemaakt. De uitkomst van deze vergelijking is een bepaalde waarde die aan een keus wordt gehecht, eventueel uitgedrukt in een bedrag dat men bereid is voor deze keuze te betalen. In feite is de hele marktwerking die leidt tot het kunnen uitdrukken van allerlei criteria in geld gebaseerd op dit soort “onvergelykbaarheden”. Het verschil tussen een monetaire kostenbaten analyse en een multicriteria methode is dan ook niet dat een KBA alleen vergelykbare en een MCA ook onvergelykbare criteria beoordeelt, het verschil is dat een KBA de verhouding in waarde tussen de criteria laat bepalen door de markt of een gesimuleerde markt, terwijl een MCA deze verhouding laat bepalen door de beslisser. Daarover meer aan het einde van dit hoofdstuk.

Hoewel standaardiseren dus niet nodig is vanwege de “onvergelykbaarheid van criteria” kan standaardiseren nog wel nuttig zijn, maar dan om de schalen van de verschillende grootheden vergelykbaarder te maken. Het volgende voorbeeld laat dit zien. Een vergelijking van de geschatte kosten van een civieltechnisch kunstwerk (expert-opinion, in gulden)¹⁹ met de inpassing in het landschap (enquêtes, schaal 1-10) bijvoorbeeld is lastig omdat de kosten van het kunstwerk in de tientallen miljoenen lopen. Stel er zijn twee alternatieve kunstwerken. De één scoort een 9 op inpassing in het landschap en kost f30.000.000,-, de ander scoort 6 op inpassing en scoort f25.000.000,-. Stel dit zijn de enige twee relevante criteria. Als de beleidsmaker heeft gekozen voor het duurdere en beter ingepaste kunstwerk, dan vindt hij de drie punten winst op inpassing meer waard dan de f5.000.000,- verlies op kosten. Als de weegfactor voor kosten op 1 wordt gesteld, dan is de

¹⁸ Praktisch argument: de kwaliteit van de beslissing wordt hoger als de beslisser weet waarover hij beslist

¹⁹ (middel waarmee de kennis vergaard wordt, eenheid waarin de criteriumscore wordt uitgedrukt)

weegfactor voor inpassing dus groter dan 1.666.667. Dit soort weegfactoren vergelijken niet prettig. Daarom is het makkelijker om beide criteria te standaardiseren. Heel voor de hand liggend is de standaardisatie van kosten naar miljoenen guldens in plaats van guldens. De kosten worden dan uitgedrukt in Megaflorijn. Dit is een heel nuttige standaardisatie met een factor 1.000.000. Er zijn verschillende methoden om tot standaardisatie te komen. Veel belangrijker is echter de afstemming tussen weegfactoren en standaardisatie. Daarom bespreek ik die eerst.

Standaardisatie kan dus wel nuttig zijn, omdat het ervoor zorgt dat weegfactoren in dezelfde orde van grootte blijven, wat het inzicht in de afweging vergroot. Tegelijkertijd is het van het grootste belang dat bij het vaststellen van weegfactoren precies bekend is wat er precies wordt afgewogen. Bij de selectie van de standaardisatiemethode zal dit aan de orde moeten komen. Voor de beslisser ondoorzichtige standaardisatiemethoden vallen bij voorbaat af. Het lijkt er op dat de meeste standaardisatiemethoden zijn gericht op het zo ver mogelijk vereenvoudigen van de cijfertjes. Bij voorkeur dienen de scores op criteria op een schaal van -1 tot 1 te liggen. Gevolg is de ontwikkeling van een aantal nutteloze standaardisatiemethoden. Deze noem ik eerst.

Er kan gedeeld worden door het rijmaximum. De scores worden daarbij gedeeld door de (absoluut gemeten) grootste score. Is deze score bijvoorbeeld 2137 arbeidsplaatsen, dan wordt vanaf nu het aantal arbeidsplaatsen gepresenteerd in delen van 2137 stuks. De vraag aan de beleidsmaker wordt bij een dergelijk systeem dan bijvoorbeeld: "Verdeel honderd punten tussen 2137 arbeidsplaatsen, 3.718 miljoen gulden, 134 zeehonden, 66 hectare natuurgebied en gemiddeld een 6.667 op inpassing in het landschap." Naar mijn mening heeft een dergelijke standaardisatie de afweging niet bepaald doorzichtiger gemaakt. De achtergrond van dit ondoorzichtig maken van de feitelijke keuze voor de beleidsmaker lijkt gelegen te zijn in de doelstelling. Het is waarschijnlijk de bedoeling de keuze mogelijk te maken door de beslisser niet te laten kiezen tussen precieze getallen, maar meer door vage abstracte begrippen, die door standaardisatie worden uitgedrukt in dimensieloze variabelen. Het gevolg is natuurlijk wel dat problemen zoals die gedemonstreerd zijn in de tabellen 1 en 2 kunnen optreden. Dezelfde kritiek kan geleverd worden op standaardisatie door te delen door het rijtotaal, intervalstandaardisatie of met behulp van vectornormalisatie, daarom zal ik deze verder niet bespreken. De enige nuttige standaardisatie (genoemd in Ministerie van financiën 1992) die dan overblijft is de deling door doel- of ideaalwaarden. Dit kan nuttig zijn omdat juist de doelen die gesteld zijn de scores in het juiste perspectief kunnen zetten. Een score van 200 zeehonden zegt mij persoonlijk niets. Maar als die score, zoals dat in de AMOEBE gebeurt, uitgezet wordt tegen de gewenste score, krijg ik veel meer inzicht in het effect van een maatregel. Juist bij die criteria waar een beslisser niet genoeg kennis heeft van het systeem om de criteriumscore op zijn merites te beoordelen is een standaardisatie met behulp van doelwaarden erg nuttig. De expert bereidt de keuze dan in feite voor door de beslisser een doelwaarde aan te reiken als referentie.

Samenvattend is standaardisatie oorspronkelijk bedoeld als methode om "onvergelijkbare" variabelen vergelijkbaar te maken door ze dimensieloos te maken. Het probleem van de onvergelijkbaarheid wordt daarmee echter alleen formeel opgelost terwijl in de praktijk de onvergelijkbaarheid blijft en de beslisser geen idee heeft hoe hij de

onvergelijkbare criteria met elkaar aan het vergelijken is. Standaardisatie kan wel nuttig zijn om de weegfactoren klein genoeg te houden en kan heel nuttig zijn om systeemkennis van een expert over te dragen op een beslisser. Voorwaarde is wel dat de beslisser zicht blijft houden op betekenis van de criteria waarvoor hij een afweging maakt.

§ II.3.3.4 Linearisatie

Bijna alle multicriteria analyses zijn lineaire modellen. Ze veronderstellen dat de werkelijkheid lineair is, of benaderen deze met een lineair model. Weegfactoren bijvoorbeeld, geven de verhouding tussen het belang van verschillende criteria aan. De veronderstelling die met een lineair model gemaakt wordt is dat deze verhouding constant is, onafhankelijk van de grootte van de scores op deze criteria. Dit zal lang niet altijd het geval zijn. Bij een werkloosheid van 17% zal werkgelegenheid misschien belangrijker gevonden worden dan het voorkomen van milieuschade. Maar als een plan voorziet in een werkgelegenheids groei van 30% van de beroepsbevolking, wordt eventuele milieuschade veel belangrijker, want wat moet je met werkgelegenheid zonder dat er mensen zijn die daar behoefte aan hebben. Toch zal een lineair model dat gecalibreerd is bij een werkloosheid van 17% voorspellen dat ook de werkgelegenheid voorbij de die 17% belangrijker is dan de milieuschade. Nu verwacht ik niet dat er plannen komen die de werkloosheidsproblematiek in één klap oplossen en een banenoverschot creëren. Maar het voorbeeld geeft aan dat lineaire modellen maar op een beperkt domein geldig zijn. Voor werkloosheid is dat domein waarschijnlijk groot genoeg, maar is dat het geval voor alle criteria? Voor elke weegfactor zal daarom een schatting gemaakt moeten worden voor welk domein ze geldig zijn, of concreet: voor welke range van scores van alternatieven op criteria het model nog opgaat.

§ II.3.3.5 Rangordening

Een aantal MCA maken gebruik van voorkeursvolgordes. Er wordt bijvoorbeeld per criterium vastgesteld wat de voorkeursvolgorde in de alternatieven is. Neem de alternatieven auto, fiets, vliegtuig en de criteria prijs en snelheid, dan is de voorkeursvolgorde op prijs: fiets > auto > vliegtuig, en op snelheid: vliegtuig > auto > fiets. Deze informatie vormt de input van de afweging. Omdat er slechts gebruik wordt gemaakt van informatie op ordinale schaal, zal informatie die op een andere schaal bekend is omgerekend moeten worden. Op zich is dat geen probleem, de omrekening is heel eenvoudig, maar er kan wel enorm veel informatie verloren gaan. Als bijvoorbeeld bekend is dat een fietsreis naar Rome vier weken duurt, een autoreis twee dagen en het vliegreis inclusief in- en uitchecken ook wel een dag duurt, dan is er in feite een score verhouding van 28 : 2 : 1, deze wordt na “ordinalisatie” 3:2:1. Het lijkt nu of een fietstocht drie keer zo lang duurt als een vliegreis. Weliswaar betekent “3:2:1” dat formeel niet, maar ook in de behandeling van deze cijfers wordt wel zo behandeld. Deze methode kan wel werken, maar alleen als de bekende gegevens echt ordinaal zijn, of wanneer deze zich “ordinaal tot elkaar verhouden”.

§ II.3.3.6 Paarsgewijze vergelijking

Sommige multicriteria methoden maken gebruik van paarsgewijze vergelijking. Dit is een ander soort paarsgewijze vergelijking dan die gebruikt wordt bij de SAATY-methode. Bij deze methoden worden de alternatieven paarsgewijs met elkaar vergeleken per criterium. Als er bijvoorbeeld drie alternatieven zijn en drie criteria, dan wordt eerst de score op criterium 1 van alternatief 1 met alternatief 2 vergeleken, daarna van de alternatieven 2 en 3, daarna van 3 en 1. Op basis van deze drie vergelijkingen wordt daarna een voorkeursvolgorde bepaald van de alternatieven voor dit criterium. Deze volgorde kan op verschillende manieren bepaald worden. Daarna kunnen deze voorkeursvolgorden per alternatief weer gebruikt worden voor een totale voorkeursvolgorde. Het vreemde van deze methoden is dat de score van een alternatief afhankelijk wordt van andere alternatieven. Het is zelfs mogelijk om door een alternatief toe te voegen, de voorkeursvolgorde van de andere alternatieven te beïnvloeden (zie bijlage 2). Het zal duidelijk zijn dat dit de betrouwbaarheid van de methode niet vergroot. Ik kom hier in hoofdstuk IV op terug.

§ II.3.3.7 Black-box berekeningsmethoden

Een aantal multicriteria analyses maken gebruik van weinig doorzichtige rekenmethoden. Zij berekenen op de één of andere manier een voorkeursvolgorde, maar hoe dat gedaan wordt is niet direct inzichtelijk en de antwoorden zijn ook niet direct narekenbaar. (Dit in tegenstelling tot methoden die gebaseerd zijn op gewogen somming, de berekening is daar doorzichtig.) Dit kan tot gevolg hebben dat een methode minder vertrouwen wekt bij de gebruiker. Bovendien kan de berekening van onzekerheid en gevoeligheid ingewikkelder worden bij dit soort methoden. Ondoorzichtige berekeningsmethoden moeten met extra voorzichtigheid benaderd worden. Vooral als niet duidelijk is waarom een methode iets op een bepaalde manier berekent en wat dit voor gevolgen heeft.

§ II.3.3.8 Overeenkomsten tussen kostenbaten analyse en gewogen somming.

Hoewel in eerste instantie gesuggereerd is dat er enorme verschillen zijn tussen kostenbaten analyses en multicriteria methoden blijken deze in de praktijk wel mee te vallen. Met name de overeenkomst tussen een KBA en de MCA “gewogen somming” is frappant. Beide methoden zijn in feite vormen van gewogen somming, het enige verschil is de manier waarop weegfactoren worden vastgesteld. Bij een KBA wordt de weging bepaald door de markt of door een gesimuleerde markt. Lukt dat niet dat wordt een criterium genegeerd. Bij een gewogen somming bepaalt de beslisser de weegfactoren. Daarnaast is het mogelijk de weegfactoren middels interviews te laten vaststellen door burgers. Zolang hier niet gevraagd wordt naar wat zij zelf over hebben voor een alternatief, maar wat zij vinden dat de staat over zou moeten hebben voor dit alternatief, blijft het principe van “one man one vote” overeind. Alle criteria doen mee. De KBA-analyse wordt daarmee meer geschikt voor afwegingen

waarbij de markt dient te beslissen, de gewogen sommering wanneer de afweging meer een zaak van de beslisser of de overheid is. Verder zijn de methoden praktisch hetzelfde.

§ II.4 Conclusies

In hoofdstuk twee zijn een groot aantal aspecten van afwegingsmethoden besproken aan de hand van modelkenmerken van (Rip, 1996). Afwegingsmethoden zijn in eerste instantie prescriptieve modellen, ze schrijven voor hoe een afweging gemaakt moet worden. Zijn ze eenmaal ingevuld, dan is er ook sprake van representatie en van een speeltuinfunctie. De ingevulde methoden kunnen ook zo gebruikt worden.

Een belangrijk punt bij de beschrijving van modellen is de manier waarop geïntegreerd wordt. Er zijn een aantal vormen van integratie besproken. Ten eerste worden verschillende criteria op de één of andere manier geïntegreerd. Deze integratie kan impliciet door de beslisser gemaakt worden, of expliciet in een Multicriteria analyse. Bij kostenbaten analyses worden alle criteria gereduceerd tot één criterium. De tweede vorm van integratie is de aggregatie van informatie tot het juiste aggregatieniveau. Alleen bij de SAATY-methode wordt deze aggregatie expliciet weergegeven. Bij kostenbaten analyse doet aggregatie er niet veel toe, omdat er maar één criterium is. Bij de andere methoden wordt verondersteld dat voor de invulling van de methode al een juist aggregatieniveau bereikt is. Ten derde is er de integratie van verschillende schalen. Sommige criteria kunnen uitgedrukt worden in een kwantitatieve schaal, anderen zullen misschien kwalitatief moeten blijven. Een afwegingsmethode zal op de één of andere manier deze verschillende schalen met elkaar moeten combineren om zo tot een eenduidig eindresultaat te komen. Tot slot is er de integratie van verschillende deelmodellen. De afwegingsmethode moet waarschijnlijk ingepast worden in een besluitvormingsproces waarbij mogelijk ook andere modellen (impactanalyse-modellen) gebruikt gaan worden. Een soepele aansluiting van de verschillende deelmodellen op elkaar is duidelijk een pré.

Aan elk model liggen vooronderstellingen ten grondslag. Deze vooronderstellingen kunnen de gebruiker van het model in een bepaalde (mogelijk ongewenste) richting dwingen. Een goede analyse van deze vooronderstellingen is daarom noodzakelijk. In hoofdstuk twee zijn algemene vooronderstellingen en per model specifieke vooronderstellingen behandeld. Tot slot zijn de verschillende manieren waarop modellen gebruikt kunnen worden behandeld. Deze zijn achtereenvolgens: ‘instrumenteel gebruik’, ‘conceptueel gebruik’, ‘strategisch gebruik’ en ‘het hanteerbaar maken van onzekerheid.’

In hoofdstuk twee zijn ook een aantal verschillende functies die afwegingsmethoden kunnen vervullen naar voren gekomen. Dit zijn de volgende:

- Rationeel beslissen
- Handleiding voor “hoe te beslissen”
- Informatie verschaffen over de preferenties van de beslisser, hulp bij het zoeken naar nog betere alternatieven
- Helpen ordenen van de preferenties van de beslisser

- Verantwoording afleggen van de keuze, representatie van een gemaakte keuze
- Inzicht verschaffen in eigen voorkeuren
- Structureren van de besluitvorming
- Expliciete integratie van verschillende “onvergelijkbare” criteria
- Reductie van alternatieven
- Communicatie
- Strategisch gebruik
- Inschatten van onzekerheid
- Gericht reduceren van onzekerheid
- Onzekerheden opsporen (scenarioanalyse) en bedenken van mogelijke reactieplannen
- Het weergeven van gevolgen voor de belangengroepen
- Het genereren van mogelijke compenserende maatregelen

Hoofdstuk III Kwaliteitscriteria voor afwegingsmethoden

In dit hoofdstuk zullen kwaliteitscriteria voor afwegingsmethoden afgeleid worden. Daarvoor worden de algemene kwaliteitscriteria van Covello gebruikt. Deze criteria, die oorspronkelijk ontwikkeld zijn om uitspraken te kunnen doen voor risk-assessment modellen, worden door (Kolkman 1996) gebruikt voor het vaststellen van algemene modelkwaliteitscriteria. In paragraaf 1 wordt het concept ‘modelkwaliteit’ uitgewerkt, in paragraaf 2 worden de modelkwaliteitscriteria gepresenteerd, in paragraaf 3 worden de algemene modelkwaliteitscriteria toegespitst op afwegingsmethoden en in paragraaf 4 volgen de samenvatting en conclusies.

§ III.1 Wat is modelkwaliteit?

Dat een gebruikt model goed moet zijn, is elke gebruiker duidelijk. Wat dat goed zijn inhoudt, verschilt per gebruiker. Meestal moet het model, uitgaande van een gecomputeriseerd model, gebruikersvriendelijk zijn, de resultaten moeten duidelijk en mooi gepresenteerd worden, met behulp van die resultaten moeten ‘goede’ voorspellingen gedaan kunnen worden, en het is natuurlijk belangrijk dat de resultaten ook aansluiten bij de problemen die met het model opgelost moeten worden. Modelbouwers en wetenschappers richten zich meestal op de ‘goede’ voorspellingen. Zij doen wat minachtend (is mijn ervaring) over de gelijke grafische user-interfaces die vanzelfsprekend alleen gemaakt worden door commerciële bureaus en waarachter derhalve een black-box-model²⁰ hangt, ongetwijfeld gevuld met achterhaalde of foutief geïnterpreteerde wetenschappelijke gegevens. Deze minachting komt voort uit de sterke oriëntatie op de modelinhoud. Een dergelijke eenzijdige oriëntatie lijkt mij niet correct, zeker niet bij modellen waarbij de representatie-functie van minder belang is. Immers, een model dat zich niet richt op een zo goed mogelijke representatie van de werkelijkheid hoeft ook niet beoordeeld te worden op de mate van representatie, de modelinhoud is dan minder van belang.

Juist de bruikbaarheid, de presentatie, de acceptatie en de communiceerbaarheid van het model kunnen dan een grote rol gaan spelen. Bovendien zijn de technieken die gebruikt worden om de modelinhoud te controleren lang niet allemaal toepasbaar op evaluatiemethoden. Calibratie (ijken, met behulp van meetgegevens) en validatie (mate van overeenkomst met de werkelijkheid controleren aan de hand van meetgegevens) maken gebruik van meetgegevens. Maar wat zijn meetgegevens bij evaluatiemethoden? Als een beslisser een evaluatiemethode gebruikt om een beslissing te nemen, kan hij dan achteraf zeggen, het model heeft de juiste beslissing genomen? Als dat het geval is, waarom had hij dan überhaupt de afwegingsmethode nodig? Als de kwaliteit van het product niet direct beoordeeld

²⁰ Een black-box model is ontoegankelijk voor de gebruiker. Dit heeft het voordeel dat de gebruiker niet kan knoeien aan het model, wat voorkomt dat een onervaren gebruiker een goed model naar de Filistijnen helpt, maar het heeft het nadeel dat het niet mogelijk is de ‘interne’ kwaliteit van het model te beoordelen.

kan worden, kan altijd nog de kwaliteit van het proces dat het product opgeleverd heeft gebruikt worden als criterium voor de productkwaliteit. Bij de validatie van de interne kwaliteit van het model moet de overstap van product- naar proceskwaliteit gemaakt worden. Of de keuze die het model maakt overeenkomt met de beste keuze kan niet direct gemeten worden, wel kan worden vastgesteld of de kwaliteit van het proces van de afweging goed is geweest.

Om bovengenoemde redenen is het beter om afwegingsmethoden op brede algemene modelkwaliteitscriteria af te rekenen. Op de U.T. wordt daarvoor gebruik gemaakt van een soort checklist van kwaliteitscriteria voor risk-assessment methoden opgesteld door Vincent T. Covello (geciteerd in Kolkman 1996) die volgens Kolkman ook goed bruikbaar is voor modellen in het algemeen. Aan de hand van deze checklist kan de kwaliteit van afwegingsmethoden, geïnterpreteerd als modellen van het keuzeproces, beoordeeld worden. Daarnaast gebruik ik door Kolkman opgestelde aanvullingen.

