



MNP Rapport 550031004/2006

**Gevoeligheidsanalyse van de bepaling van de  
duurzaamheidsindex**

Bijlage III bij het Methoderaapport Duurzaamheidsverkenning  
(MNP Rapport 550031001/2006)

Peter S.C. Heuberger en Peter H.M. Janssen

Contact:

Peter S.C. Heuberger

Peter.Heuberger@mnp.nl

## **SAMENVATTING**

De indicatoreselectie en indexconstructie binnen de Duurzaamheidsverkenning zijn aan een gevoeligheids- en onzekerheidsanalyse onderworpen. Met name de rol van schaling en weging zijn onderzocht. Verdere onderwerpen zijn o.a. de gevoeligheid t.g.v. onnauwkeurigheid in de basisgegevens, aantallen indicatoren, uitsplitsing naar domeinen, en proxy-keuze.

De belangrijkste conclusies zijn dat de indexconstructie met name op de enquête gebaseerd is en niet zozeer op een achterliggend systeemraamwerk; en dat de index sterk gevoelig is voor de keuze van het aantal indicatoren en de schaling.

**INHOUD**

III.1	Introductie .....	5
III.2	Basismateriaal .....	7
III.2.1	Inleiding .....	7
III.2.2	Standaard constructiemethode .....	8
III.2.3	Kanttelingen .....	8
III.2.3.1	Periode en indicatorbouw .....	9
III.2.3.2	Schaling langs de Y-as en ijkpunt .....	10
III.2.3.3	Positieve en negatieve werking .....	10
III.2.3.4	Weegfactoren.....	10
III.2.3.5	Keuze indicatoren/proxy's .....	11
III.2.4	Varianten op de duurzaamheidsindex.....	11
III.2.5	Aantal indicatoren en Index per domein.....	12
III.2.6	Redundantie .....	15
III.3	Schaling en versterking .....	17
III.3.1	Alternatieve schalingsmethoden .....	17
III.3.2	Versterking .....	18
III.3.3	Distance to Targets .....	19
III.4	Invloed van individuele indicatoren .....	21
III.4.1	Bijdrage per indicator .....	21
III.4.2	Gevoeligheid per indicator .....	21
III.5	Onzekerheidsanalyse .....	25
III.6	Conclusies .....	29
III.7	Referenties.....	31

## Lijst van Tabellen

Tabel 1. Vraagstukken, indicatoren en weegfactoren per wereldbeeld, .....	7
Tabel 2. Totaal Scores per domein en per wereldbeeld uit de 53 vragen in de TNS-NIPO Waarden en Normen-enquête. ....	14
Tabel 3. Totaal Scores per domein en per wereldbeeld, zoals gebruikt bij de bepaling van de duurzaamheidsindex. ....	14
Tabel 4. Principale componenten van gecentreerde en gestandaardiseerde indicatoren. ....	15
Tabel 5. Voorbeeld van verschil in versterking tussen min-max schaling en schaling op basis van gemiddelden. ....	17
Tabel 6. Transformatie per indicator X voor verschillende schalingsmethoden. ....	19
Tabel 7. Indicatoren gebruikt voor illustratie van Distance to Targets benadering.....	20
Tabel 8. Gemiddelde procentuele bijdrage per jaar, per indicator .....	23
Tabel 9. Standaarddeviaties voor weegfactoren bij vraagstukken.....	26
Tabel 10. Relatieve onzekerheidsgrenzen , gebruikt voor Monte Carlo analyse (in %). ....	27

## Lijst van Figuren

Figuur 1. Indicatoren uit het Ecologisch Domein, gebruikt voor duurzaamheidsindex. ....	33
Figuur 2. Indicatoren uit het Economisch Domein, gebruikt voor duurzaamheidsindex. ....	34
Figuur 3. Indicatoren uit het Sociaal-Culturele Domein, gebruikt voor duurzaamheidsindex. ....	35
Figuur 4. Verschillen in bevolkingsgerelateerde indicatoren en de corresponderende duurzaamheidsindex. ....	36
Figuur 5. Varianten op de berekening van de duurzaamheidsindex.....	37
Figuur 6. Duurzaamheidsindex met gelijk aantal (8) indicatoren per wereldbeeld en de duurzaamheidsindex uitgesplitst naar domein.....	38
Figuur 7. Duurzaamheidsindex voor de situatie dat, (1) alle 25 indicatoren worden meegenomen, (2) voor elke wereldbeeld de belangrijkste 12 indicatoren worden meegenomen en de uitsplitsing naar domein in dit laatste geval. ....	39
Figuur 8. Duurzaamheidsindices voor drie verschillende schalingsmethodieken.....	40
Figuur 9. Quotiënt van versterkingsfactoren voor verschillende schalingsmethoden. ....	40
Figuur 10. Versterkingsfactoren voor de standaard methode en de methode gebaseerd op gemiddelden. ....	41
Figuur 11. Illustratie van distance to targets benadering in vergelijk met standaardmethode en schaling op basis van gemiddelde. ....	41
Figuur 12. Bijdrage individuele indicatoren aan duurzaamheidsindex met schaling zoals in DV.42	
Figuur 13. Bijdrage individuele indicatoren aan duurzaamheidsindex met schaling op basis van het gemiddelde.....	43
Figuur 14. Duurzaamheidsindices voor drie verschillende schalingsmethodieken en de lineaire benadering van deze indexen op basis van lineaire regressie (gestippelde lijnen). ....	44
Figuur 15. Bijdragen in gevoeligheid voor de indicatoren <i>Stikstofgebruik in kunstmest</i> , <i>Levensverwachting</i> en <i>Aantal misdrijven per persoon</i> . ....	45
Figuur 16. Onzekerheidsmarges in duurzaamheidsindex voor standaardmethode en de methode gebaseerd op gemiddelden. ....	46
Figuur 17. Onzekerheidsmarges in verschillen tussen duurzaamheidsindices per wereldbeeld voor standaardmethode en de methode gebaseerd op gemiddelden. ....	47