§ III.2 Modelkwaliteit volgens Covello

Covello maakt onderscheid tussen twee soorten kwaliteitscriteria, namelijk interne en externe modelkwaliteitscriteria. Interne criteria beoordelen de kwaliteit van het model zelf, de externe kwaliteitscriteria meten de kwaliteit van de inpassing van het model in zijn omgeving. De ‘goede’ voorspellingen slaan met name op interne criteria, de gelijke grafische user-interface zou goed kunnen scoren op externe criteria. Vertaald en aangevuld kunnen de kwaliteitscriteria volgens Covello als volgt weergegeven worden:

§III.2.1 Interne modelkwaliteit

1. logica
 - a) Kan de methode verdedigd worden op basis van theoretische argumentatie?²¹
 - b) Wat zijn de aannames en zijn deze aannemelijk?
 - c) Komt de praktische toepassing van de methode overeen met de achterliggende theorie? (Zijn de afwijkingen acceptabel?)
2. compleetheid
 - a) Theoretisch: Zijn alle probleemaspecten opgenomen?
 - b) Functionele validatie: Worden alle bekende systeemgedragingen ook gemodelleerd?
 - c) Operationeel: In hoeverre wordt informatie weggelaten omdat het nauwelijks verkrijgbaar is, of te duur? Is dit verantwoord?
3. accuratesse

²¹ Het gaat hierbij om de argumentatie van de ontwerper van een methode, duidelijk niet om de argumenten die in dit verslag aangedragen worden om voor een methode te kiezen. De veronderstelling hier is dat als een ontwerper een goede argumentatie heeft voor zijn ontwerp, het waarschijnlijk is dat het ontwerp zelf ook goed zal zijn.

- a) Is de betrouwbaarheid van de resultaten groot genoeg?
- b) Ligt er te veel nadruk op bepaalde modelonderdelen?
- c) Gevoeligheid voor vereenvoudigingen en aannames, hoe groot is de impact van vooronderstellingen?
- d) Hoe groot is de gevoeligheid voor fouten in de modelinput?
- e) Is er gekozen voor het juiste aggregatieniveau?

§III.2.2 Extern modelkwaliteit

1. acceptatie
 - a) In hoeverre past het model binnen bestaande instituten en processen?
 - b) In hoeverre wordt het model als eerlijk, ethisch verantwoord en rationeel gezien?
 - c) Zijn ook mensen zonder technische achtergrond in staat het model te begrijpen?
 - d) In hoeverre is de gebruiker vertrouwd met en heeft hij vertrouwen in dit type modellen?
 - e) Sluit de presentatievorm aan bij de belevingswereld van de beslisser?
 - f) Aanvaardt de gebruiker de sturende elementen (scripts) in het model?
 - g) Wordt het model ook geaccepteerd door actoren (op alle voorgaande punten) die (indirect) door het model beïnvloed worden en die de gebruiker kunnen beïnvloeden?
2. praktische bruikbaarheid
 - a) Heeft de gebruiker de kennis in huis om het model goed en verantwoord te gebruiken?
 - i) Modelafhankelijk
 - a) Inzichtelijkheid
 - b) documentatie
 - c) gebruikersvriendelijkheid
 - d) robuustheid (m.b.t. input en bediening.)
 - ii) gebruikerafhankelijk
 - a) opleiding
 - b) ervaring
 - b) Zijn de benodigde rekenfaciliteiten aanwezig?
 - c) Is er genoeg tijd om het model te laten draaien, is het snel genoeg?
 - d) Is de benodigde input beschikbaar? Is de methode flexibel ten aanzien van de soort input?
3. effectiviteit
 - a) Zijn de modelresultaten nuttig?
 - i) Komen de modelleergrootheden overeen met de grootheden waarover de beslisser informatie behoeft?
 - ii) Is het model geldig op het domein waarop de gebruiker het wil toepassen?
 - b) Worden de onzekerheden in de modeluitkomsten expliciet weergegeven?
 - c) Is de methode bruikbaar voor nieuwe en andere problemen?
 - d) Zijn inzichten en conclusies generaliseerbaar?

- e) Is de methode effectief en efficiënt te koppelen aan andere methodes en modellen?

De criteria in deze checklist behoeven een verdere uitwerking om ze bruikbaar te maken voor afwegingsmethoden. Een groot deel van de externe en een aantal interne criteria zijn bovendien sterk afhankelijk van de wensen van rijkswaterstaat. Tot slot is ook het belang dat aan verschillende criteria gehecht moet worden afhankelijk van de opdrachtgever. In het volgende hoofdstuk zal ik deze criteria uitwerken.

§ III.3 De betekenis van modelkwaliteitscriteria voor afwegingsmethoden.

In dit tweede deel van hoofdstuk III zal ik de in het eerste deel beschreven modelkwaliteitscriteria uitwerken. Ik wil de algemene modelkwaliteitscriteria concretiseren voor afwegingsmethoden. Daarnaast is ook gebruik gemaakt van gesprekken met rijkswaterstaat, om een indicatie te hebben van de richting die zij op willen met afwegingsmethoden.

Modelkwaliteitscriteria zijn bedoeld voor modellen in het algemeen. Omdat afwegingsmethoden tot een specifieke groep modellen behoren is een nadere aanscherping van de criteria voor afwegingsmethoden nuttig. Een aantal criteria zal minder relevant blijken, van andere criteria zal het toepassingsgebied verder geconcretiseerd kunnen worden en aan sommige criteria zal het één en ander toegevoegd moeten worden. De criteria zullen één voor één behandeld worden. Elk punt begint met de uitwerking van het criterium voor de selectie van een afwegingsmethode, eventueel gevolgd door aandachtspunten voor het gebruik van een afwegingsmethode.

§III.3.1 Interne criteria

§ III.3.1.1 Logica

Dit hele verslag is een argumentatie voor de keuze van een methode. Bij het punt logica gaat het onder andere om de argumentatie voor een model. Dit gaat echter om een andere argumentatie, namelijk om een intern criterium: De eis van een goede, logisch consistente argumentatie zoals die aan wetenschap en wetenschappelijke modellen wordt gesteld. Het probleem is dat afwegingsmethoden niet ontwikkeld zijn om de keuze die gemaakt wordt zo goed mogelijk te representeren. Het zijn zoals in hoofdstuk 2 naar voren kwam, eerder prescriptieve modellen, ze zeggen hoe een keuze gemaakt behoort te worden. Het doel van een prescriptief model is dus helemaal niet het bereiken van een hoge mate van gelijkheid met de in het model afgebeelde werkelijkheid. In (Ministerie van Financiën 1992) worden nauwelijks argumenten gegeven waarom een keuze voor een bepaald model gemaakt zou moeten worden. Er worden hoogstens wat technische voor- en nadelen van de verschillende methoden

genoemd. Toch is argumentatie ook van belang voor prescriptieve modellen, namelijk om mensen te overtuigen dat dit een goede manier van werken is.

Ik wil dit punt op drie manieren meenemen in de beoordeling van methoden. Ten eerste mag de theoretische argumentatie niet ontbreken of duidelijk aanvechtbaar zijn. Ten tweede is het belangrijk om alle vooronderstellingen achter de afwegingsmethoden te kennen en de acceptatie van een vooronderstelling te kunnen beargumenteren. Bij de keuze voor een afwegingsmethode zullen de specifieke vooronderstellingen van de methode beoordeeld moeten worden. Onacceptabele vooronderstellingen maken een onacceptabele methode. Ten derde zullen de methoden in de praktijk ook moeten voldoen aan de theoretische vooronderstellingen. Dit lijkt heel voor de hand liggend, maar de in hoofdstuk II genoemde discrepantie tussen de theorie van standaardiseren en de praktijk valt onder dit punt.

Daarnaast zal na de keuze van een afwegingsmethode de manier waarop de methode wel en niet gebruikt mag worden gedetailleerd beschreven en beargumenteerd moeten worden. Een bespreking van de vooronderstellingen kan gevonden worden in hoofdstuk II. De voor alle afwegingsmethoden geldende vooronderstellingen zijn beargumenteerd in paragraaf II.3.4: *geography of responsibilities*.

In de vorige twee alinea's werd in feite alleen de afwegingsmethode als prescriptief model behandeld. Als een afwegingsmethode in gebruik en ingevuld is, wordt in feite gebruik gemaakt van een representatief of speeltuinmodel (zie ook paragraaf II:3:1). Ook dit ingevulde model zal logisch consistent moeten zijn. Dit vraagt om een goed beargumenteerde keuze voor criteria, een goede argumentatie voor de manier waarop de scores op criteria voorspeld worden, en, indien een MCA gebruikt wordt, om een argumentatie voor de keuze en grootte van de weegfactoren.

§ III.3.1.2 Compleetheid

Bij het criterium compleetheid komt in de eerste plaats het representatieve karakter van het model naar voren. Volgens Rip is het belangrijk dat bij een representatief model de structuur van het model overeenkomt met die van de gemodelleerde werkelijkheid. De structuur van de afweging wordt grotendeels bepaald door het gekozen afwegingsmodel. Om het model dus een goede weergave te laten zijn van de te nemen beslissing moet de structuur van het model overeenkomen met de structuur van de te nemen beslissing. Deze structuur is echter juist weer afhankelijk van de beslisser, die een model kiest om zo zijn beslissing te kunnen structureren. Juist vanwege de onbekendheid met mogelijke structuren en met welke structuur nu goed zal zijn voor een specifieke situatie is het niet mogelijk om van tevoren vast te stellen welke structuur het beste met de "werkelijke" structuur overeenkomt. Daarom lijkt het mij het beste om de verschillende structuren te presenteren zodat bij de keuze voor een model duidelijk is wat mogelijke structuren zijn. Bij de bespreking van dit onderdeel in §IV.1.4 wordt de betekenis van het begrip structuur in het geval van afwegingsmethoden verder duidelijk.

Verder onderscheidt dit criterium vooral monetaire methoden van andere afwegingsmethoden. Monetaire methoden beperken zich tot het financiële aspect van de

afweging. Is de afweging daarmee compleet genoeg? Presentatiemethoden beperken zich tot de presentatie van de scores zonder ze onderling te vergelijken. Weer komt de vraag op, is dit genoeg? Tenslotte is er het onderscheid tussen MCA die wel en die niet om kunnen gaan met kwalitatieve criteria. Zijn alleen kwantitatieve criteria voldoende of maken kwalitatieve criteria de afweging pas compleet? Al deze vragen moeten door Rijkswaterstaat beantwoord worden.

In een gesprek met de opdrachtgever bleek het volgende:

- Alleen monetaire aspecten zijn niet voldoende. De themagroep 'gebruik' volgt een driesporenplan. Eén van die sporen behelst de selectie van economische en ecologische criteria. Het is niet de bedoeling om de ecologische criteria in guldens uit te drukken
- Presentatiemethoden schieten tekort. Het is wel de bedoeling om expliciet af te wegen. Een groot deel van de problematiek waar de themagroep gebruik mee te maken heeft bestaat uit het afwegen zelf. Presentatiemethoden bieden hiervoor geen oplossing.
- De groep die zich richt op het zoeken van criteria geeft aan dat ook kwalitatieve criteria gebruikt zullen worden. De criteria worden waarschijnlijk in een soort 'criteriaboom' gezet, met bovenin de meer abstracte en onderin de meer concrete criteria.

Dit betekent dat een complete methode in staat moet zijn tot vergelijking van alle soorten criteria met niet alleen verschillende grootheden maar ook met verschillende schalen.

Voor de ingevulde afwegingsmethode is op het gebied van compleetheid vooral van belang dat alle relevante criteria in de methode opgenomen zijn. Een complete set criteria is noodzakelijk voor een complete afweging. Het probleem is dat de compleetheid van de afwegingsmethode mede afhankelijk is van de belanghebbenden. Met elke nieuwe belanghebbende kunnen er nieuwe argumenten voor of tegen worden ingebracht. Deze nieuwe argumenten betekenen dat er ook nieuwe criteria nodig zijn voor een complete afweging. Een flexibele methode die het mogelijk maakt om nieuwe criteria te blijven toevoegen lijkt mij daarom noodzakelijk.

Tegelijkertijd zal voor elk nieuwe alternatief onderzoek gedaan moeten worden om de scores op alle criteria te kunnen invullen. Als het mogelijk is om al deze scores op criteria te voorspellen met één betrouwbaar model (of BOS), hoeft dit geen probleem te zijn. Als bij elk onderzoek expert opinion of enquêtes nodig zijn, dan kan dit een kostbare zaak worden. Het zou daarom handig zijn om de gevoeligheid van de uitkomsten van de afweging voor scores op elk criterium te kunnen meten. Zo kan vastgesteld worden of het weglaten van moeilijk te berekenen criteria veel schade doet aan de betrouwbaarheid van de afwegingsmethode. Met behulp van een gevoeligheidsanalyse kan dan vastgesteld worden welke informatie wel en welke niet weggelaten kan worden. De resultaten worden hiermee beter dan de andere mogelijkheden, namelijk moeilijke of dure criteria direct weglaten of alle criteria berekenen ongeacht kosten of moeite. De criteriagroep moet zich, in mijn ogen, dan ook in eerste instantie bezig houden met het inventariseren van alle relevante criteria, daarna moet pas geselecteerd worden op de mate waarin de criteria berekend of gepresenteerd kunnen worden. Met een gevoeligheidsschatting kan dan onderzocht worden hoe belangrijk de criteria in de afwegingsmethode zullen zijn en of ze de kosten van onderzoek waard zijn.

§ III.3.1.3 Accuratesse

De betrouwbaarheid van de uitkomsten van de afwegingsmethode zijn nauwelijks te valideren. Zoals eerder gezegd zijn voor een klassieke validatie meetgegevens nodig. Het gecallibreerde (geijkte) model moet een voorspelling doen voor een keuze met bekende afloop, waarna deze voorspelling vergeleken wordt met de bekende afloop. De afwijking tussen gemodelleerde en werkelijke keuze is dan een maat voor de betrouwbaarheid van het model. Zeker met meerdere vergelijkingen kan dan een statistisch betrouwbaarheidsinterval worden opgesteld, bijvoorbeeld dat het model in 95% van de gevallen minder dan 10% afwijkt van de werkelijke situatie. Echter, afwegingsmethoden zijn normatief. Ze voorspellen niet hoe een keuze gedaan gaat worden, ze schrijven voor hoe een keuze gedaan moet worden. Van vergelijken met eerder gemaakte keuzes kan geen sprake zijn. Stel dát er een verschil is tussen een keuze uit het verleden en de uitkomst van de afwegingsmethode voor dezelfde keuze. Is dan de afwegingsmethode fout of had in het verleden beter een andere keuze gemaakt kunnen worden?

Directe meting van de betrouwbaarheid van de methode kan dus problematisch worden. Het is echter ook mogelijk om de betrouwbaarheid van het model indirect te meten. De betrouwbaarheid van de berekeningen van scores op criteria kan ingeschat worden, voor de vaststelling van weegfactoren kunnen betrouwbaarheidsintervallen worden opgesteld, de compleetheid van de set criteria kan ingeschat worden. Een inventarisatie van alle onzekerheden in afwegingsmethoden is gemaakt in (Pouwels 1996). Al deze schattingen kunnen weer een inschatting van de totale betrouwbaarheid van het model opleveren. Ik zal dit later verder uitwerken. Hoe betrouwbaar de resultaten moeten zijn is afhankelijk van het doel waarmee de afwegingsmethode wordt aangewend.

Omdat de modellen bijna allemaal lineair zijn, komt er nauwelijks een grote nadruk op bepaalde modelonderdelen voor. Uitzondering is de verwachtingswaarde methode, het is zeer goed mogelijk dat daarbij het hoogst gewaardeerde criterium onevenredig veel aandacht krijgt.

Een ander probleem is het verlies van informatie dat optreedt bij een aantal multicriteria analyses. Het gaat daarbij om methoden waarbij kwantitatieve schalen omgezet worden in een kwalitatieve, waarbij alleen de volgorde in belang (bij weegfactoren) en de volgorde in grootte (bij gestandaardiseerde criteriumscores) worden meegenomen in de berekening. Zeker als de scores kwantitatief en precies bekend zijn, is dit een enorm verlies aan informatie. (A=10.000, B= 9.000, C=12) heeft hetzelfde resultaat als (A=10.000, B=14, C=7) als het gaat om de volgorde (Eerst A, dan B, dan C). In beide gevallen wordt evenveel belang aan B gehecht, onterecht, zoals blijkt. Een dergelijke vereenvoudiging van beschikbare informatie is alleen toelaatbaar als de andere criteria zeer onprecies bekend zijn en zwaarder meewegen.

In hoofdstuk twee zijn al een aantal vooronderstellingen (scripts) die bij het gebruik van een afwegingsmethode gemaakt worden geanalyseerd. Hoe erg het is om deze vooronderstellingen te maken is weer sterk afhankelijk van het doel waarvoor de afwegingsmethoden worden aangewend. Ook hiervoor zal overlegd worden met Rijkswaterstaat.

De keuze voor het juiste aggregatieniveau wordt vooral gemaakt bij het vaststellen van de criteria. De criteriagroepen onderkennen het probleem van het vaststellen van het juiste

aggregatieniveau. Men is met name alert op het gelijk houden van de aggregatieniveaus van de criteria van de twee verschillende groepen. Vanuit Rijkswaterstaat is aangegeven dat het mogelijk moet zijn om op verschillende niveaus af te wegen. Dit omdat men een brede doelgroep van gebruikers op het oog heeft, die elk op een ander niveau en in andere organisaties kunnen werken. Een ander belangrijk punt zijn de betalingsproblemen die op gaan treden als ruimtelijke problemen vertaald gaan worden in scores op niet-ruimtelijke criteria. De hier gepresenteerde afwegingsmethoden kunnen niet om gaan met ruimte en tijd. Dit betekent dat ruimtelijke en temporele patronen gevangen zullen moeten worden in vrij statische criteria.

De groei van zeegras in de Eems-Dollard waarbij op een gegeven moment niet alleen op de zeer geschikte maar ook minder geschikte gebieden het zeegras langzaam maar zeker uitbreidt en verdicht, onder de invloed van langdurig gunstige omstandigheden, is moeilijk te vangen in oppervlak zeegras. “Oppervlak zeegras waar?”, is de gerechtvaardigde vraag. Bij problemen waar ruimte en tijd een grote rol spelen zou het daarom misschien beter te zijn om gebruik te maken van een soort presentatiemethode, waarbij ruimte en tijd in de presentatie zijn opgenomen. Ik denk dan bijvoorbeeld aan simulatie-modellen van een gebied. Als er echter toch expliciet afgewogen moet worden dan zal er een vertaling naar statische één-dimensionale criteria plaats moeten vinden. Hoe deze vertaling precies ingekleed wordt is aan de criteriagroep.

§III.3.2 Externe criteria

§ III.3.2.1 Acceptatie

Het criterium acceptatie is een erg moeilijk criterium voor toepassing op afwegingsmethoden voor de Waddenzee, omdat nauwelijks in te schatten is wat de toekomstige gebruiker accepteert, er is immers nauwelijks een beeld van de toekomstige gebruiker. Hoe goed een model past binnen bestaande processen en instituten is niet zo relevant, omdat de afwegingsmethode juist moet leiden tot een nieuwe handelingspraktijk. In hoeverre het model als eerlijk, rationeel of ethisch verantwoord ervaren zal worden is ook moeilijk te zeggen. Feit is dat het model bij rijkswaterstaat vandaan komt en dat andere gebruiksgroepen het alleen om die reden al minder acceptabel zouden kunnen vinden. Daarom is het belangrijk om te voorkomen dat de indruk gewekt wordt dat het model de gebruiker in een richting gaat sturen. Want die richting is dan ongetwijfeld die van rijkswaterstaat en het model is dus niet eerlijk want het wordt strategisch gebruikt, enz. enz. Acceptatie vraagt dus mogelijk om een grote flexibiliteit (=weinig sturing) van het model jegens de gebruiker. Hoewel de modellen over het algemeen eenvoudig van opzet zijn²² zullen er altijd mensen zijn die de technieken niet begrijpen, niet willen begrijpen, wantrouwen of in het algemeen vijandig staan tegenover rationele expliciete afwegingen. De vraag is, hoe moeilijk mag de techniek zijn, hoeveel kennis bij de gebruiker verondersteld mag worden. Misschien is dit punt

²² In de ogen van een technisch geschoolde academicus.

de reden dat monetaire kostenbaten analyses nog zo populair zijn, ze zijn ontzettend eenvoudig: je telt gewoon alle guldens op, het alternatief met de meeste guldens levert het meeste op. In ieder geval zullen de ingewikkelder methodes, vooral als de berekeningen onzichtbaar zijn, bij de techniekwantrouwendes minder en bij de techniekgelovigen meer vertrouwen opwekken. Overigens is hier niet alleen de algemene technische (wiskundige) kennis van belang, ook de specifieke kennis van dit soort methoden doet er toe. Als mensen eenmaal aan een bepaalde methode gehecht zijn, dan is het moeilijk om een andere te gaan gebruiken.

De gebruikers zullen ook de ingebakken scripts in het model moeten accepteren. Gebruik van het model dwingt hen, soms opzettelijk, tot bepaald handelen. Door een afwegingsmethode te gebruiken moet de gebruiker zich bloot geven, bereid zijn af te wegen, gaan onderhandelen, onderzoek doen naar zijn eigen motieven, eventueel alternatieven verzinnen, er moet onderzocht, gerekend en voorspeld worden. Kortom, de gebruiker moet nog al wat. Onacceptabele scripts maken ook een methode onacceptabel.²³

Het rekening houden met de gebruiker blijft erg moeilijk, ook in het ontwerp van een afwegingsmodel, juist vanwege de onbekendheid van de gebruiker. Omdat de acceptatie door de gebruiker een grote faalfactor is en omdat deze acceptatie nauwelijks voorspelbaar is, zou het goed zijn om tijdens het gebruik nog (deskundig) te kunnen sleutelen aan de gebruikte methode om zo de gebruiker meer tegemoet te komen. Als de gebruiker zelf gaat sleutelen kan de kwaliteit van het model enorm teruglopen, omdat te weinig rekening gehouden wordt met andere kwaliteitscriteria.

§ III.3.2.2 Praktische bruikbaarheid

Omdat de modellen relatief eenvoudig zijn en slechts enkele afwegingsmethoden rekenintensief zijn, is het gebruik van een computer technisch gezien meestal niet nodig. Computers zijn echter zo ingeburgerd en de presentatie van de resultaten wordt zoveel versneld en verbeterd dat een gecomputeriseerde versie van het model wel aantrekkelijk is. Dit hoeft geen gebruikersproblemen op te leveren. De modellen zijn ook zo eenvoudig dat met behulp van een computer de berekening bijna direct klaar is. Het onderzoek dat het model vereist om het te kunnen laten draaien kost veel meer tijd. Ik denk dan met name aan het berekenen van de scores van alternatieven op criteria. Men moet bereid zijn om tijd en geld in de afweging te steken.

Een ander punt is de input. Als de criteria waarop beoordeeld gaat worden is vastgesteld, zit de gebruiker ook vast aan gevraagde input. Als deze input niet beschikbaar blijkt zou het model gewoon door moeten kunnen blijven draaien. Ook de flexibiliteit ten aanzien van de soort input is belangrijk. Monetaire kosten baten analyses kunnen alleen monetaire input aan.

²³ Overigens, het gaat hier om een andere accepteerbaarheid van scripts dan die genoemd onder het criterium logica. Daar gaat het om acceptabel in de zin van: logisch gezien in overeenstemming met de randvoorwaarden waarbinnen de methode gebruikt gaat worden. Hier gaat het om acceptabel voor de gebruiker.

Ook multicriteria analyses verschillen van elkaar in de flexibiliteit ten aanzien van de vereiste soort input.

Voor het ontworpen model geldt dat het, als het aan de gebruiker wordt gepresenteerd, inzichtelijk moet zijn, voorzien van goede documentatie over wat er gebeurt, hoe en waarom. Ook gebruikersvriendelijkheid en robuustheid zijn criteria voor een goed geaccepteerd en bruikbaar model. Mocht er gekozen worden voor een ingewikkelder model dan is er misschien een cursus “omgaan met afwegingsmethoden” nodig om verantwoord modelgebruik te bevorderen.

§ III.3.2.3 Effectiviteit

Een erg belangrijk criterium is de effectiviteit. Ook dit criterium wordt sterk beïnvloed door de wensen van de opdrachtgever. Maar op dit punt is veel meer duidelijkheid te verschaffen over die wensen. Waar acceptatie en praktische bruikbaarheid moeilijk te voorspellen zijn door het grote aantal potentiële gebruikers is de effectiviteit vooral afhankelijk van de doelstellingen van een relatief kleine groep opdrachtgevers.

Ten eerste moeten de resultaten van de afwegingsmethode natuurlijk aangewend kunnen worden voor de doelen die de opdrachtgever met de afwegingsmethode heeft. Maar er is meer. Ook de exercitie van het modelbouwen kan al nuttig zijn voor de opdrachtgever, bijvoorbeeld omdat hiermee de problematiek gestructureerd wordt, of omdat dit een reden is voor verschillende partners om rond de tafel te gaan zitten en met elkaar te praten. Aan het eind van hoofdstuk twee heb ik een opsomming gemaakt van alle mogelijke functies die het gebruik en het ontwerp van een afwegingsmethode kan vervullen. Een andere voorwaarde voor een nuttig gebruik van de afwegingsmethode is dat de criteria overeenkomen met de criteria die de gebruiker interessant vindt. Dit gaat verder dan het compleetheidcriterium. Als het model de gebruiker dwingt om na te denken over allerlei criteria die hem toch niet interesseren kost hem dit nodeloos tijd. Daarnaast is het van belang dat het model geldig is op het domein waarop de gebruiker het wil inzetten. Dit is bijvoorbeeld van belang bij linearisaties (zie hoofdstuk II).

Van grote invloed op de effectiviteit van een model is of de onzekerheden in de modeluitkomst expliciet worden weergegeven. Vooral als belangrijke keuzes op basis van het model gemaakt gaan worden of als het model gebruikt gaat worden voor voorspellingen is inzicht in de onzekerheden van de modeluitkomsten van het grootste belang. Het verschilt nogal of een model voorspelt dat alternatief 1 één miljoen kost en alternatief 2 anderhalf of dat alternatief 1 gemiddeld één miljoen kost met een variatie van achthonderdduizend tot 5 miljoen en alternatief twee precies anderhalf miljoen. Een organisatie met een budget van anderhalf miljoen en een sterk risicomijdende strategie kiest ongetwijfeld alternatief twee, op voorwaarde dat de onzekerheid over de uitkomst (één en anderhalf miljoen) wordt vermeld. Daarnaast kan een onzekerheidsanalyse samen met een gevoeligheidsanalyse ook zeer nuttig zijn voor het sturen van onderzoek. In geen enkele afwegingsmethode worden de onzekerheden expliciet weergegeven. Bij sommige afwegingsmethoden wordt wel wat aangemodderd met kansen, maar daarbij worden statistische verdelingen verondersteld die enorm van de werkelijkheid kunnen afwijken, zodat de uitspraak dat alternatief 1 in 80% van

de gevallen beter scoort dan alternatief 2 óók kan betekenen dat alternatief 1 waarschijnlijk ergens tussen de 45% en de 92% beter scoort. In het onderzoek van (Pouwels 1996) worden vijf mogelijke vormen van onzekerheid onderscheiden. Een goede afwegingsmethode zou deze vijf vormen van onzekerheid en hun invloed op het resultaat moeten inschatten en weergeven. Omdat geen enkele besproken afwegingsmethode de onzekerheid in de resultaten weergeeft moet er nog een onzekerheidsanalyse worden toegevoegd. Het criterium kan daarom beter worden vervangen door een criterium: "In hoeverre de afwegingsmethode geschikt is voor een onzekerheidsanalyse?"

Het zou prettig zijn als de methode, nadat deze ingevuld is voor een specifiek beleidsvraagstuk, ook bruikbaar is voor andere problemen. Als bijvoorbeeld de criteria of de weegfactoren overgenomen kunnen worden naar een nieuw probleem. Dit scheelt tijd en geld. Aan de andere kant betekent dit wel dat er problemen kunnen optreden met het domein waarop het model geldig is. Een zo breed mogelijke selectie van criteria zou dus kunnen bijdragen aan een brede toepasbaarheid van de afwegingsmethode. Voor de weegfactoren is dit moeilijker, omdat dit tijd- en plaatsafhankelijke variabelen zijn. Deze zouden per probleem erg kunnen verschillen. Voor de scores is overzetten al nauwelijks mogelijk, hoogstens kan opgedane ervaring en kennis gebruikt worden zodat scores sneller berekend kunnen worden.

De inzichten en conclusies uit een opgestelde afwegingsmethode zijn nauwelijks generaliseerbaar. Daarvoor is het model te specifiek voor één situatie opgesteld. Dit is, in mijn ogen, ook niet erg belangrijk. De manier van behandelen van problemen kan wel erg nuttig zijn en bruikbaar voor tal van andere situaties (mocht de methode bevallen).

Wat de koppeling van de afwegingsmethode aan andere modellen betreft, voor het gebruiksgemak is deze niet erg belangrijk. Het gaat hier niet om een groot model. Omdat het model voor de gebruiker doorzichtig moet blijven, zullen er waarschijnlijk een beperkt aantal criteria gebruikt worden. Deze kunnen daarom best met de hand ingevoerd worden, dit hoeft geen onoverkomelijk probleem te zijn. Sterker nog, door een mens als intermediair te gebruiken is het ook mogelijk om een kwaliteitsschatting van de input te maken en die mee te geven aan het model. Vanuit de afwegingsmethode geredeneerd doet de mogelijkheid tot koppeling aan andere modellen er niet zo toe. Al zou de koppeling aan een spreadsheet of een tekstverwerker in verband met de presentatie van gegevens wel makkelijk zijn, maar dat is een softwarematig probleem en is niet onderscheidend voor de keuze van een afwegingsmethode. Zoals eerder vermeld lopen parallel aan de ontwikkeling van een afwegingsmethode nog twee andere projecten, namelijk de selectie van criteria en de ontwikkeling van een beslissingsondersteunend systeem (BOS). Wat de koppeling hier aan betreft, de criteria stellen eisen aan de afwegingsmethode, deze zijn al genoemd, en aan het BOS. Het BOS moet zoveel mogelijk informatie geven over de gestelde criteria. Een koppeling andersom, waarbij de selectie van criteria afhankelijk wordt gemaakt van de technische beperkingen van een BOS lijkt mij niet verstandig, omdat juist inzicht in de niet te modelleren belangen en in de onzekerheden die deze meebrengen voor de beslissing belangrijk is voor het maken van een goede beslissing. Wat mij betreft krijgen de criteria in zoverre prioriteit dat eerst onderzocht wordt wat de belangen zijn en dat daarna pas gekeken wordt in hoeverre deze belangen voorspelbaar zijn en makkelijk zijn in te passen in afwegingsmethode en BOS. De directe koppeling van de afwegingsmethode aan het BOS lijkt me daarom niet noodzakelijk. Ten

eerste zal het BOS waarschijnlijk niet alle informatie kunnen leveren. Ten tweede hoeft ook niet alle informatie uit het BOS rechtstreeks in de afwegingsmethode gestopt te worden, een menselijk intermediair als interpretator van de gegevens kan heel nuttig zijn.

§ III.4 Conclusies en aanbevelingen

In hoofdstuk III zijn de algemene modelkwaliteitscriteria van Covello besproken in het licht van afwegingsmethoden. Een aantal criteria werden geconcretiseerd, een aantal konden worden genegeerd. Voor sommige criteria werd de opdrachtgever om invulling gevraagd, de gesprekken hierover zijn verwerkt. Het resultaat is een lijst met concrete criteria voor de keuze van een afwegingsmethode voor de opdrachtgever en een aantal aanbevelingen voor het ontwerp, de invulling en het gebruik van de nog te kiezen afwegingsmethode.

De criteria zijn de volgende:

- De argumentatie voor de methode moet niet aanvechtbaar zijn.
- De methode moet ook daadwerkelijk op de argumentatie gebaseerd zijn.
- Gemaakte vooronderstellingen moeten acceptabel zijn en passen binnen de beoogde functie van de afwegingsmethode. De scripts moeten geaccepteerd kunnen worden.
- De structuur van de afwegingsmethode moet overeenkomen met de gewenste structuur.
- Alle soorten criteria (verschillende grootheden, verschillende schalen) moeten vergeleken kunnen worden.
- De methode moet flexibel zijn, er moeten later nieuwe criteria of alternatieven toegevoegd kunnen worden.
- De betrouwbaarheid van het model.
- De mate van informatieverlies. (Ten opzichte van de precisie waarmee scores aangeleverd kunnen worden.)
- De methode moet de functies die de opdrachtgever ten doel heeft gesteld kunnen vervullen.
- Er moet een onzekerheidsanalyse toegevoegd kunnen worden.

Aanbevelingen:

- De gebruiksmogelijkheden van de afwegingsmethode, vooral voor zover het toepassingsdomein wordt beperkt door vooronderstellingen, zal bij de gebruiker bekend moeten zijn, om zo oneigenlijk gebruik te voorkomen.
- De criteria moeten zo compleet mogelijk zijn.
- Te veel (niet belangrijk gevonden) criteria echter, benadelen indirect de effectiviteit van de methode.
- Een gevoeligheidsanalyse naar de criteria voordat er onderzoek naar scores wordt gedaan kan kostenefficiëntie verhogen.

- De ruimtelijke spreiding en tijd moeten ook opgenomen worden in de criteria, bijvoorbeeld door voorkomen op gewenste plaatsen zwaarder mee te wegen dan voorkomen op minder gewenste plaatsen.
- Acceptatie kan verhoogd worden door het mogelijk te maken later nog te sleutelen aan een methode. Alleen deskundigen kunnen verantwoord sleutelen, kennis van de consequenties van veranderingen is noodzakelijk.
- Voor het ontworpen model geldt dat het, als het aan de gebruiker wordt gepresenteerd, inzichtelijk moet zijn, voorzien van goede documentatie over wat er gebeurt, hoe en waarom. Ook gebruikersvriendelijkheid en robuustheid zijn criteria voor een goed geaccepteerd en bruikbaar model. Mocht er gekozen worden voor een ingewikkelder model dan is er misschien een cursus “omgaan met afwegingsmethoden” nodig om verantwoord modelgebruik te bevorderen.
- Er moet een domein vastgesteld worden waarop de afwegingsmethode geldig is, zodat verkeerd gebruik (buiten dat domein) voorkomen kan worden.
- Om onzekerheden weer te kunnen geven is het praktischer wanneer input voor de afwegingsmethode niet exact maar met kansverdelingen worden weergegeven.
- Met onzekerheids- en gevoeligheidsanalyse kan onderzoek gestuurd worden naar de onzekerste en invloedrijkste variabelen. Daarmee kan onzekerheid zo efficiënt mogelijk gereduceerd worden.
- Een menselijke intermediair tussen rekenmodellen en afwegingsmethode kan de onzekerheidsschatting van de resultaten aanzienlijk verbeteren omdat menselijke interpretatie van de uitkomsten van een rekenmodel kan leiden tot betere inschatting van de onzekerheden in de modeloutput.

Hoofdstuk IV De toetsing van de afwegingsmethoden

In hoofdstuk II zijn een aantal modelkenmerken van afwegingsmethoden besproken. In hoofdstuk III zijn een aantal criteria ontwikkeld waarop afwegingsmethoden getoetst kunnen worden op hun kwaliteit. In dit hoofdstuk vindt de confrontatie plaats van de afwegingsmethode met hun toetsingscriteria. In de eerste paragraaf worden per criterium de relevante afwegingsmethoden besproken. In de tweede paragraaf wordt vervolgens weer per afwegingsmethode kort de criteria besproken. Dit ter vergroting van het overzicht. In de derde paragraaf volgen samenvatting en conclusies, inclusief een overzichtstabel met daarin de resultaten van dit hoofdstuk.

§ IV.1 Beoordeling per criterium.

§IV.1.1 De argumentatie

Het eerste criterium gaat om de kwaliteit van de argumentatie. De argumentatie voor de methode mag niet aanvechtbaar zijn. De meeste methoden zitten wel redelijk logisch in elkaar. In een aantal methodes blijken toch aanvechtbare bewerkingen voor te komen. Deze worden nu één voor één behandeld.

De concordantieanalyse, om te beginnen, ziet er in eerste instantie heel aardig uit. Met behulp van weegfactoren, kwalitatieve scores en paarsgewijze vergelijking wordt op twee manieren een prioriteitsvolgorde bepaald. Als deze identiek zijn, dan moet het haast wel een goede volgorde zijn. Deze laatste suggestie wordt nog eens versterkt door de naamgeving van de beide manieren, concordantie en discordantie. Het lijkt alsof het probleem op twee tegenovergestelde wijzen wordt benaderd. Maar wat gebeurt er echt? Zoals gezegd bestaat een concordantieanalyse uit twee delen, de concordantie- en de discordantie analyse. Bij de concordantieanalyse worden de ingevoerde kwantitatieve scores alleen kwalitatief gebruikt. Er wordt namelijk paarsgewijs vergeleken, waarbij alleen telt of een score hoger of lager is dan de score van andere alternatieven op één criterium. Er treedt dus een groot informatieverlies op. Bij de discordantie analyse wordt wél rekening gehouden met de kwantitatieve scores, daar echter wordt weer helemaal niets gedaan met de weegfactoren. Dit heeft tot gevolg dat relatief onbelangrijke criteria net zo veel invloed hebben op het resultaat als de belangrijkste criteria! De concordantieanalyse is dus een verkapt kwalitatieve methode en de discordantie analyse is gewoon niet compleet. Als beide methoden toevallig dezelfde prioriteitsvolgorde hebben zegt dit dus helemaal niets over de werkelijke geprefereerde volgorde.

Een ander voorbeeld van een logisch gezien aanvechtbare methode is de meerdimensionale schaalanalyse. Bij deze methode worden kwalitatieve scores omgezet in

kwantitatieve, met behulp van een grafische methode. Dit is op zich al bevreedend, want er wordt blijkbaar verondersteld dat er informatie toegevoegd kan worden aan de kennis, zonder meer kennis te vergaren. Het is onduidelijk waarom veronderstelt zou moeten worden dat rangschikken van punten op een rij iets te maken zou hebben met de veronderstelde scores. Er wordt vervolgens gesteld dat bij een groter aantal alternatieven en criteria de puntenconfiguratie stabiel wordt. Misschien wordt de puntenconfiguratie wel stabiel, maar betekent dit dat bij een groter aantal alternatieven en criteria de kennis over de scores van alternatieven en criteria groter wordt? Gelukkig is er ook een rekenvoorbeeld opgenomen in (Ministerie van Financiën 1992). Van een oorspronkelijke set scores weergegeven in onderstaande tabel wordt hier een kwalitatieve set gemaakt. Dit is niet moeilijk, verarming van de informatie vormt geen probleem.

Kwantitatief	Alternatief 1	Alternatief 2	Alternatief 3
Kosten (Mfl)	40	60	80
Reistijdwinst	2	1	3
Verlies natuur	3	1	2
Minder slachtoffers	4	5	10

Tabel 3 De oorspronkelijke scores

Kwalitatief ²⁴	Alternatief 1	Alternatief 2	Alternatief 3
Kosten (Mfl)	1	2	3
Reistijdwinst	2	1	3
Verlies natuur	3	1	2
Minder slachtoffers	3	2	1

Tabel 4 Gekwalificeerde scores

Met behulp van een ééndimensionaal ontvouwingsmodel (een speciale ééndimensionale variant van meerdimensionale schaalanalyse), worden deze kwalitatieve scores omgezet in kwantitatieve scores. Het resultaat uit (Ministerie van financiën 1992) staat in onderstaande tabel.

Weer kwantitatief ²⁵	Alternatief 1	Alternatief 2	Alternatief 3
Kosten (Mfl)	2	6	10
Reistijdwinst	2,5	1,5	0,5
Verlies natuur	5,5	1,5	2,5
Minder slachtoffers	9	5	1

Tabel 5 Met behulp van meerdimensionale schaalanalyse "berekende" kwantitatieve scores

²⁴ De getallen zijn rangordegetallen en dus in wezen kwalitatief

²⁵ Waarbij opgemerkt zij dat lage scores goed zijn en hoge scores slecht. Daardoor is bijvoorbeeld criterium 4 omgedraaid, de hoogste score is nu de laagste geworden, omdat een grotere vermindering verkeersslachtoffers beter is.

Het zal duidelijk zijn dat van de oorspronkelijke verhoudingen niets meer over is. Het meest duidelijk is dit te zien bij criterium vermindering van de slachtoffers. Waar oorspronkelijk het beste alternatief tweeënhalve keer zo goed was als het slechtste is deze verhouding nu 1 op 9. Het had nog erger gekund. Als bijvoorbeeld de oorspronkelijke scores op criterium 2 niet: 2,1,3 waren geweest maar: 200,1,201 dan was het resultaat nog steeds hetzelfde geweest. Gezien de onmogelijkheid om informatie (wet van behoud van informatie, informatica) toe te voegen zonder die ergens vandaan te halen lijkt het verstandiger om niet gebruik te maken van dit soort “opwaarderings-”methoden, vooral ook omdat de veronderstelde precisie toeneemt en een beslisser op basis van dit soort informatie wellicht met een te grote zekerheid een (verkeerde) beslissing neemt.

Dan is er nog de verwachtingswaarde methode die ook gebruikt wordt om kwalitatieve data om te zetten in kwantitatieve. Er wordt zomaar verondersteld dat scores meestal een bepaald patroon zullen volgen. Daarom moet kwalitatieve data maar gewoon in dit patroon gedrukt worden. Deze veronderstelling is speculatief. Er zou ook een lineaire schaal gebruikt kunnen worden, of juist een omgekeerde niet-lineaire schaal met veel belangrijke en weinig onbelangrijke gewichten. Eventueel is ook een S-curve mogelijk, volgens het principe van de toe- en afnemende meeropbrengsten, waardoor en relatief veel gemiddelde gewichten zijn. De keuze voor een verwachtingsschaal is in mijn ogen volstrekt arbitrair en zou alleen toegepast moeten worden als de beslisser denkt dat deze schaal overeenkomt met zijn voorkeuren. Bovendien gaat ook hier de wet van behoud van informatie op.

Tot slot zijn er nog de methoden die gebruik maken van paarsgewijze vergelijking per criterium. (Dit in tegenstelling tot de SAATY-methode, die door middel van paarsgewijze vergelijking de waard van weegfactoren bepaald.) Zoals opgemerkt in § II.3.3.6 is de voorkeursvolgorde van de alternatieven hier beïnvloedbaar door het toevoegen van een nieuw alternatief. Dit maakt de methode niet alleen uiterst gevoelig voor manipulatie, het maakt de methode ook onbetrouwbaar. Dit is het geval bij de concordantieanalyse, de regimemethode en de evamix methode.

§IV.1.2 De realisatie van de theorie in het model

Het derde criterium is dat de methode ook daadwerkelijk op de argumentatie gebaseerd moet zijn. Dit criterium speelt met name bij de multicriteria methoden waar gestandaardiseerd moet worden. Zoals opgemerkt in § II.3.3.3 wordt formeel door standaardisatie vergelijking van “onvergelijkbare” criteria mogelijk, terwijl in de praktijk de standaardisatie niets verandert aan het feit dat “onvergelijkbare” criteria met elkaar vergeleken worden, maar dit alleen verdoezelt. Tegelijkertijd is echter beargumenteerd dat deze theorie ook helemaal niet nodig is en dat de veronderstelde onvergelijkbaarheid in werkelijkheid helemaal geen probleem hoeft te zijn. Daarmee vervalt dus de uitvoering en de noodzaak van de argumentatie. Het enige wat daardoor verandert is dat de “plicht” tot standaardisatie vervalt en dat standaardisatie ten zeerste wordt afgeraden, tenzij alle informatie over de standaardisatie gebruikt wordt bij het toekennen van gewichten of er gebruikt wordt gemaakt van standaardisatie naar streefwaarden, zoals dat bijvoorbeeld bij de AMOEBE het geval is.

Een andere methode die hier genoemd moet worden is de kostenbaten analyse. Formeel is deze methode gebaseerd op het Pareto-criterium. Pareto stelt in zijn welvaartstheorie dat als “het nut” van een individu toeneemt én het nut alle andere individuen niet afneemt, neemt de welvaart toe. Op zich is dit een heel redelijk criterium, vooropgesteld dat je toenemende jaloezie niet accepteert als een achteruitgang van de welvaart. Maar tegelijkertijd is het niet een heel bruikbaar criterium, omdat er alleen een uitspraak gedaan kan worden over de welvaart in win-win situaties. Als de één wint en de ander verliest, kan met behulp van het Pareto-criterium niets gezegd worden. Daarom wordt het criterium vaak opgerekt; er is ook sprake van optimale situatie als de verliezer gecompenseerd wordt door de winnaar, en de winnaar na compensatie er nog steeds op vooruit gaat. Vervolgens wordt gesteld dat in een KBA deze compensatie niet opgenomen hoeft te worden, een potentiële compensatie is voldoende. Dit wordt de “potentiële compensatietest van Kaldor en Hicks” genoemd, (Nieuwkamer 1995). Omdat in de KBA vervolgens op geen enkele wijze deze compensatie naar voren komt en deze ook niet bij de alternatieven genoemd wordt, blijkt er in de praktijk helemaal niets meer van het Pareto-criterium terecht te komen, waar onder andere een aantal van de problemen in §II.3.1 genoemd uit voortkomen. Met name het ontbreken van een rechtvaardige verdeling van welvaart, dat in het oorspronkelijke Pareto-criterium tot op zekere hoogte nog gewaarborgd wordt, vervalt. Bij Pareto kunnen de rijken wel rijker worden, maar nooit ten koste van de armen.²⁶ Bij KBA is dit zeer goed mogelijk op voorwaarde dat de rijken meer rijker worden, dan de armen armer. Aan de andere kant is het Pareto-criterium erg conservatief: als in het verleden een verdelingsfout is gemaakt, dan kan deze niet teruggedraaid worden zonder het Pareto-criterium te schenden, tenzij de totale productie toeneemt. Strikt vasthouden aan Pareto betekent bijvoorbeeld dat er niets gedaan kan worden om een eerder milieuschadelijk handelen te beknotten, tenzij dit ook positief is voor de gebruiker of dit betaald wordt vanuit een toename van het BNP. Concreet betekent dit dat een verbod op lozing van bepaalde giftige chemicaliën wel een enorme verbetering van de milieukwaliteit geeft, maar tegelijkertijd een schending van het Pareto-criterium betekent, omdat de welvaart van de fabrikant achteruit gaat, omdat hij nu extra kosten moet maken. De compensatietest maakt KBA een stuk minder conservatief, wat, mits goed gebruikt, zeker als een voordeel gezien mag worden.

§IV.1.3 Vooronderstellingen en Scripts

Gemaakte vooronderstellingen moeten acceptabel zijn en passen binnen de beoogde functie van de afwegingsmethode. De scripts moeten geaccepteerd kunnen worden. Omdat scripts in feite dwingende gevolgen zijn van gemaakte vooronderstellingen is de kwaliteit van een model in grote mate afhankelijk van de accepteerbaarheid van de scripts in het model. Deze scripts zijn besproken in hoofdstuk II. Voor de monetaire evaluatiemethoden zijn twee scripts van belang. Deze zijn: “Geld is het enige criterium of alle andere criteria zijn in geld uit

²⁶ Overigens, de armen kunnen ook niet minder arm worden ten koste van de rijken, Pareto gaat altijd uit van de status-quo.

te drukken.” en “Al het geld is evenveel waard.” Beide scripts leveren problemen op. Zoals eerder aangegeven kan voor de problematiek in de Waddenzee niet alleen geld als criterium gebruikt worden, ook niet na omrekening. Daarnaast is er het script dat alle geld evenveel waard zou zijn. Ook dat gaat niet op voor de Waddenzee. Een aantal afwegingen zouden niet bepaald moeten worden door een markt, maar op democratische wijze. Toch zou een goede analyse van de kosten en baten heel nuttig zijn. Maar dan alleen op die criteria op de markt gewogen zouden moeten worden. De scripts van de kostenbaten analyse beperkt deze methode dus tot een gedeelte van de afweging.

Voor presentatiemethode geldt het script dat de beslisser zelf in staat is om zijn afweging te maken. De afweging hoeft daarom niet expliciet gemaakt te worden, een presentatie van de gegevens is voldoende. Uit gesprekken met Rijkswaterstaat is gebleken dat ook dit script niet acceptabel is. Juist de expliciete integratie van verschillende criteria is belangrijk. Wel heel nuttig is het concept van een analyse van de gevolgen voor verschillende actoren zoals die bij de community impact analyse naar voren komt. Juist met behulp van een dergelijke analyse kan de rechtvaardigheid van een alternatief meer aandacht krijgen.

Voor de multicriteria methoden zijn de meeste scripts beschreven. De eerste script, en deze is geldig voor alle multicriteria methoden, is dat alle criteria vergelijkbaar zijn. Dit script kan grote problemen geven als een multicriteria methode wordt toegepast op een situatie waar criteria voorkomen die niet vergelijkbaar zijn. Ik denk bijvoorbeeld aan het criterium bescherming tegen overstroming in de Waddenzee. De bevolking moet hoe dan ook beschermd worden tegen overstroming. Als veiligheid echter als criterium mee wordt genomen in een afwegingsmethode, dan moet ineens veiligheid afgewogen worden tegen andere criteria. Dit maakt het in potentie mogelijk om niet het verplichte veiligheidsniveau te waarborgen omdat dat “meer” oplevert. Dit kan natuurlijk voorkomen worden door de weegfactor voor veiligheid extreem hoog te kiezen, maar dit heeft weer tot gevolg dat het veiligste ontwerp per definitie gekozen wordt en de andere criteria in feite geen invloed meer hebben tenzij er twee ontwerpen zijn met exact dezelfde veiligheid. Toch is dit probleem vrij eenvoudig op te lossen. De criteria moeten dan gesplitst worden in afweegbare criteria en harde criteria (die waaraan de oplossing in ieder geval moet voldoen). De alternatieven worden dan eerst getest op de harde criteria, van de alternatieven die daaraan voldoen kan vervolgens de prioriteitsvolgorde bepaald worden met behulp van een MCA. Dezelfde methode kan trouwens ook goed toegepast worden op de andere afwegingsmethoden.

Multicriteria methoden maken altijd gebruik van weegfactoren. Dat dwingt de gebruiker dus ook tot het vast stellen van gewichten. Dit is niet makkelijk. Zeker niet als de keuze voor weegfactoren later verantwoord moet worden. Maar weegfactoren zijn natuurlijk noodzakelijk voor een goed model. Een model zonder weegfactoren veronderstelt namelijk dat alle criteria even belangrijk zijn. Bij een KBA worden de weegfactoren bepaald door de markt. Dit maakt de keuze voor een KBA een stuk aantrekkelijker, omdat dan altijd verwezen kan worden naar de markt of naar onderzoek naar de markt. Maar ook voor weegfactoren kan onderzoek gedaan worden. Zoals in paragraaf II.4.3.2 gesteld kan de SAATY-methode de keuze van gewichten een stuk vereenvoudigen. Bovendien kan hiermee een indicatie gegeven worden van de consistentie van de gewichtenset. Dit kan de beslisser een stuk meer duidelijkheid geven over zijn eigen voorkeuren.

Methoden die gebruik maken van een de omrekening van kwalitatieve naar kwantitatieve criteria zouden niet gebruikt moeten worden. De veronderstelling dat dit überhaupt mogelijk of wenselijk zou zijn kan niet geaccepteerd worden. Omdat dit punt al bij andere criteria behandeld wordt ga ik er nu niet verder op in.

Veel multicriteria methoden maken gebruik van standaardisatie. Zoals eigenlijk al duidelijk geworden is uit paragraaf II.4.3.4 is standaardisatie alleen acceptabel als duidelijk blijft hoe er gestandaardiseerd wordt. Standaardisatie heeft alleen nut om te schalen, en daarmee de vergelijkbaarheid te vereenvoudigen, of om de beslisser te helpen scores in perspectief te zien. Met name de “Goals achievement matrix methode” is hierop gericht. Bij kwalitatieve methoden is het erg moeilijk om exact de betekenis van een weegfactor te achterhalen. Door de extra bewerkingen is een weegfactor niet direct te koppelen aan een verhouding tussen twee scores. Dat hoeft echter geen probleem te zijn omdat kwalitatieve methoden toch al geen exacte antwoorden geven. De significantie van de antwoorden zullen zo laag zijn, dat een indicatie van de betekenis van een weegfactor genoeg is om de weegfactor vast te stellen. Extra moeilijk wordt het bij de evamix methode, vanwege de dubbele standaardisatie. Maar ook daar is het in principe wel mogelijk om de verhoudingen tussen de criteria te achterhalen. Dit heeft echter als nadeel dat de evamix methode helemaal niet meer gebruikt wordt zoals hij bedoeld is. De dubbele standaardisatie komt voort uit de behoefte om kwalitatieve en kwantitatieve criteria gescheiden te beoordelen. Door nu terug te rekenen wat de oorspronkelijke verhoudingen waren tussen kwalitatieve en kwantitatieve criteria, moeten als nog kwalitatieve en kwantitatieve criteria met elkaar vergeleken worden.

De meeste MCA methoden zijn lineair. Dit kan problemen opleveren bij grote verschillen in scores tussen alternatieven. Het kan zijn dat weegfactoren en scores afhankelijk zijn van elkaar, en dat een grote verandering van een score dus ook een verandering van een weegfactor betekent. Maar, zoals eerder gesteld in paragraaf II.4.5 hoeft dit geen probleem te zijn zolang er een domein vastgesteld wordt waarop weegfactoren geldig zijn. Vallen scores buiten dat domein, dan moet voor die score een nieuwe weegfactor vastgesteld worden. Zolang rekening gehouden wordt met het domein waarop weegfactoren geldig zijn is linearisatie acceptabel.

Dan zijn er nog de methoden die gebruik maken van paarsgewijze vergelijking per criterium. (Dit in tegenstelling tot de SAATY-methode, die door middel van paarsgewijze vergelijking de waarde van weegfactoren bepaalt.) Zoals opgemerkt in paragraaf II.4.3.8 is de voorkeursvolgorde van de alternatieven hier beïnvloedbaar door het toevoegen van een nieuw alternatief. Dit maakt de methode niet alleen gevoelig voor manipulatie, het maakt de methode ook onbetrouwbaar. Dit is het geval bij de concordantieanalyse, de regimemethode en de evamix methode. Methodes die gebruik maken van paarsgewijze vergelijking kunnen daarom beter niet gebruikt worden.

Tot slot is er de black-box berekening. Als deze methoden al gebruikt worden, dan zal in ieder geval nagegaan moeten worden dat ze werkelijk betrouwbaar zijn. Zoals eerder aangetoond is dit bij de meerdimensionale schaalanalyse niet het geval. Ook de SAATY-methode is vrij ondoorzichtig in de berekening, maar de uitkomsten zijn hier beter. Van de goals-achievement matrix methode is het nog niet duidelijk hoe er precies gerekend moet worden. Bovendien is de integratie van de verschillende actoren ook nog een ondoorzichtig

proces. Zijn alle actoren even belangrijk, of moet ook aan hen weegfactoren toegekend worden?

§IV.1.4 De structuur van de afwegingsmethode

Het vierde criterium stelt dat de structuur van de afwegingsmethode moet overeenkomen met de gewenste structuur. Met de structuur van een methode wordt hier bedoeld de manier waarop de besluitvorming plaatsvindt als gevolg van gebruik van de methode. Er worden daarbij twee dimensies onderscheiden, namelijk: de manier waarop omgegaan wordt met verschillende belangengroepen, de actoren en de manier waarop de integratie tussen de verschillende criteria plaatsvindt.

In de eerste dimensie worden twee opties onderscheiden, onder de eerste optie vallen alle methodes waarbij alleen de criteria naast elkaar gezet worden. Bij de tweede methode wordt niet alleen naar de criteria gekeken maar ook naar de verdeling van de gevolgen over de actoren en de waardering van verschillende criteria door actoren. Hier moeten niet alleen de criteria geïntegreerd worden maar ook de actoren. Er wordt verondersteld dat de actoren zelf de criteria integreren, wie vervolgens de actoren integreert (Wat is belangrijker dat de visserij tevreden is of de milieubeweging?) wordt bepaald door de tweede dimensie. Deze tweede dimensie maakt onderscheidt tussen impliciete en expliciete integratie. Impliciete integratie wordt altijd verricht door de beslisser, en is kenmerkend voor presentatiemethoden. Expliciete integratie kan verricht worden door de beslisser, op basis van berekeningen en metingen aan de markt en op basis van de mening van burgers. Marktberekeningen gaan daarbij uit van individuele behoeftes en een individuele bestedingsruimte. Als de beslisser of de burger (met behulp van enquêtes) integreert, wordt er uitgegaan van een collectieve uitgave (want gedaan door de overheid) en men gaat dus ook uit van een collectieve behoefte. De methoden worden dan als volgt ingedeeld (ZOZ):

	Impliciete integratie	Expliciete integratie		
	Collectief door beslisser	Collectief door beslisser	Individueel door de markt	Collectief door burgers
Alleen op basis van criteria	Scorekaart, AMOEBE	MCA m.u.v. Goals Achievement	KBA en KEA	MCA m.u.v. Goals Achievement.
Op basis van criteria per actor	Community impact	Goals Achievement	Onwaarschijnlijk ²⁷	Onwaarschijnlijk ²⁶

Tabel 6 De mogelijke structuren

De structuur die de beslisser gewenst acht moet vervolgens bepalen welke methode gekozen wordt.

§IV.1.5 Alle soorten criteria moeten vergeleken kunnen worden.

Alle soorten criteria (verschillende grootheden, verschillende schalen) moeten vergeleken kunnen worden. De kostenbaten en de kosteneffectiviteit analyse kunnen alleen met monetaire criteria omgaan. Andere criteria moeten omgerekend worden. Zoals opgemerkt aan het einde van hoofdstuk twee verschillen ze daarin niet eens zo heel veel van een gewogen somming. Alleen criteria die niet omrekenbaar zijn in geld (waarvoor geen markt is en ook geen markt is te simuleren) passen niet in een kostenbaten analyse. De AMOEBE kan ook niet omgaan met kwalitatieve informatie. Hiervoor geldt eigenlijk hetzelfde als voor kwantitatieve multicriteria methoden. Deze opmerkingen worden direct hieronder gemaakt.

Een aantal multicriteria methoden kunnen alleen met kwantitatieve gegevens overweg. Dit zijn de gewogen somming en de concordantieanalyse. Dit hoeft geen probleem te zijn. Voorwaarde is dan dat er gewerkt wordt met een onzekerheidsanalyse. Er kan dan een schatting gemaakt worden van waarde van de kwalitatieve scores, waarbij een vrij groot onzekerheidsinterval aan de score wordt toegekend. Voorkeursvolgordes kunnen vervangen worden door bijvoorbeeld beoordelingscijfers (bijvoorbeeld op een schaal van 1 tot 10), een cijfer 8 betekent dan van 7 tot 9, of nog veiliger, van 6 tot 10. Deze methoden vragen dus om kleine aanpassingen in de input, maar dat hoeft geen groot probleem te zijn. Een aantal methodes zijn in feite zuiver kwalitatief, maar omdat kwantitatieve informatie altijd kwalitatief

²⁷ Deze twee onwaarschijnlijke combinaties zijn theoretisch wel mogelijk, maar zullen waarschijnlijk nauwelijks praktische relevantie hebben. De markt laten bepalen hoe belangrijk actoren zijn wordt bijvoorbeeld min of meer toegepast bij het tientjeslidmaatschap van publieke zenders. De collectieve bepaling door burgers met behulp van kijkcijfers. Maar zenders zijn nog redelijk vergelijkbaar, hoe burgers dat bij actoren in de Waddenzee zouden moeten doen...

te maken is, kunnen deze methoden de kwantitatieve informatie best aan. Dit geeft natuurlijk wel informatieverlies, maar dat wordt elders in dit hoofdstuk al besproken.

§IV.1.6 Flexibiliteit ten aanzien van de input.

De bruikbaarheid van een methode neemt enorm toe als hij flexibel is ten aanzien van de input. Er zijn een aantal methoden die telkens weer opnieuw moeten beginnen met rekenen als de input verandert. Dit kan lastig zijn. Voor een computer maakt het niet uit, de meeste methoden zijn zo eenvoudig dat het verschil in rekentijd nauwelijks waarneembaar zal zijn op de computer. De kwalitatieve methoden zijn veel minder gevoelig voor de input. Zolang er een voorkeursvolgorde aangegeven wordt, werkt de methode. Er hoeft dan ook veel minder onderzoek gedaan te worden. In die zin zijn de methoden een stuk flexibeler. Verder lijkt dit punt mij niet erg onderscheidend. Zeker niet gezien de grote nadruk van het thema “gebruik” op kennisystemen.

De meeste methodes laten nieuwe criteria en alternatieven wel toe. Bij sommige methodes betekent dit wel dat alle al gemaakte berekeningen opnieuw gedaan moeten worden. Vooral bij ingewikkelde berekeningen of bij niet geautomatiseerd en niet automatiseerbaar werk is dit een nadeel. Bij andere methoden kunnen de nieuwe criteria of alternatieven gewoon toegevoegd worden.

Kostenbaten analyses en presentatiemethoden kunnen zonder problemen nieuwe alternatieven of criteria aan. Er zijn een aantal multicriteria analyses die wel problemen geven bij nieuwe criteria. Dit geldt voor alle methoden waarbij gekozen wordt voor een standaardisatie van criteria of weegfactoren, als deze standaardisaties afhankelijk zijn van minimum of maximum rij- of kolomwaarden of bij standaardisaties waarbij de totale waarde van de rij of kolom gelijk moet zijn aan, meestal, één. Het toevoegen van een alternatief kan hier in sommige gevallen zelfs leiden tot een nieuwe voorkeursvolgorde in de oorspronkelijke alternatieven! De concordantieanalyse zal ook over gedaan moeten worden, dat wil zeggen, het discordantiegedeelte. Het concordantiegedeelte blijft onveranderd. Voor een computer is dit geen enkel probleem. Ook bij de verwachtingswaarde methode, de evamix methode, een meerdimensionale schaalanalyse, moeten alle berekeningen opnieuw gemaakt worden. Ook hier is dat in een geautomatiseerde versie geen probleem.

De enige methode die echt meer tijd gaat kosten is de SAATY-methode, vanwege de opgebouwde structuur en de paarsgewijze vergelijking. De methode is al bewerkelijker dan de meeste andere methoden, dit wordt veroorzaakt door de extra aandacht die gegeven wordt aan het vaststellen van weegfactoren. Als er vervolgens aan de criteriastructuur gesleuteld wordt, moet een deel van de weegfactorenprocedure weer herhaald worden. Het volgende voorbeeld maakt dit duidelijk. Stel er is een criteriaboom ontworpen waarbij de eerste ordecriteria bijvoorbeeld milieu, economie en sociale structuur zijn opgenomen, en milieu bestaat uit de elementen bodemkwaliteit, waterkwaliteit, kwaliteit van de lucht en kwaliteit van de biosfeer. Als dan voor alle criteria weegfactoren zijn vastgesteld en de hele boom is doorgerekend en wil men dan bij milieu geluidsoverlast voor bewoners toevoegen, dan moet er niet alleen opnieuw weegfactoren voor de onderdelen van milieu vastgesteld worden, maar de verhouding tussen milieu, economie en sociale structuur zal óók veranderen, omdat milieu nu

niet alleen natuurlijk milieu is maar ook menselijk milieu en zo meer betekenis heeft gekregen en dus ook zwaarder zal moeten wegen.

§IV.1.7 De betrouwbaarheid van het model.

Een belangrijk criterium voor de kwaliteit van een model is zijn betrouwbaarheid. (Pouwels 1996, pag. 69) noemt in zijn onderzoek naar de bruikbaarheid van prioriteitsstellingsmethoden voor integraal waterbeheer vijf oorzaken van onbetrouwbaarheid. Deze zijn achtereenvolgens:

- Maatregel- en criteriumonzekerheid
- Effectonzekerheid
- Gewichtonzekerheid
- Methodegevoeligheid
- Interpretatieonzekerheid

Al deze onzekerheden en gevoeligheden hebben effect op de betrouwbaarheid van het model. Ze zijn alleen lang niet allemaal onderscheidend. Ze worden in deze paragraaf elk apart besproken

§ IV.1.7.1 Maatregel- en criteriumonzekerheid

De maatregel- en criteriumonzekerheid ontstaat vooral doordat niet alle maatregelen (alternatieven) en criteria zijn meegenomen in het model. De keuze van het model is hier niet van invloed, behalve in de gevallen waarbij een model niet om kan gaan met bepaalde criteria of alternatieven. Dit punt is al opgenomen in het criterium: “alle criteria moeten opgenomen zijn.”

§ IV.1.7.2 Effectonzekerheid

De effectonzekerheid ontstaat doordat de effecten van maatregelen niet exact voorspeld kunnen worden. Zeker in biologische of sociale systemen is kan deze onzekerheid groot worden. Maar ook deze onzekerheid is onafhankelijk van de soort afwegingsmethode. Wel kan het zijn dat een bepaalde methode om meer en (daarmee meestal onbetrouwbaardere) informatie vraagt. Dit geldt bijvoorbeeld voor de kostenbaten analyse. Daarvoor is een uitgebreid onderzoek nodig waarbij ook nog eens de data zo verwerkt moet worden dat er één integrale weefactor (de prijs) uit wordt gehaald²⁸. Natuurlijk is het eenvoudiger om de beslisser zelf een waardering voor de verschillende elementen te geven. En waarschijnlijk is

²⁸ Waarbij zij opgemerkt dat er, qua betrouwbaarheid, een enorm verschil is tussen op de markt ontstane prijs en de bij een onderzoek ingeschatte waarde of prijs van een niet aan marktwerking onderhevig produkt.

deze inschatting ook betrouwbaarder. Maar als het gaat om de marktwaarde in plaats van om de persoonlijke voorkeur van de beslisser, dan is de KBA wel geschikter.

Een ander belangrijk punt is de mate waarin fouten veroorzaakt door de onzekerheid over een effect gepropageerd worden in het eindresultaat. Als de kosten in werkelijkheid vijftig procent hoger blijken te liggen, hoeveel invloed heeft dit op het resultaat. Waarschijnlijk is de afwijking hier minder dan vijftig procent. Maar het kan zijn dat bij het ene model deze afwijking 20% is en bij het andere 5%. De propagatie van de fout is dan veel kleiner in het tweede model. Een robuuster model is minder gevoelig voor dit soort fouten. Anderzijds kan deze robuustheid ook als een gebrek gezien worden, het model is dan blijkbaar niet in staat om onderscheid te maken tussen kleine verschillen. Deze gevoeligheid wordt natuurlijk voor een heel groot deel bepaald door de ingestelde weegfactoren. Het eindresultaat is gevoeliger voor effecten die zwaar meewegen. Omdat robuustheid ten aanzien van de effecten dus niet noodzakelijk een positieve eigenschap is, is het beter om door middel van een gevoeligheidsanalyse overall de gevoeligheid voor afwijkingen te onderzoeken. Deze kan standaard bij een gecomputeriseerd model meegeleverd worden. Effecten die een grote invloed op het eindresultaat hebben kunnen dan preciezer geschat worden om zo de onzekerheid over het resultaat te verkleinen. Kwalitatieve methoden versterken fouten enorm of negeren ze volkomen. Kwalitatieve methoden zijn discreet. Stel dat de score van variant A net een heel klein beetje beter is dan die van variant B op criterium 1. Een kleine fout in variant A of B, waardoor B beter wordt dan A, zorgt dan al voor relatief grote afwijkingen. Als A duidelijk veel beter is dan B dan kan heeft een flinke afwijking in A geen enkel effect. Discrete²⁹ (kwalitatieve) methoden zouden daarom alleen ingezet moeten worden bij duidelijke verschillen in scores, óf het moet mogelijk zijn om een even goede score neer te zetten, dus niet verplicht rangschikken.

Al met al is deze vorm van onzekerheid niet echt onderscheidend. Er zijn wel verschillen tussen methoden, maar deze worden veroorzaakt door verschillende doelstellingen. Alle methoden die marktwerking simuleren worden minder betrouwbaar door een toenemende effectonzekerheid. Waar het om de propagatie van fouten gaat zijn kwalitatieve methoden mogelijk slechter, mogelijk beter dan andere methoden. In ieder geval zijn ze onzekerder.

§ IV.1.7.3 gewichtonzekerheid

Gewichtonzekerheid speelt op het eerste gezicht alleen bij MCA en niet bij KBA en presentatiemethoden. Toch is dit niet het geval. Bij KBA is er onzekerheid over de waarde van een score van een criterium in geld uitgedrukt. Is er voor een bepaalde vorm van dijkonderhoud 1000 manuren per jaar nodig, dan is het niet direct duidelijk hoeveel dit gaat kosten, zeker niet over een paar jaar. Deze kosten per manuur hebben een zelfde functie als een gewicht bij een MCA zou hebben, dus er is wel degelijk sprake van gewichtonzekerheid. Alleen is deze bij een KBA al geïntegreerd in de score, zodat in feite de effectonzekerheid bij

²⁹ Hier in de zin dat de scores discreet zijn. Dit in tegenstelling tot het onderscheid dat in de inleiding van hoofdstuk II gemaakt wordt tussen discreet en continu waar het de oplossing betreft.

KBA structureel hoger is dan bij MCA. Dit wordt vervolgens gecompenseerd door de afwezigheid van een gewichtonzekerheid

Maar niet alleen bij KBA speelt gewichtonzekerheid mee. Zelfs voor presentatiemethoden is dit het geval. Ter illustratie het volgende voorbeeld:

Presentatie A	Alternatief I	Alternatief II
Kosten	fl. 1.000.000	fl. 1.400.000
Milieuschade	Groot	Beperkt
Werkgelegenheid	10 manjaar	12 manjaar

Tabel 7 Scoretabel met weinig aandacht voor het milieu.

Presentatie B	Alternatief I	Alternatief II
Kosten	fl. 1.000.000	fl. 1.400.000
Aantasting van het landschap	Groot	Beperkt
Horizonvervuiling	Veel	Geen
Verstoringsen	Veel	Beperkt
Grondwatervervuiling	Klasse III	Geen
Geluidsoverlast	60dB	35dB
Biodiversiteit	Klein	Groot
Werkgelegenheid	10 manjaar	12 manjaar

Tabel 8 Scoretabel met veel aandacht voor het milieu.

Beide overzichten laten dezelfde situatie zien. Maar ongetwijfeld zal de detaillering van het begrip in de tweede presentatie milieuschade in de tweede situatie ook meer de aandacht vestigen op de milieuproblemen, waarmee toch het gewicht van dit criterium versterkt wordt. In de keuze van de detaillering van criteria wordt dus ook al gekozen voor het gewicht dat de criteria zullen krijgen. Toch zal het effect van gewichtonzekerheid minder sterk meespelen bij presentatiemethoden, omdat het door überhaupt niet de bedoeling is om een voorkeursvolgorde te berekenen. De uitzondering is ook hier weer de AMOEBE. De grafische representatie suggereert dat alle criteria even zwaar meetellen. Dit kan een grote afwijking zijn van de werkelijkheid. Het kan met name problemen geven met de (onbewuste) interpretatie van de gepresenteerde gegevens. Dit punt valt ten dele ook onder interpretatieonzekerheid, maar zal daar niet nog eens behandeld worden.

De MCA die gebruik maken van kwalitatieve gewichten bieden weinig inzicht in de grootte van gewichten. De onzekerheid over de precieze waarde van een gewicht is vrij groot, daarom wordt er geen precieze waarde gegeven. Wat de kwantitatieve methoden betreft, met de verwachtingswaarde methode waarbij gewichten “berekend” worden zullen een grote gewichtonzekerheid hebben, omdat de verdeling volstrekt willekeurig gekozen wordt. Eén methode moet hier nog genoemd worden, de SAATY-methode. Het toekennen van gewichten kan heel lastig zijn en erg arbitrair overkomen. Een onderdeel van de SAATY-methode bestaat uit het vereenvoudigen van dit toekennen van gewichten door middel van

paarsgewijze vergelijking. De methode wordt beschreven in paragraaf II:4.3.2: weegfactoren. Omdat de methode het toekennen eenvoudiger maakt én de consistentie van de invoer gecheckt kan worden, zal de SAATY-methode, waar het het vaststellen van gewichten betreft, waarschijnlijk tot betrouwbaarder resultaten kunnen leiden.

Extra gevoelig is de goals-achievement methode. Deze vraagt om een inschatting van de weegfactoren die andere actoren zouden toekennen aan de criteria. Het zal duidelijk zijn dat deze gewichten moeilijker te achterhalen zijn of, indien ze door anderen worden geschat, onbetrouwbaarder zullen zijn dan gewichten die de beslisser zelf invult. Daarom zullen de gewichten die in een goals-achievement methode gebruikt worden minder betrouwbaar zijn dan die in bijvoorbeeld een gewogen sommering nodig zijn. Is de beslisser echter sterk gericht op de andere actoren en is het voor de kwaliteit van de beslissing van belang dat een beslisser rekening houdt met hun wensen, dan zal het uiteindelijke resultaat van een goals-achievement methode wel beter zijn.

§ IV.1.7.4 Methodegevoeligheid

Ook methodegevoeligheid is niet onderscheidend. Methodegevoeligheid is juist kenmerkend voor alle methoden. Zowel in (van der Veen 199?) als in (Pouwels 1995) wordt gesuggereerd dat methodegevoeligheid kenmerkend is voor MCA. Dit in tegenstelling tot KBA. Dit komt echter voort uit de vooronderstelling dat MCA een groep methoden is en KBA één methode. De groep MCA methoden zouden allemaal hetzelfde antwoord moeten geven, een KBA geeft maar één antwoord. MCA geven echter niet dezelfde uitkomst en zijn dus methodegevoelig, KBA niet. Echter de uitkomst van een KBA is sterk afhankelijk van de gekozen waarderingmethode (zie §II.3.1) en is dus wel degelijk methodegevoelig. Ten tweede is er geen enkele reden om MCA als een groep te beschouwen die telkens tot hetzelfde antwoord zouden moeten komen, maar de KBA daar buiten te laten staan. KBA is natuurlijk net zo methodegevoelig als MCA, want KBA kan tot andere uitkomsten leiden dan andere afwegingsmethoden. De enige conclusie is die hieruit kan volgen is dat alleen presentatiemethoden niet methodegevoelig zijn, ze zijn eerder methodeloos (waar het om prioriteitsvolgorde gaat.) Alle andere methoden zijn methodegevoelig.

§ IV.1.7.5 Interpretatieonzekerheid

De laatste vorm van onbetrouwbaarheid ontstaat door interpretatieonzekerheid. Het is onduidelijk hoe de uitkomsten van de methode geïmplementeerd moeten worden in bestaand of nieuw beleid. Het moment waarop de afwegingsmethode in het afwegingsproces gebruikt wordt en de functie die de hij in dat proces moet vervullen zijn van belang. Naarmate de gekozen afwegingsmethode beter past bij de rol die hij moet vervullen zal de interpretatieonzekerheid kleiner worden. Het vervullen van een functie wordt echter al besproken bij een volgend criterium. Voor een goede interpretatie is het belangrijk dat de afwegingsmethode ook precies weergeeft wat de resultaten van de afwegingsmethode zijn. Een

te grote precisie in de presentatie leidt tot een te grote zekerheid bij de beslisser, een te lage precisie totodeloze onzekerheid. Nu is geen enkele in (Ministerie van Financiën 1992) of in (Pouwels 1995) gepresenteerde afwegingsmethode voorzien van een degelijke onzekerheidsanalyse. Geen van de methoden laat dan ook zien hoe groot de precisie van de uitkomsten is. Daarom is de afwezigheid van een indicatie van de precisie niet onderscheidend. Dit punt zal weer ter sprake komen bij de behandeling van het criterium: “Er moet een onzekerheidsanalyse toegevoegd kunnen worden.”

Toch zijn er een aantal methoden die een veel grotere precisie suggereren dan ze in werkelijkheid hebben. Dit geldt voor de kostenbaten analyses voor zover er in de presentatie geen rekening gehouden is met de significantie van de cijfers. De kostenbaten analyse komt toch al betrouwbaarder over vanwege de brede acceptatie van economische wegen van belang.

Ik kan mij voorstellen dat een presentatie waarbij het ene alternatief duidelijk duurder is dan het andere betrouwbaarder overkomt dan een presentatie waarbij het ene alternatief minder punten heeft gekregen dan het andere. Durder suggereert toch dat er een prijs betaald moet gaan worden. Ik spreek hier het vermoeden uit dat het appèl dat uitgaat van een kostenbaten analyse sterker is dan van een gemiddelde MCA. Het zal, mijns inziens, moeilijker zijn om een duurder dan om een slechter scorend alternatief te kiezen. Als dat het geval is, dan zal een KBA ook een hogere precisie moeten hebben, wil de methode dit appèl waar kunnen maken. Over deze precisie is echter weer weinig te zeggen, hoogstens dat deze bij geen enkele methode erg groot zal zijn, als het gaat om een complexe keuze als in de Waddenzee. Omdat het hier slechts om een vermoeden gaat zal ik de laatste opmerkingen niet meenemen in de afweging.

Ook veel multicriteria methoden suggereren een veel grotere precisie dan ze werkelijk kunnen garanderen. Dit is het meest duidelijk bij de meervoudige schaalanalyse en de verwachtingswaarde methode, omdat deze kwantitatieve informatie suggereren terwijl in werkelijkheid gebruik is gemaakt van kwalitatieve data. Een heel sterk voorbeeld hiervan is de regimemethode, waarbij een “analytische oplossing” als mogelijkheid wordt gesuggereerd, terwijl ook de numerieke oplossing een analytische kansberekening suggereert die in feite gebaseerd is op aannames die in de meeste gevallen ver van de werkelijkheid af zullen liggen. Voor een bespreking hiervan zie bijlage 4. Maar ook de evamix methode maakt gebruik van kwalitatieve data terwijl de uitkomst kwantitatief lijkt (tot op drie cijfers significant?!). Deze methoden hebben dus een kleinere interpretatiebetrouwbaarheid, de kans dat ze verkeerd geïnterpreteerd worden is groter.

§IV.1.8 De mate van informatieverlies.

Bij modelgebruik mag niet te veel informatieverlies optreden. Informatieverlies treedt alleen op bij onverantwoorde standaardisatie en bij multicriteria methoden die ontworpen zijn voor kwalitatieve criteria die gebruikt worden voor kwantitatieve criteria. De gebruikte criteria

zullen zeker niet alleen kwalitatief zijn³⁰. De methoden waarmee informatie verloren gaat zijn: de regimemethode, de verwachtings waardemethode, permutatiemethode, meerdimensionale schaalanalyse en mogelijk de goals achievement matrix methode³¹. Maar ook de concordantieanalyse is, voor het concordantie gedeelte tenminste, kwalitatief. Deze methoden zijn derhalve minder geschikt voor gebruik in de Waddenzee op basis van dit criterium.

§IV.1.9 Het voldoen aan functies.

De methode moet de voor hem beoogde functies kunnen vervullen. Aan het einde van hoofdstuk twee zijn een aantal mogelijke functies van afwegingsmethoden opgesomd. Het was oorspronkelijk de bedoeling om met behulp van interviews vast te stellen welke functies de opdrachtgever vervuld wenst te zien worden. Dit is echter niet gelukt. Het erg moeilijk om functies vast te stellen zonder te weten wat er precies kan en in welke combinaties. De meeste afwegingsmethoden kunnen voor meer dan de helft van de functies ingezet worden. Het is dan ook de vraag of dit een onderscheidend criterium moet worden. Misschien is het beter om een afwegingsmethode te selecteren met kennis over de mogelijke functies. Daarom is in bijlage 3 een tabel opgenomen met daarin de mogelijke functies die elke afwegingsmethode zou kunnen vervullen. Deze lijst is ook nuttig te gebruiken om meerdere methoden te combineren.

§IV.1.10 De toevoeging van een onzekerheidsanalyse.

Om de betrouwbaarheid van de uitkomst van het model in te kunnen schatten is het noodzakelijk om een onzekerheids- en een gevoeligheidsanalyse toe te voegen. Sommige afwegingsmethoden laten dit makkelijker toe dan anderen. Voor kostenbaten analyse, de gewogen somming en de op gewogen somming gebaseerde SAATY-methode, de verwachtingswaarde methode en de meerdimensionale schaalanalyse, is het uiterst eenvoudig om een gevoeligheids- of een onzekerheidsanalyse te maken. Het zijn beide eenvoudige lineaire modellen. Voorwaarde is wel (in het geval van een onzekerheidsanalyse) dat de gebruiker een inschatting maakt van de onzekerheid in de input en de weegfactoren dan wel de monetaire waardering. Ook voor presentatiemethoden is het natuurlijk heel makkelijk, op voorwaarde dat de onzekerheid in de input bekend is. Met de andere methoden ligt dat wat moeilijker. Voor kwalitatieve methoden is het bijvoorbeeld moeilijker om een onzekerheidsanalyse toe te voegen vanwege de discontinuïteit van de data. Alternatief A scoort beter dan alternatief B. Dit blijft zo, ook als B een beetje verbetert. Pas als alternatief B veel verbetert kan B ineens beter worden dan A. Dit discontinue karakter van de kwalitatieve afwegingsmethode maakt het moeilijker om een onzekerheidsanalyse toe te voegen. Het kan

³⁰ Alleen al gezien het feit het derde spoor van het thema gebruik waar ook dit onderzoek in opgenomen is bestaat uit de ontwikkeling van een beslissingsondersteunend model waarmee kwantitatieve voorspellingen gedaan worden. Maar ook de “criteria-groep” heeft bij monde van Daniëlle Hirsch al aangegeven dat de criteria zowel kwalitatief als kwantitatief van aard zullen zijn.

³¹ Dit is helemaal afhankelijk van hoe er mee gerekend wordt.

wel, maar het nut van een dergelijke analyse wordt een stuk kleiner. De bedoelde kwalitatieve methoden zijn de Evamix-, de regime- en de permutatiemethode en de concordantie analyse.

§ IV.2 Bespreking van de kwaliteit van elke methode

In deze paragraaf zal de bespreking van de voorgaande paragraaf nog eens samengevat worden per methode. Dit ter bevordering van de leesbaarheid.

§IV.2.1 Monetaire Evaluatiemethoden

§ IV.2.1.1 Kostenbaten analyse

De kostenbaten analyse komt voort uit algemene economische theorie en wordt mede daarom breed geaccepteerd. De methode is aangepast om ook toepasbaar te zijn voor collectieve goederen. Het is op deze toepasbaarheid waar de meeste kritiek zich op richt. De methode is goed onderbouwd, maar deze onderbouwing suggereert dat de methode gebaseerd is op de welvaartstheorie van Pareto, terwijl deze op het belangrijkste punt verlaten is. Gevolg hiervan is dat de methode de gebruiker dwingt tot twee vooronderstellingen, die in het bedrijfsleven misschien wel acceptabel zijn, maar in sociaal-economische context, zeker als het gaat om milieu en natuur, waarschijnlijk niet. Deze zijn: “Al het geld is evenveel waard” en “geld is het enige criterium.” Ondanks deze tekortkomingen kan de monetaire kostenbaten analyse heel nuttig zijn, mits deze wordt aangevuld met een andere methode zodat wel rekening gehouden kan worden met niet-monetaire criteria en zodat afgeweken kan worden van marktwerking waar dat nodig is.

De structuur van de KBA analyse kenmerkt zich door de expliciete integratie van verschillende criteria, waarbij de markt bepaalt in welke verhouding de criteria gewogen moeten worden. Daarnaast wordt niet rekening gehouden met de verdeling van de lusten en de lasten over de verschillende actoren, er wordt alleen op scores op de criteria afgerekend. Met de keuze voor marktwerking als verdeelsleutel wordt de keuze van criteria beperkt tot monetair waardeerbare criteria. Nieuwe criteria en nieuwe alternatieven kunnen probleemloos ingepast worden. Wel is het zo dat de KBA veel onderzoek vraagt, zeker als marktprijzen benaderd moeten worden voor producten die niet aan marktwerking onderhevig zijn.

Waarschijnlijk wordt de betrouwbaarheid van de KBA overschat door de gebruiker. Dit kan betekenen dat met een te grote zekerheid beslissingen worden genomen. Er treedt geen informatieverlies op, in zoverre, dat er geen informatie wordt omgezet van een hogere naar een lagere orde. Doordat alleen monetair waardeerbare criteria gebruikt kunnen worden zal er mogelijk wel informatieverlies optreden daar waar een aantal criteria genegeerd moeten worden omdat zij niet monetair waardeerbaar zijn. Tot slot is het zeer goed mogelijk om een onzekerheidsanalyse toe te voegen aan een KBA.

De methode is nuttig, goed toepasbaar en breed geaccepteerd. Wel zullen bij gebruik een aantal aanvullende methoden gebruikt moeten worden om de gebreken van KBA op te vangen.

§ IV.2.1.2 Kosteneffectiviteits analyse

De kosteneffectiviteitsanalyse is nauwelijks behandeld in dit verslag. Voor een groot deel komt deze methode overeen met de KBA, vooral waar het gaat om de kritiek. Het grootste nadeel van de KEA is dat in feite alleen vergeleken kan worden tussen alternatieven met dezelfde kosten of met dezelfde criteriascores, of met maar één ander criterium naast de kosten. Wordt niet aan één van deze voorwaarden voldaan, dan kan er in feite al niet meer vergeleken worden zonder dat er een MCA- of KBA-methode wordt toegepast. In de meeste gevallen zal daarom de KEA een KBA zijn, waarbij de baten door de kosten worden gedeeld in plaats dat de kosten van de baten worden afgetrokken. In dat geval gaan alle opmerkingen over KBA ook op voor de KEA.

§IV.2.2 Presentatiemethoden

§ IV.2.2.1 Scorekaartmethode

De scorekaartmethode gaat er duidelijk van uit dat de beslisser zelf in staat is om de verschillende criteria te integreren. Zeker bij een klein aantal alternatieven en criteria zal dit waar zijn.³² Bovendien voelt de beslisser zich niet in een bepaalde richting geduwd. Bij grotere aantallen criteria en alternatieven wordt een scorekaartmethode minder overzichtelijk. De structuur van de presentatiemethode kenmerkt zich doordat er alleen criteria worden weergegeven, geen actoren, en doordat de beslisser zelf de criteria moet integreren.

De scorekaartmethode is uiterst flexibel, alle criteria kunnen worden meegenomen en ze kunnen op elk moment worden ingepast. De enige begrenzing is de ruimte en de overzichtelijkheid. Er mogen niet te veel criteria of te lange beschrijvingen gebruikt worden. De betrouwbaarheid van de methode is volledig afhankelijk van de input. Wel is het zo dat er extra nadruk aan een bepaald criterium gegeven kan worden door dit criterium op te splitsen in meerdere deelcriteria en deze stuk voor stuk te presenteren. Omdat ook meerdere criteria samengevoegd kunnen worden tot één nieuw criterium, waarmee hun aandacht vermindert, mag de vraag gesteld worden wanneer de juiste nadruk op criteria gelegd wordt.

Er treedt geen informatieverlies op, tenzij informatie ingekort moet worden om de overzichtelijkheid te bewaren. Er hoeft bij deze methode geen onzekerheidsanalyse toegevoegd te worden, wel is het aanbevelenswaardig om bij de gepresenteerde gegevens ook informatie over de zekerheid van de uitspraak toe te voegen. Dit kan voor een heel groot deel al gedaan worden door de presentatie (aantal significante cijfers, gebruik van plusjes en

³² (Hobbs et al 1992) noemt hier de grens van 10 criteria bij 10 alternatieven.

minnetjes in plaats van exacte cijfers, woorden als: ‘ongeveer’, ‘waarschijnlijk’ en ‘misschien’ en het gebruik van voetnoten).

Deze methode kan heel nuttig zijn als er een beperkt aantal alternatieven en criteria zijn en de beslisser op zijn intuïtie kan en durft te vertrouwen. Bij complexere beslissingen of bij een behoefte aan een meer expliciete integratie kan beter een KBA of MCA gekozen worden.

§ IV.2.2.2 Community impact analyse

The community impact analyse onderscheidt zich met name van de scorekaartmethode door de andere structuur. De community impact analyse is gericht op de impact van alternatieven op de verschillende actoren. Het verdelingsaspect wordt daarmee nadrukkelijk meegenomen. Tegenover deze extra informatie staat het nadeel dat nu niet direct het totaalplaatje duidelijk is. Community impact analyse zou dan ook goed naast een presentatiemethode gebruikt kunnen worden. Verder gaan alle opmerkingen over de scorekaartmethode ook op voor de community impact analyse.

Deze methode is erg nuttig als er behoefte is aan extra informatie over de al dan niet rechtvaardige verdeling van baten en lasten over verschillende actoren.

§ IV.2.2.3 AMOEBE

Het belangrijkste onderscheid tussen de AMOEBE en de scorekaartmethode is de standaardisatie die bij een amoebe is toegevoegd. Een ander opvallend aspect is de grafische presentatie. Door de standaardisatie wordt er niet alleen gepresenteerd, impliciet wordt er ook een te halen doelstelling vastgesteld. Het gaat dan om een kwantitatieve doelstelling. Niet: Er moeten meer zeehonden komen, maar er horen 4000 zeehonden te zijn. Omdat dit soort kwantitatieve doelstellingen gesteld moeten worden, is de AMOEBE alleen geschikt voor kwantitatieve criteria. Door de grafische presentatie wordt gesuggereerd dat alle criteria even belangrijk zijn. Alle criteria zijn tenslotte even breed. Er wordt in de AMOEBE al min of meer rekening gehouden met onzekerheid doordat er een bandbreedte gesteld kan worden voor de doelstelling. Doel is dan niet een 100% bereiken van de doelstelling, maar bijvoorbeeld tussen de 75% en de 200%. Voor het overige is de AMOEBE vergelijkbaar met de scorekaartmethode.

Omdat de methode op de meeste punten minder scoort dan de scorekaartmethode, zou deze laatste mijns inziens beter gebruikt kunnen worden voor afweging in de Waddenzee. Wel nuttig en toepasbaar in andere methoden is de aan een doelstelling gerelateerde standaardisatie.

§IV.2.3 Multicriteria analyse

§ IV.2.3.1 Gewogen Sommering

De gewogen sommering lijkt, zoals eerder opgemerkt, veel op de KBA. Het voornaamste verschil is de vaststelling van weegfactoren. Deze worden vastgesteld door de beslisser, eventueel op basis van enquêtes onder de bevolking of de actoren. Deze toekenning van gewichten zonder een gedegen wetenschappelijk onderzoek, zoals bij KBA, maken de onderhandelingspositie van de beslisser zwakker. Nu zijn expliciet de subjectieve voorkeuren van de beslisser opgenomen, in plaats van de veel objectievere markt. Een handig hulpmiddel voor de gewogen sommering zou de techniek zijn die de SAATY-methode gebruikt om het vaststellen van gewichten te vereenvoudigen. De gewogen sommering is lineair van aard. Daarom zal ook een domein vastgesteld moeten worden waarop de gebruikte gewichten geldig zijn. Alleen alternatieven die binnen dit domein blijven kunnen beoordeeld worden, alternatieven die er buiten vallen moeten een speciale behandeling krijgen om ze vergelijkbaar te houden. Wat er precies zal moeten gebeuren om deze alternatieven vergelijkbaar te maken is een punt voor verder onderzoek.

De structuur van gewogen sommering kenmerkt zich, net als alle andere MCA met uitzondering van de goals achievement methode, door expliciete integratie van criteria, het zij door de beslisser, het zij door burgers, en een afweging op basis van criteria, zonder dat de verdeling over actoren van baten en lasten aan de orde komt. Deze laatste kan overigens wel opgenomen worden, maar dan als extra criterium. In principe kunnen alle soorten criteria meegenomen worden, maar dan moeten de kwalitatieve scores wel omgezet worden in kwantitatieve scores. Als er gewerkt wordt met een onzekerheidsanalyse hoeft dit nauwelijks een probleem te zijn omdat dan alleen voorkeursvolgordes omgezet hoeven te worden in bijvoorbeeld beoordelingscijfers. De gewogen sommering is erg flexibel, nieuwe criteria en alternatieven kunnen probleemloos worden toegevoegd. Over de betrouwbaarheid geen opmerkingen. De betrouwbaarheid zou wel eens kunnen toenemen door gebruik van de SAATY-methode. Er treedt geen informatieverlies op en er kan probleemloos een onzekerheidsanalyse worden toegevoegd.

De methode kan nuttig zijn en is erg helder voor de gebruiker. Dit is zeker een goede mogelijkheid voor toepassing in de Waddenzee.

§ IV.2.3.2 *Verwachtingswaarde methode*

De verwachtingswaarde methode is een aanvulling op de gewogen sommering waarbij kwalitatieve gewichten of criteria omgerekend kunnen worden naar kwantitatieve gewichten en criteria zonder dat daarbij extra informatie ingewonnen hoeft te worden. De methode baseert deze mogelijkheid op de vooronderstelling dat setjes scores en setjes weegfactoren altijd een bepaald patroon volgen. Deze vooronderstelling is mijns inziens onhoudbaar. Bovendien is het extra onderzoek dat nodig is om kwalitatieve criteria om te zetten in kwantitatieve criteria in de vorm van bijvoorbeeld waarderingscijfers zo minimaal dat er geen grote noodzaak is om deze methode te gebruiken. Kortom, deze methode kan beter niet gebruikt worden.

§ IV.2.3.3 Concordantie analyse

De concordantiemethode is zeker aanvechtbaar. De methode bestaat in feite uit twee onderdelen, waarbij bij het eerste onderdeel zuiver kwalitatief is, óók bij een kwantitatieve invoer. Het tweede gedeelte is wel kwantitatief, maar houdt weer geen rekening met gewichten. De theorie hierachter is dat als je met deze twee niet-complete systemen tot hetzelfde antwoord komt, dat dit antwoord dan waarschijnlijk wel goed is. Er is echter geen enkele reden om te veronderstellen dat als twee incomplete systemen, die niet complementair zijn, tot hetzelfde antwoord komen, dat dit antwoord wel goed zal zijn. De twee systemen vormen immers ook samen niet een compleet systeem. Daarnaast wordt ook nog eens gebruik gemaakt van een vorm van paarsgewijze vergelijking die uitermate gevoelig is voor manipulatie: de toevoeging van een extra alternatief kan gevolgen hebben voor de prioriteitsvolgorde in de andere alternatieven! Qua structuur komt de Concordantieanalyse overeen met de gewogen somming. De methode kent dezelfde problemen (en oplossingen) als de gewogen somming ten aanzien van de criteria die gebruikt kunnen worden. Zoals gezegd is de methode niet flexibel ten aanzien van de toevoeging van nieuwe alternatieven, deze kunnen de voorgaande analyse wijzigen. Bij beide gedeelten van de concordantiemethode treedt informatieverlies op. In het eerste worden kwantitatieve scores kwalitatief gemaakt, in het tweede wordt de gewichtenset niet gebruikt. Tot slot is de toevoeging van een onzekerheidsanalyse minder zinvol bij het kwalitatieve gedeelte, vanwege het discontinue karakter van een kwalitatieve methode. Al met al lijkt deze methode mij nuttig nog betrouwbaar. Niet gebruiken.

§ IV.2.3.4 Evamix methode

De Evamix methode houdt expliciet rekening met het verschil tussen kwalitatieve en kwantitatieve criteria. Het is jammer dat daar vervolgens twee keer een standaardisatie bij nodig is. Bij de tweede standaardisatie en het daaropvolgende vaststellen van gewichten, is het nauwelijks mogelijk om nog zicht te houden op de betekenis van die gewichten. Het is niet mogelijk om onder deze standaardisatie uit te komen zonder daarbij een bijna nieuwe methode te ontwerpen. De structuur van de Evamix methode is weer gelijk aan die van de gewogen somming. Door dat expliciet rekening gehouden wordt met kwantitatieve én kwalitatieve criteria, kunnen ook alle grootheden met elkaar vergeleken worden. De methode is minder flexibel ten aanzien van de input. Ook bij de evamix methode wordt gebruikt gemaakt van paarsgewijze vergelijking, zodat ook deze gevoelig is voor manipulatie. Ook deze methode is onbetrouwbaar en veel te gevoelig voor manipulatie. Niet toepassen dus.

§ IV.2.3.5 Regimemethode

De structuur van de regimemethode is weer gelijk aan die van de gewogen somming. Net als de voorgaande twee methoden maakt ook de regimemethode gebruik van paarsgewijze vergelijking is alleen daarom al onbetrouwbaar en gevoelig voor manipulatie. Daarnaast is de methode kwalitatief van aard, zodat er informatieverlies optreedt als er wel kwantitatieve informatie beschikbaar is. Bovendien suggereert de methode een veel grotere precisie dan in werkelijkheid gegarandeerd kan worden. Tot slot is het minder zinvol om hier een onzekerheidsanalyse toe te voegen, vanwege de kwalitatieve aard van de methode. Dit alles lijkt mij reden genoeg om de methode niet te gebruiken.

§ IV.2.3.6 Permutatiemethode

Ook de permutatiemethode heeft een structuur die gelijk is aan die van gewogen somming. De methode is kwalitatief, waardoor er informatie verloren gaat. Daarnaast is de methode weinig doorzichtig. Er gebeurt van alles, het komt wel overeen met de intuïtie, maar waarom het precies zo zou moeten wordt mij niet duidelijk. Ik heb geen fouten kunnen aantonen in de berekening, zoals bij voorgaande methoden, maar helemaal lekker zit het toch niet. Dit zal waarschijnlijk voornamelijk vastzitten op het omzetten naar kwalitatieve scores. In principe kunnen alle soorten criteria meegenomen worden. De berekeningen nemen enorm toe met het aantal alternatieven. (Aantal mogelijke rangordeningen: $n = A!$, waarbij $A =$ het aantal Alternatieven). Dit betekent dat al heel snel computers nodig zullen zijn en dat bij grote aantallen alternatieven de methode te bewerkelijk wordt. Het toevoegen van een onzekerheidsanalyse is hier moeilijker omdat de methode kwalitatief van aard is én vanwege het gebrekkige doorzicht in de methode. Hoewel deze methode minder slecht is dan de drie voorgaande, zou ik hem, vanwege het informatieverlies, toch niet gebruiken in de Waddenzee.

§ IV.2.3.7 Meerdimensionale schaalanalyse

Deze methode is een uitbreiding op andere methoden, waarmee weegfactoren berekend kunnen worden. Zoals aangetoond in paragraaf §IV.1.1 is deze methode volstrekt onbetrouwbaar en wordt de interpretatieonzekerheid alleen maar vergroot met deze methode. Verder is er geen enkel voordeel aan het gebruik van deze methode verbonden, zodat het gebruik van deze methode in alle gevallen afgeraden moet worden.

§ IV.2.3.8 Goals achievement matrix methode

Deze methode heeft een duidelijk andere structuur dan alle voorgaande multicriteria analyses. In plaats van een integratie van criteria worden hier de gevolgen voor verschillende actoren geïntegreerd door de beslisser. In (Ministerie van Financiën 1992) blijft onduidelijk hoe dit precies zou moeten gebeuren. Er zijn meerdere mogelijkheden. De mogelijkheid die hier wordt uitgewerkt is kwalitatief en er gaat informatie verloren. Dat is jammer en kan

waarschijnlijk ook wel anders. Omdat er in deze bronnen te weinig informatie beschikbaar is over de methode en hij te veel afwijkt van andere methoden is hier geen oordeel over de methode opgenomen. Het idee om expliciet rekening te houden met actoren is wel interessant, al heeft dit als nadeel dat straks aan actoren gewichten zouden moeten worden toegekend, wat politiek misschien moeilijk te verkopen is. (Wij vinden de belangen van de horeca vier keer zo belangrijk als die van de visserij...) Verder onderzoek in dit type methoden kan zeker de moeite waard zijn.

§ IV.2.3.9 SAATY-methode

De SAATY-methode is een gebaseerd op gewogen sommering. Hier worden echter een aantal elementen aan toegevoegd. Ten eerste wordt het vaststellen van weegfactoren vereenvoudigd door hier gebruik te maken van paarsgewijze vergelijking van criteria. Op basis van deze vergelijking kunnen vervolgens weegfactoren vastgesteld worden. Bovendien kan de consistentie van de paarsgewijze vergelijking bepaald worden, waarmee ook een inschatting gemaakt kan worden van de zekerheid en de helderheid waarmee de beslisser de weegfactoren vaststelt. Daarnaast moet er een criteriastructuur gemaakt worden. De integratie van verschillende niveaus van criteria wordt daarmee expliciet gemaakt én gebruikt om de berekening van weegfactoren te vereenvoudigen. Tot slot kunnen de scores op criteria ook geschat worden door de gebruiker. Deze onderdelen zijn echter onafhankelijk van elkaar te gebruiken. Het schatten van scores door de beslisser is niet zo zinvol als ook gebruik gemaakt kan worden van modellen en experts. De andere twee aspecten, expliciete integratie en vaststellen van weegfactoren zijn mogelijk nuttige verbeteringen van de gewogen sommering.

De methode werkt en is goed en correct onderbouwd.³³ De methode is minder flexibel, vanwege de paarsgewijze vergelijking en de opgebouwde criteriastructuur. De betrouwbaarheid van het uiteindelijke model wordt verhoogd doordat de gewichtonzekerheid wordt verkleind. De SAATY-methode staat de toevoeging van een onzekerheidsanalyse niet in de weg. Sterker nog, de berekende consistentie kan een hulpmiddel zijn bij het maken van een inschatting van de onzekerheid in de gewichten. De SAATY-methode zal in complexere gevallen een goede aanvulling zijn op gewogen sommering.

§ IV.3 Samenvatting en conclusies hoofdstuk IV

In dit hoofdstuk zijn de afwegingsmethoden getoetst aan de in hoofdstuk III opgestelde kwaliteitscriteria. Een heel aantal afwegingsmethoden heeft deze toets niet doorstaan. De resultaten van deze toets worden weergegeven in een tabel. Om de informatie zo eenvoudig mogelijk samen te vatten zal ik gebruik maken van een presentatiemethode. De criteria komen in een verticale kolom te staan, de alternatieve methoden op een horizontale rij. Hierdoor

³³ Wat ook te verwachten was, Saaty is wiskundige...

ontstaat een matrix. In deze matrix kan alle informatie over de alternatieven en de criteria samengevat worden. Het criterium functionaliteit is niet opgenomen in de kolom omdat er over de functionaliteit nog geen uitspraken gedaan kunnen worden. Er is gebruik gemaakt van een ordinale set van 5 symbolen: --, -, 0, +, ++. In deze symbolen is niet alleen de score maar ook het belang van de score verwerkt. Een "--" geeft niet alleen aan dat een alternatief hier relatief (t.o.v. de andere alternatieven) slecht scoort, maar ook dat het erg vervelend is dat het alternatief hier slecht scoort. Een 0 kan betekenen dat het criterium er voor dit alternatief niet toe doet, dat het alternatief heel gemiddeld scoort of dat het alternatief wel goed of slecht scoort, maar dat dit niet zo belangrijk is vanwege mogelijke compenserende maatregelen. Tot slot is er een kolom toegevoegd met een eindoordeel voor de methode. Dit is een oordeel over de toepassingsmogelijkheden van de methode voor gebruik door Rijkswaterstaat in de Waddenzee. De betekenis van de woorden is als volgt:

- Aanvullen: Als onderdeel wel bruikbaar, mits aangevuld met andere methoden.
- Beperkt: Alleen onder bepaalde omstandigheden bruikbaar.
- Niet: Niet gebruiken.
- Mogelijk: Nader onderzoek naar deze methode is noodzakelijk om de mogelijkheden vast te stellen.

Methode	argumentatie	De theorie in de praktijk	Scripts	Alle criteria kunnen opgenomen worden	Flexibiliteit ³⁴	Betrouwbaarheid	Informatieverlies	Onzekerheidsanalyse	Eindresultaat
KBA	+	-	0 ³⁵	-	0	-	+	+	Aanvullen
KEA	+	-	0	-	0	-	+	+	Aanvullen
Scorekaart methode	+	+	-	++	+	0	+	+	Beperkt
Community impact analyse	+	+	-	++	+	0	+	+	Beperkt
AMOEBE	+	+	-	0	0	-	+		Niet
Gewogen Sommering	+	+	+	+	0	-	+	+	Aanvullen
Verwachtingswaarde methode	--	+	-	+	0(-)	-	--	+	Niet
Concordantie analyse	--	-	-	+	0(-)	--	-	0	Niet
Evamix methode	+	-- ³⁶	-	++ ³⁷	+(-)	--	+	0	Niet
Regime methode	--	+	-	+	+(-)	--	--	0	Niet
Permutatie methode	+	+	+	+	+(-)	--	--	0	Niet
Meerdimensionale schaalanalyse	--	+	--	+	+	-	--	+	Niet
Goals achievement matrix methode	+	+	+	+	0(--)	--	0	+	Mogelijk
SAATY-methode	+	+	+	+	-(--)	0	+	+	Aanvullen

Tabel 9: de resultaten van de toetsing van afwegingsmethoden.

³⁴ Tussen haakjes staat de score die de methode zou krijgen bij een niet-geautomatiseerd systeem.

³⁵ Door de scripts zijn de KBA en de KEA slechts geschikt voor een gedeeltelijke toepassing. Zij vragen om aanvulling met een andere methode.

³⁶ Deze dubbele min vanwege de verplichte dubbele standaardisatie.

³⁷ Een dubbele plus omdat de evamix methode expliciet rekening houdt met kwantitatieve en kwalitatieve criteria.

Hoofdstuk V Concept ontwerp van een afwegingsmethode

§ V.1 Inleiding

In dit laatste hoofdstuk wil ik aan de hand van een casus al de bovenstaande informatie integreren tot een concept ontwerp van een afwegingsmethode voor gebruik in de Waddenzee. Dit is niet het beste ontwerp voor alle beleidsproblemen, maar een illustratie van de mogelijkheden die voorgaande analyse biedt bij de keuze van een afwegingsmethode. Ik wil daarbij onderdelen van verschillende methoden in elkaar passen, zodat de sterke punten van deze methoden gecombineerd worden en de zwakke punten opgevangen door het gebruik van andere methoden. De basis voor dit ontwerp is de gewogen somming, vanwege de doorzichtigheid van de methode en de flexibiliteit van het ontwerp, waardoor allerlei andere methoden hier ingepast kunnen worden. In paragraaf 2 zal het ontwerp geschetst worden, inclusief alle voorbereidingen en maatregelen die nodig zijn om de methode te laten werken. In paragraaf 3 bespreek ik de methode aan de hand van de in de voorgaande hoofdstukken besproken modelkwaliteitscriteria.

§ V.2 Het ontwerp

Dit ontwerp wordt gepresenteerd in een stappenplan waarin alle onderdelen van de methode inclusief ondersteunende werkzaamheden naar voren komen. Niet alle onderdelen zijn echter even goed uitgewerkt. Voor de methode zelf zijn die grotendeels terug te vinden in voorgaande hoofdstukken. Omdat de ondersteunende werkzaamheden buiten dit onderzoek vallen wordt er hier ook niet veel over gezegd.

§V.2.1 Voorbereidend werk

Als eerste zal er natuurlijk een probleem geformuleerd moeten worden, gevolgd door een doelstelling. Alle alternatieven zijn er op gericht om de doelstellingen te realiseren en de problemen op te lossen. Zeker bij grote breed ervaren problemen kan het verstandig zijn om hier al de actoren te betrekken bij probleemsigalering en bij het vaststellen van te bereiken doelen.

Daarnaast moeten alle actoren die belang hebben bij de doelrealisering en alle actoren die beïnvloedt gaan worden door de alternatieven geïdentificeerd worden. Deze zijn nodig om straks alle criteria te vinden. Tegelijkertijd kunnen de actoren niet geïdentificeerd worden zonder dat er zicht is op de alternatieven. Omdat je alternatieven niet goed kunt ontwerpen

zonder te weten aan welke criteria ze moeten voldoen, is er hier een iteratief proces nodig.

Door alternatieven te genereren worden mogelijk nieuwe actoren betrokken die mogelijk nieuwe eisen, problemen of doelen mee zullen brengen, waardoor de eerste stappen van de besluitvormingsprocedure nogmaals doorlopen moeten worden.

§V.2.2 Criteria

De tweede stap in dit schema is het vaststellen van criteria. Alle mogelijke criteria mogen hier genoemd worden, later worden ze wel geordend. Met behulp van de criteria kan worden vastgesteld of de doelen in voldoende mate bereikt worden, of het probleem voldoende wordt opgelost, of er geen onacceptabele schadelijke neveneffecten zijn en ze zijn een uitdrukkingsvorm van alle mogelijke argumenten die actoren voor of tegen alternatieven in kunnen brengen. Daarom zullen ook de actoren onderzocht moeten worden op de criteria die zij belangrijk vinden.

Vervolgens kunnen deze criteria geordend worden in een criteriaboom, zoals die bij de SAATY-methode gebruikt worden. De onderste criteria zijn dan het meest operationeel, deze zijn meetbaar. Verder naar boven worden steeds groepen criteria geaggregeerd tot een criterium van een hoger abstractieniveau. Helemaal bovenaan staan de te bereiken doelstellingen. Door deze boom van boven naar beneden en van beneden naar boven te doorlopen kunnen extra criteria ontdekt worden die vergeten zijn maar zeker opgenomen moeten worden. Het is voor de kwaliteit van het model van het grootste belang dat de set criteria compleet is. Bovendien is het in verband met de geringere flexibiliteit van de SAATY-methode van belang om na het rekenen niet meer te veel aan de boom zelf te hoeven sleutelen.

Als deze criteria zijn vastgesteld moet vervolgens worden onderzocht om wat voor soort criteria het gaat. Sommige criteria zijn in feite harde randvoorwaarden. Alternatieven moeten aan deze criteria voldoen, ze zijn niet afweegbaar tegen andere criteria. Deze criteria moeten niet in de criteriaboom worden opgenomen, maar elk alternatief zal aan al deze criteria moeten voldoen. Voldoet een alternatief niet dan komt het per definitie niet in aanmerking voor uitvoering en het hoeft dus niet verder meegenomen te worden in de afweging. Voorbeelden van harde randvoorwaarden zijn: wettelijke regelingen, veiligheidsrisico's voor omwonenden of eerder genomen besluiten waarop nu niet teruggekomen moet worden.

Daarnaast kunnen interne, publieke en marktafhankelijke criteria onderscheiden worden. Met marktafhankelijke criteria bedoel ik die criteria waarvan de waarde door de markt bepaald kan worden. De kosten van dijkverhoging kan vergeleken worden met de kosten van de aanleg van deltawerken. Die kosten worden in feite door de markt van vraag en aanbod (arbeidskosten, materiaalkosten e.d.) bepaald. Veel andere criteria kunnen omgerekend worden naar monetaire criteria (Zie Rudolf S. de Groot).

Van een aantal criteria zal echter geen prijs berekend kunnen of mogen worden. Deze noem ik de publieke criteria. Het gaat mij dan om criteria die de belangen van "het volk" behartigen of een collectief goed waar iedereen evenveel recht op hoort te hebben, onafhankelijk van inkomen of bijvoorbeeld iets als de intrinsieke waarde van natuur.

Tot slot zijn er nog de interne criteria. Criteria die vanuit de organiserende organisatie gesteld worden. Te denken valt dan aan budget, tijd, organisatorische capaciteit en bestuurlijke competentie. Deze indeling is noodzakelijk om straks duidelijk vast te stellen op welke manier de criteria behandeld moeten worden.

Tot slot moet er per groep criteria vastgesteld worden in hoeverre zij een representatie vormen van het criterium op een hoger abstractieniveau. Als het criterium “milieu” op een lager abstractieniveau wordt vertegenwoordigd door grond, water, lucht, flora en fauna, dan kan de vraag gesteld worden of dit de complete representatie van het criterium milieu is. Als de beslisser ook geluidsoverlast tot het milieu rekent, en dit niet in “lucht” is inbegrepen, dan is voor hem milieu niet compleet gerepresenteerd. Door op dit moment een schatting te maken van de representatie in procenten per groep kan later een inschatting gemaakt worden van de mate van representatie van het totale model. Als er ook nog genoteerd wordt welke criteria voor het meest gemist worden, dan is er ook een aangrijpingspunt voor verbetering, als straks blijkt dat de representatie te gering is.

§V.2.3 Alternatieven

Als alle criteria duidelijk zijn en gecategoriseerd kunnen er alternatieve oplossingen worden gegenereerd. Dit is een creatief proces dat volledig buiten dit onderzoek valt. In ieder geval is het belangrijk om de ruimte open te laten om de alternatieven aan te passen en te verbeteren, ook nadat volgende fases van de besluitvorming al doorlopen zijn.

§V.2.4 Impact analyse en voorbereidend werk

Om de alternatieven te kunnen beoordelen zullen eerst de consequenties van het uitvoeren van het alternatief bekend moeten zijn. Deze moeten voorspeld worden. Omdat een voorspelling nooit helemaal zeker is, moet deze voorspelling bij voorkeur gegeven worden in een geschat betrouwbaarheidsinterval.³⁸ Met behulp van dit betrouwbaarheidsinterval kan er straks een indicatie gegeven worden van de onzekerheid in het eindresultaat van de afwegingsmethode.

Als er inzicht is in de orde van grootte van de scores dan kan er ook een domein vastgesteld worden waarop de weegfactoren geldig moeten zijn. Dit domein is precies zo groot dat alle scores er binnen vallen.

Daarnaast kunnen nu de criteria waar nodig gestandaardiseerd worden. Met name operationele criteria zullen ver van de belevingswereld van de beslisser af staan. Een standaardisering naar doelbereiking, zoals die bijvoorbeeld in de AMOEBE en in de “goals achievement methode” voorkomt, kan een enorme verbetering in de begrijpelijkheid en in de hanteerbaarheid van deze criteria betekenen. Daarnaast kunnen de eenheden van waarin de

³⁸ Ook betrouwbaarheidsintervallen zijn niet zeker, omdat hier een schatting wordt gemaakt van de kansverdeling, het gemiddelde en de verwachte afwijking daarvan.

scores uitgedrukt worden zo gekozen worden (of gestandaardiseerd) dat ze makkelijk hanteerbaar zijn voor de beslisser. Het gaat er daarbij om dat de criteria vooral in ongeveer dezelfde orde van grootte vallen, zodat de beslisser straks niet het overleven van een lieveheersbeestje hoeft te vergelijken met een groei van de werkgelegenheid met 50.000 banen. Daarnaast kunnen op dit moment de eerder genoemde standaardisaties gemaakt worden.

Tot slot is dit het moment om een community impact analyse te maken. Door te onderzoeken hoe de kosten en baten van de alternatieven verdeeld worden over de verschillende actorgroepen kunnen de alternatieven gecontroleerd worden op rechtvaardigheid van deze verdeling. Bovendien kan nu een inschatting gemaakt worden van de politieke haalbaarheid van alternatieven, er kan onderzocht worden waar de meeste weerstand ondervonden zal worden en er kunnen compenserende maatregelen bedacht worden om zo actoren over de streep te trekken.

§V.2.5 Invullen afwegingsmodel

De volgende stap is het invullen van het afwegingsmodel. Daarvoor moeten eerst weegfactoren worden vastgesteld. Voor elk type criterium gebeurt dit op een andere manier. (Zie §V.2.2 voor deze criteriumtypen.)

De zogenaamde marktconforme criteria kunnen behandeld worden als met een KBA. De kosten en baten worden berekend van alle scores op deze criteria, en met behulp van deze kosten en baten kunnen ook de onderlinge verhoudingen berekend worden. Daarmee ontstaat in feite al een stelsel van weegfactoren, dat zich in feite onttrekt aan de strakke hiërarchie van de SAATY-criteriaboom. Dit is erg handig, omdat hierdoor extra informatie over de verhoudingen tussen criteria uit verschillende criteriagroepen ontstaat, waarmee later weer de consistentie van de hele criteriaboom gecheckt kan worden. Ik kom hier later op terug. Na de marktconforme criteria kunnen verhoudingen tussen de publieke criteria worden vastgesteld. Deze worden bij voorkeur niet alleen aan elkaar, maar óók aan marktconforme criteria binnen dezelfde criteriagroep gerelateerd. Het publiek of de beslisser bepaald zo dus ook hoeveel de publieke criteria waard zijn in verhouding tot marktconforme criteria.

De verhoudingen tussen interne criteria kunnen op dezelfde manier vastgesteld worden als de publieke criteria. Maar het is ook mogelijk om deze alleen tot elkaar te relateren om ze dan straks niet mee te wegen, maar in plaats daarvan de alternatieven te beoordelen op hun efficiëntie: de verhouding tussen de maatschappelijke opbrengst en de interne kosten.

Tot slot kunnen, als alle verhoudingen tussen de verschillende criteria zijn vastgesteld met behulp van SAATY weegfactoren berekend worden. Daarnaast kan er een berekening van de consistentie van de ingevulde verhoudingen gemaakt worden. Tot slot moet er voor de weegfactoren een kansinterval opgesteld worden. Dit kan door direct bij de vaststelling van de verhoudingen een interval te gebruiken en deze consequent door te rekenen. Ook is het mogelijk om een verhouding van 1:3 te beschouwen als een verhouding tussen 1:2 en 1:4, waarmee ook recht gedaan wordt aan het discontinue karakter van de verhoudingen zoals die

met SAATY gekozen kunnen worden.³⁹ Uiteindelijk resultaat van deze bewerking is een verzameling weegfactoren (met elk een eigen betrouwbaarheidsinterval) voor de criteria. Deze set weegfactoren kan nog eens op betrouwbaarheid gecontroleerd worden met behulp van de verhoudingen die al bekend zijn voor de marktconforme criteria. Als deze verhoudingen na berekening met de SAATY-methode overeenkomen met de verhoudingen die direct berekend zijn met behulp van de KBA, dan zijn de verhoudingen redelijk consistent ingevuld. Als de weegfactoren redelijk consistent blijken te zijn kan er met een gewogen sommering een prioriteitsvolgorde berekend worden.

§V.2.6 Gevoeligheids- en onzekerheidsanalyse

Als alle voorgaande stappen zijn ingevuld rest alleen nog de gevoeligheids- en onzekerheidsanalyse en de interpretatie van de resultaten. De gevoeligheidsanalyse is vooral bruikbaar voor de verbetering van de resultaten, de onzekerheidsanalyse geeft een indicatie van de betrouwbaarheid van de resultaten en van het risico dat er een verkeerde beslissing genomen wordt.

De eenvoudigste manier om een gevoeligheids- en een onzekerheidsanalyse uit te voeren is met behulp van een Monte-Carlo simulatie. Er van uitgaande dat alle scores en alle weegfactoren uitgedrukt zijn in kansverdelingen kan er door telkens voor alle scores en weegfactoren random een getal uit de kansverdeling te trekken en daarmee de gewogen sommering uit te voeren, een schatting gemaakt worden van de gevoeligheid voor de input en de betrouwbaarheid van de resultaten. Als alternatief II negentig van de honderd keer als beste naar voren komt en in de andere tien gevallen een heel goede tweede is, dan zal alternatief II een goede keuze zijn. Als er grote onzekerheid is over de keuze dan kan met behulp van de gevoeligheidsanalyse gezocht kunnen worden naar de grootste bronnen van onzekerheid, zodat er deze onzekerheid efficiënt gereduceerd kan worden door verder onderzoek op de meest gevoelige punten.

Een andere betrouwbaarheidscheck is de compleetheid van de criteria. Als voor alle criteriagroepen de mate van representatie is vastgesteld dan kan nu eenvoudig de totale representatie vastgesteld worden. Dit is een indicator voor de compleetheid van het model. Als het model compleet en betrouwbaar genoeg is of een grotere compleetheid of betrouwbaarheid niet haalbaar is, dan kunnen hier de resultaten geïnterpreteerd worden en gebruikt als argumentatie voor een keuze.

³⁹ Bij SAATY komen allen de verhoudingen 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9 en andersom voor. Tussenvallende waarden kunnen niet worden ingevuld.

§V.2.7 Tabel van herkomst

In de onderstaande tabel is de herkomst van bovengeschetst stappenplan weergegeven. Per onderdeel uit §II.1.1 worden voor dat onderdeel gebruikte methoden of aanbevelingen genoemd. Voor een overzicht van mogelijke onderdelen van methoden zie bijlage 7.

Stappen (Zie begin hoofdstuk II)	bronnen
Extra stap: probleemstelling, doelstelling	Beleidsprocessen I en II (bestuurskunde)
Criteria	Aanbevelingen hoofdstuk IV
Alternatieven	geen
Scores van alternatieven op criteria	Aanbevelingen hoofdstuk IV
Omrekening van scores naar meer bruikbare scores	Gewogen sommering, KBA
Standaardisatie van de scores	Evt. Met standaardisatie naar doelwaarden
Indeling van scores per actor	Community impact analyse
Extra stap: categorisering van de criteria	Nieuwe oplossing
Vaststellen van weegfactoren	KBA, SAATY
Berekening van weegfactoren	SAATY
Berekening van een voorkeursvolgorde in de alternatieven	Gewogen sommering
Extra stap: onzekerheids- en gevoeligheidsanalyse	Aanbevelingen hoofdstuk IV
Presentatie van de scores	Gewogen sommering, community impact
Presentatie van de berekening	Gewogen sommering
Presentatie van de voorkeursvolgorde	Gewogen sommering

Tabel 10 De gebruikte methoden per onderdeel, analoog aan de onderdelen uit §II.1.1.

§ V.3 Bespreking van de methode

In deze paragraaf zal de in de vorige paragraaf geschetste afwegingsmethode getoetst worden aan de in hoofdstuk IV opgestelde criteria. Niet alle criteria zullen uitputtend behandeld worden, omdat hier gekozen is uit bestaande methoden. De methode bestaat uit de beste delen van de kostenbaten analyse, de gewogen sommering, de SAATY-methode, daarnaast is de community impact analyse toegevoegd om een completer beeld van de alternatieven te geven. Bovendien is rekening gehouden met de aanbevelingen uit hoofdstuk III.

De argumentatie voor de methoden is houdbaar voor deze toepassing, zoals te zien is in hoofdstuk IV. De gemaakte vooronderstellingen zijn of acceptabel, of zijn ondervangen door gebruik te maken van andere methoden. Extra vooronderstellingen die gemaakt wordt is dat al deze methoden naadloos op elkaar aansluiten, wat mijns inziens een acceptabele

vooronderstelling is, en dat de beslisser uiteindelijk bepaalt hoe belangrijk de marktconforme criteria zijn. De beslisser bepaalt immers de verhouding tussen marktconforme en publieke criteria. Ook deze vooronderstelling lijkt mij acceptabel, alleen al omdat het noodzakelijk is om de beslisser deze verhouding te laten bepalen in plaats van de markt. Juist omdat er criteria zijn waarvan de waarde niet door een markt bepaald kan of mag worden wordt er gekozen voor een methode waarbij er naast een KBA gebruik wordt gemaakt van een andere methode. Het is niet meer dan logisch dat er vervolgens ook op een andere manier (anders dan door middel van een KBA) bepaald wordt hoe de verhouding is tussen deze criteria en de marktconforme criteria.

De methode voldoet aan meerdere structuren. Door de community impact analyse wordt er rekening gehouden met actoren en met de verdeling van baten en lasten. Deze worden impliciet meegenomen in de afweging. De criteria worden expliciet afgewogen, waarbij zowel de markt als de beslisser en mogelijk het publiek deze afweging bepalen.

Alle soorten criteria kunnen vergeleken worden, maar de criteria moeten wel tot kwantitatief van aard zijn. Echter, zoals eerder opgemerkt, kunnen kwalitatieve criteria vrij eenvoudig gekwantificeerd worden door de juiste vragen te stellen.

De methode is minder flexibel, door het gebruik van de SAATY-methode. Deze vraagt al meer aandacht dan de meeste andere methoden, en het proces van het vaststellen van weegfactoren moet grotendeels overgedaan worden als de criteriaboom moet worden aangepast. Een toevoeging van een alternatief is geen probleem, tenzij de scores voor dit alternatief buiten het vastgestelde domein voor de weegfactoren valt, in dat geval moeten mogelijk de weegfactoren aangepast worden.

De betrouwbaarheid van het model is zo groot mogelijk. De SAATY-methode verhoogt de betrouwbaarheid van weegfactoren, door de identificatie van verschillende soorten criteria en de verschillende aanpak van deze criteria wordt de betrouwbaarheid van de weegfactoren nog eens verbeterd. Daarnaast maakt de presentatie met een betrouwbaarheidsinterval en de onzekerheidsanalyse de kans op interpretatiefouten kleiner, waarmee de interpretatieonzekerheid afneemt.

Er vindt geen informatieverlies plaats. En er is een onzekerheids- en een betrouwbaarheidsanalyse toegevoegd. Deze methode biedt daarmee een aantal verbeteringen ten opzichte van de eerder besproken methoden. Het nadeel van deze methode is tijd die het gebruik kost en de kleinere flexibiliteit. De methode zou daarom alleen gebruikt moeten worden bij grote complexe beslissingen.

Lijst van gebruikte literatuur

De Bruin, J., de Graaf, P.J.F., 1991. Wadden actie plan, Ecologische toestandsbeschrijving - de Waddenamoebe. Rijkswaterstaat, dienst getijdewateren, rijkswaterstaat nota GWWS-91.060

De Groot, R.S., 1994. Evaluation of environmental functions as a tool in planning, management and decision-making. Thesis Wageningen

Kolkman 1996, Modelleren van Integrale Civieltechnische Systemen 2, reader. Dictaat 221320, Universiteit Twente.

Ministerie van Financiën, 1992. Evaluatiemethoden een introductie, 4e herziene druk. Afdeling Beleidsevaluatie en –instrumentatie van het Ministerie van Financiën, Sdu Uitgeverij, Platijnstraat Den Haag.

Nieuwkamer, R.L.J., 1992. Decision support for river management. Proefschrift Technische Universiteit Twente. ISBN 90-9008008-2.

Pearce, D.W., Turner, R. K., 1990. Economics of natural resources and the environment. The John Hopkins university press, Baltimore. ISBN 0-8018-3978-4

Pouwels, I., 1996, De Bruikbaarheid van Prioriteitsstellingmethoden voor Integraal Waterbeheer. Deelrapport 1: Knelpunten in de Bruikbaarheid van Prioriteitsstellingmethoden.

Rip, A., 1996. Filosofie en sociologie van modellen. Dictaat 162340, Universiteit Twente.

De Roode, F.J., 1997. Milieufilosofie en het Waddenzee-beleid, Op zoek naar richtlijnen voor de belangenafweging tussen menselijk gebruik en de natuur in de Waddenzee. Verkrijgbaar bij V.N. de Jonge, RIKZ-Haren.

Smit, W.A., 1995. Veiligheid en risicomanagement I, bronnenboek. Dictaat 162362b, Universiteit Twente.

Van der Veen, A., Green, C., Pflügner, W., Wierstra, E., 199?. Appraisal methods as tools in evaluating responses to long term climate changes: Cost-benefit analysis and multi-criteria analysis. Scheduled to appear in the Journal of coastal engineering.

Bijlagen

Bijlage I, de onderzochte methoden

- Monetaire Evaluatiemethoden
 - Kostenbaten analyse
 - Kosteneffectiviteits analyse
- Presentatiemethoden
 - Scorekaartmethode
 - Community impact analyse
 - AMOEBE
- Multicriteria analyse (MCA)
 - Gewogen Sommering
 - Verwachtingswaarde methode
 - Concordantie analyse
 - Evamix methode
 - Regimemethode
 - Permutatiemethode
 - Meerdimensionale schaalanalyse
 - Goals achievement matrix methode
 - SAATY-methode⁴⁰

⁴⁰ Alleen een bepaald onderdeel van de SAATY methode is opgenomen, namelijk de berekening van de weegfactoren. Verder is SAATY in principe gelijk aan gewogen sommering waarbij de scores van alternatieven op criteria door de beslisser zelf geschat moeten worden.

Bijlage 2, de functies die afwegingsmethoden zouden kunnen vervullen

1. Rationeel beslissen
2. Handleiding voor “hoe te beslissen”
3. Informatie verschaffen over de preferenties van de beslisser, hulp bij het zoeken naar nog betere alternatieven
4. Verantwoording afleggen van de keuze, representatie van een gemaakte keuze
5. Inzicht verschaffen in eigen voorkeuren.
6. Structureren van de besluitvorming
7. Expliciete integratie van verschillende “onvergelijkbare” criteria
8. Reductie van alternatieven
9. Communicatie
10. Strategisch gebruik
11. Inschatten van onzekerheid
12. Gericht reduceren van onzekerheid
13. Onzekerheden opsporen (scenarioanalyse) en bedenken van mogelijke reactieplannen.
14. Het weergeven van gevolgen voor de belangengroepen
15. Het genereren van mogelijke compenserende maatregelen

Methode	1	2	3	4	5	6	7	8	9	10	11 ⁴¹	12	13	14	15
KBA	*	*		*				*	*	*					
KEA	*	*		*				*	*	*					
Scorekaartmethode	*								*	*					
Community impact analyse	*					*			*	*				*	
AMOEBE			*						*	*					
Gewogen Sommering	*	*	*	*	*	*	*	*	*	*					
Verwachtingswaarde methode ⁴²		*				*				*					
Concordantie analyse		*		*		*	*	*	*	*					
Evamix methode	*	*	*	*	*	*	*	*	*	*					
Regimemethode	*	*		*		*	*	*	*	*					
Permutatiemethode	*	*		*		*	*	*	*	*					
Meerdimensionale schaalanalyse		*				*				*					
Goals achievement matrix methode	*	*		*		*		*	*	*				*	*
SAATY-methode			*		*	*	*								

Tabel 11 Overzicht van functies die afwegingsmethoden zouden kunnen vervullen

⁴¹ Bij 11, 12 en 13 ontbreken sterretjes omdat bij geen enkele methode een onzekerheids- en gevoeligheidsanalyse wordt uitgevoerd.

⁴² De verwachtingswaarde methode, de meerdimensionale schaalanalyse en de SAATY-methode zijn aanvullingen op de gewogen sommering. Daarom wordt hier alleen de meerwaarde van deze methoden beoordeeld. De verwachtingswaarde methode geeft bijvoorbeeld een extra mogelijkheid tot strategisch gebruik, vandaar het sterretje bij punt 10.

Bijlage 3, de concordantie analyse, de proef op de som.

uitgangssituatie

Als uitgangssituatie is gekozen voor een set van drie alternatieven (A1, A2, A3) en vier criteria (C1 – C4).

Criteria\Alternatieven	A1	A2	A3
C1	40	60	80
C2	25	30	20
C3	2	1,5	1,75
C4	4	5	10

Tabel 12 De uitgangssituatie

Deze criteria worden vervolgens gestandaardiseerd en er worden gewichten toegekend.

Criteria\Alternatieven	A1	A2	A3	Gewichten
C1	-0,50	-0,75	-1	0,4
C2	0,83	1,00	0,67	0,2
C3	-1,00	-0,75	-0,88	0,1
C4	0,40	0,50	1,00	0,3

Tabel 13 De scores na standaardisatie

Vervolgens wordt er een concordantieverzameling bepaald (Zie Ministerie van financiën 1992 voor technische details).

$$K_{1,2} = 0,4$$

$$K_{2,1} = 0,2 + 0,1 + 0,3 = 0,6$$

$$K_{1,3} = 0,4 + 0,2 = 0,6$$

$$K_{3,1} = 0,1 + 0,3 = 0,4$$

$$K_{2,3} = 0,4 + 0,2 + 0,1 = 0,7$$

$$K_{3,2} = 0,3$$

Met deze verzameling kunnen scores bepaald worden met het volgende resultaat:

$$K_1 = K_{1,2} + K_{1,3} - K_{2,1} - K_{3,1} = 0$$

$$K_2 = K_{2,1} + K_{2,3} - K_{1,2} - K_{1,3} = 0,6$$

$$K_3 = K_{3,1} + K_{3,2} - K_{1,3} - K_{2,3} = -0,6$$

Hieruit blijkt, de prioriteitsvolgorde is: $A_2 > A_1 > A_3$

Als er nu een vierde alternatief (A4) wordt toegevoegd met scores: -0,25; 0,5; -0,80; 0,75, dan wordt de extra concordantieverzameling:

$$K_{1,4} = 0,2$$

$$K_{4,1} = 0,4 + 0,1 + 0,3 = 0,8$$

$$K_{2,4} = 0,2 + 0,1 = 0,3$$

$$K_{4,2} = 0,4 + 0,3 = 0,7$$

$$K_{3,4} = 0,2 + 0,1 + 0,3 = 0,6$$

$$K_{4,3} = 0,4 = 0,4$$

De nieuwe scores worden daarmee:

$$K_1 = 0 + 0,2 - 0,8 = -0,6$$

$$K_2 = 0,6 + 0,3 - 0,7 = 0,2$$

$$K_3 = -0,6 + 0,6 - 0,4 = -0,4$$

$$K_4 = 0,8 + 0,7 + 0,4 - 0,2 - 0,3 - 0,6 = 0,8$$

Zodat de prioriteitsvolgorde wordt: $A_4 > A_2 > A_3 > A_1$. Opvallend is dat A_1 en A_3 van positie zijn gewisseld. Blijkbaar kan door toevoeging van een alternatief de voorkeursvolgorde in de andere alternatieven wijzigen! Daarmee wordt het ook mogelijk om, doormiddel van het toevoegen van slim gekozen alternatieven, de methode zo te manipuleren dat het uiteindelijke resultaat overeenkomt met een vooraf gewenst resultaat.

Bijlage 4, de regimemethode, de proef op de som

Bij de regimemethode worden scores kwalitatief gemaakt in een paarsgewijze vergelijking, het enige dat telt is of een alternatief beter of slechter scoort dan het alternatief waarmee wordt vergeleken, op één criterium. Er moeten daarnaast kwalitatieve gewichten zijn vastgesteld. Er zijn twee methodes om de prioriteitsvolgorde te bepalen, de analytische en de numerieke. Beide methoden suggereren een grotere precisie dan ze in werkelijkheid garanderen.

Het analytische voorbeeld

Criteria\Alternatieve n	A1	A2	Gewichten
C1	0,99	1	1e
C2	0,98	1	2e
C3	1	0	3e

Tabel 14 De uitgangssituatie

Dit levert de volgende regimematrix op:

A1-----→ A2: -1 -1 +1

Omdat de gewichten 1 en twee belangrijker zijn dan drie (de gewichten staan kwalitatief weergegeven in een prioriteitsvolgorde) kan uit deze serie afgeleid worden dat alternatief A2 de voorkeur geniet. Immers, gewicht 3 is minder belangrijk dan de gewichten 1 en 2 en alternatief 2 scoort beter op deze twee criteria. Dit wordt de analytische methode genoemd. Echter, deze redenering gaat alleen op als de -1 scores op de criteria 1 en twee een even groot zin als de +1 op criteria 3. Dit is echter duidelijk niet het geval! De ‘analytische’ uitspraak $A2 > A1$ is dus helemaal niet zo zeker als de term “analytisch” suggereert.

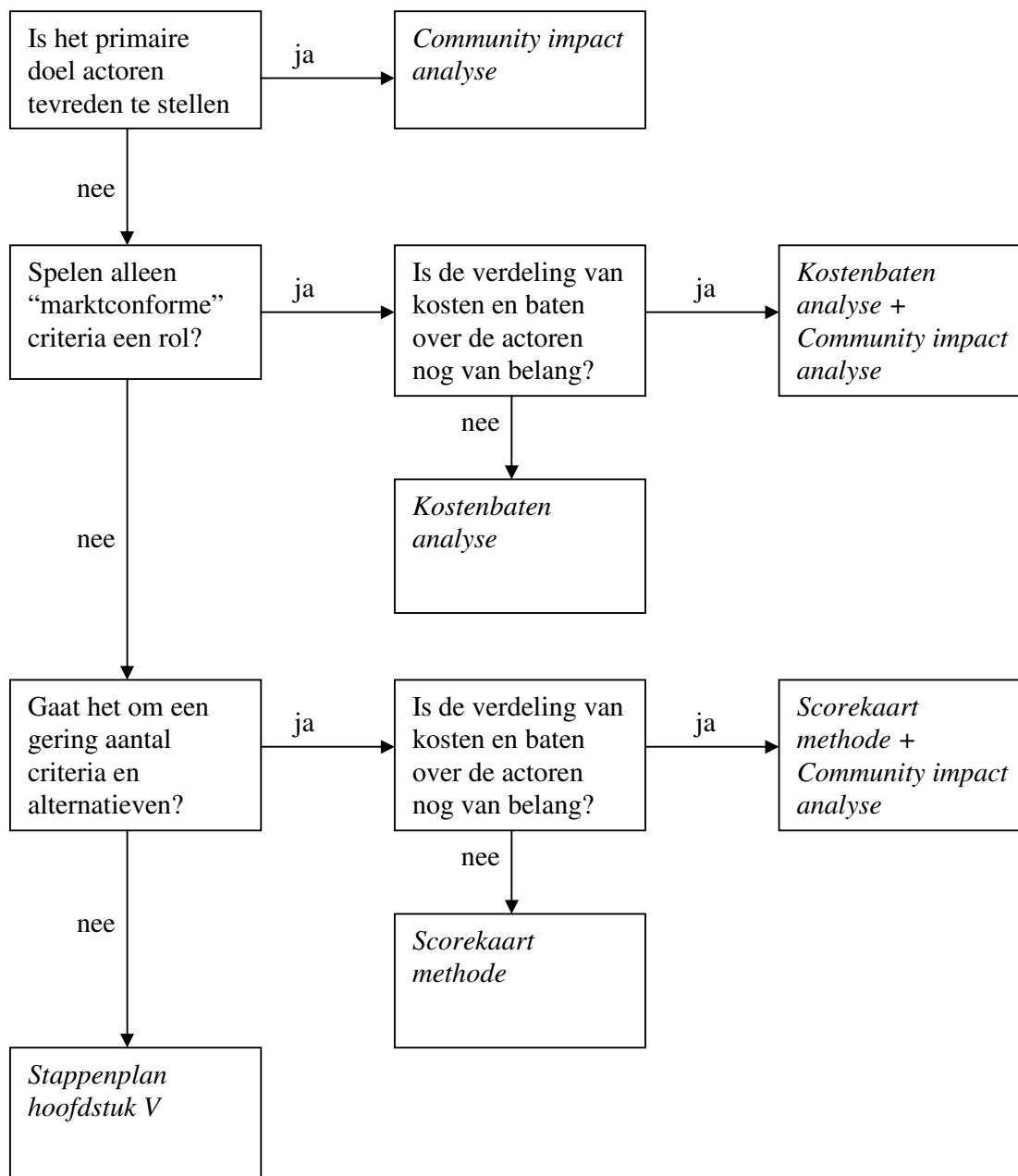
Soms kan er geen analytische uitspraak gedaan worden. Dit zou bijvoorbeeld het geval zijn als A1 ook beter op C2 zou scoren. In dat geval is A2 beter op het zwaarst wegende criterium en A1 op de twee minder zwaar wegende criteria. Als Criterium 1 zwaarder weegt dan 2 en 3 samen, wint A2, anders A1. Omdat nu niet analytisch is vast te stellen welk alternatief wint moeten er kansen worden ingeschat. Ook dit kan numeriek of weer analytisch. Het laatste overigens alleen bij een beperkt aantal criteria. Daarom wordt meestal de numerieke methode gebruikt. Daarbij worden er steeds random gewichten combinaties getrokken, die wel voldoen aan de prioriteitsvolgorde waarin de gewichten horen te staan. Met een groot aantal trekkingen kan nu, op basis van de wet van de grote getallen, een kansinterval geschat worden voor de alternatieven. Dit suggereert dat er een kans berekend is van x% dat alternatief A1 beter scoort dan A2. Dit is echter geenszins het geval, weer omdat de scoreverschillen weer even groot zouden moeten zijn. Zolang dit niet het geval is wijken ook de gepresenteerde kansen sterk af van de werkelijkheid.

Bijlage 5, de input van rijkswaterstaat

Rijkswaterstaat heeft tijdens gesprekken een aantal punten genoemd die gebruikt zijn in dit onderzoek. Dit zijn de volgende punten:

- Er is niet een duidelijk definieerbare gebruiker. Verschillende gebruikers met verschillende achtergronden op verschillende (aggregatie-)niveaus en vanuit verschillende instellingen zullen de methode moeten kunnen gebruiken.
- Er is niet één beslisser, in de meeste gevallen zullen meerdere beslissers tegelijkertijd van de methode gebruik moeten kunnen maken. (Dit punt is in dit onderzoek niet verder uitgewerkt.)
- Er kan niet alleen van monetaire criteria gebruik gemaakt worden, ook niet monetaire criteria moeten verwerkt kunnen worden.
- De criteria zullen niet alleen kwantitatief zijn, de methode moet ook met kwalitatieve criteria om kunnen gaan.
- Een methode die alleen de een de gevolgen van een alternatief presenteert voldoet niet. De criteria zullen ook expliciet geïntegreerd moeten kunnen worden.
- Waarschijnlijk zal de criteriagroep de criteria in criteriabomen plaatsen, met boven in criteria van een hoger abstractieniveau dan onderin.

Bijlage 6, een eenvoudige determinatietabel om de meest geschikte methode te selecteren.



Bijlage 7, een functionele decompositie van afwegingsmethoden

Stappen (Zie begin hoofdstuk II)	Opties
Omrekening van scores naar meer bruikbare scores	KBA en KEA: omrekenen naar geld; verwachtingswaarde methode*, meerdimensionale schaalanalyse*: kwantificering van kwalitatieve data met behulp van extra informatie.
Standaardisatie van de scores	Deling door rijmaximum*; deling door rijtotaal*; vectornormalisatie*; intervalstandaardisatie*; deling door doel- of ideaalwaarden.
Indeling van scores per actor	Community impact analyse; goals achievement matrix methode.
Vaststellen van weegfactoren	KBA: markt; SAATY: paarsgewijze vergelijking; directe toekenning via enquêtes of door analyse van verleden; extreme gewichtenmethode; random gewichtenmethode; verwachtingswaardemethode*; interactieve methoden; hypothetische gewichtensets. ⁴³
Berekening van een voorkeursvolgorde in de alternatieven	Gewogen somming; KEA; Concordantie analyse*; discordantie analyse*; regimemethode*; evamix methode*; permutatie methode*; Goals achievement matrix methode.
Presentatie	Verschilt per methode, meestal in ieder geval een scoretabel.

* Gebruik van deze methode wordt in dit rapport afgeraden.

⁴³ Voor alle methoden behalve KBA geldt dat de gewichten toegekend kunnen worden door de beslisser, maar ook door burgers of actoren. Dit is afhankelijk van wie gevraagd wordt, niet van de manier waarop gewichten worden vastgesteld.