## III.1 Introductie

In deze bijlage wordt een analyse gegeven van de samenstelling van de Duurzaamheidsindex<sup>1</sup> (DI) uit de Duurzaamheidsverkenning (DV) (RIVM, 2004; Nagelhout, 2006). Het doel van de studie is te onderzoeken in hoeverre de DI gevoelig is voor (veranderingen in) de onderliggende data en de keuze van de constructiemethode (schaling, weging, indicatorselectie). Daarbij zullen ook verschillen tussen wereldbeelden aan de orde komen. Ten slotte wordt geprobeerd een antwoord te geven op de vraag wat de resultaten van deze studie voor consequenties hebben op uitspraken in de DV.

Deze bijlage is als volgt opgebouwd:

In Hoofdstuk 2 wordt de constructie van de DI in het kort besproken, waarbij tevens een overzicht van de onderliggende indicatoren wordt gegeven. Tevens wordt aangegeven wat het resultaat is indien de duurzaamheidsindex wordt uitgesplitst naar de drie onderliggende domeinen (“de drie P’s”), het Ecologische, Economische en Sociaal-culturele domein. Daarnaast zal de rol van de weegfactoren aan de orde worden gesteld en komt de samenhang in de basisdata aan de orde.

Hoofdstuk 3 behandelt een aantal alternatieven voor de constructie van de duurzaamheidsindex, te weten schaling op basis van gemiddelden en schaling op basis van standaarddeviaties. Elke constructiemethode leidt tot andere versterkingsfactoren voor de individuele indicatoren. Tevens wordt in dit hoofdstuk middels een voorbeeld aangegeven wat het effect zou zijn indien een zogenaamde “Distance to Targets” benadering gevolgd zou worden.

In Hoofdstuk 4 wordt een gevoeligheidsanalyse gepresenteerd, waarbij onderzocht wordt in hoeverre de duurzaamheidsindex gevoelig is voor veranderingen in indicatorwaarden. Eerst wordt aangegeven wat de bijdrage is van de diverse indicatoren aan de duurzaamheidsindex.

Hoofdstuk 5 is gewijd aan een eenvoudige onzekerheidsanalyse, waarbij zowel aan indicatorwaarden als aan de weegfactoren onzekerheidsbanden worden toegekend. Door middel van Monte-Carlo Sampling wordt vervolgens een beeld gegeven van de corresponderende onzekerheidsmarge in de duurzaamheidsindex.

De auteurs bedanken Dick Nagelhout (MNP) voor zijn medewerking aan dit rapport, in diverse hoedanigheden, vraagbaak, dataleverantie, kritische reflectie etc.

---

<sup>1</sup> In feite is er sprake van 4 duurzaamheidsindices, per wereldbeeld (A1, A2, B1, B2) één.



## III.2 Basismateriaal

### III.2.1 Inleiding

Voor de DI, zoals in de DV gerapporteerd, zijn 25 indicatoren gebruikt, met gegevens voor de periode 1970-2000. Voor elk van de indicatoren zijn op basis van enquêtes weegfactoren per wereldbeeld bepaald. In deze notitie wordt de gevoeligheid van de bepaling van de weegfactoren op de berekening van de DI niet rechtstreeks onderzocht, omdat de daarvoor benodigde data niet voorhanden zijn. Wel is bij de onzekerheidsanalyse (paragraaf III.5) een onzekerheidsmarge aan elk van de weegfactoren toegekend. In Tabel 1 worden de indicatoren en weegfactoren weergegeven. In feite zijn er maar 23 verschillende indicatoren, omdat zowel het *werkloosheidspercentage* als het *aantal misdrijven per persoon* twee keer voorkomen. Voor de DV zijn per wereldbeeld slechts een beperkt aantal indicatoren meegenomen (de acht belangrijkste aangevuld met indicatoren waarvan de score voor een wereldbeeld significant verschillend is van de rest van de score voor de rest van Nederland, zie ook paragraaf III.2.5). De weegfactoren van de indicatoren die niet zijn meegenomen zijn in de tabel rood gedrukt. Merk op dat er diverse indicatoren zijn die niet meegenomen zijn, een hogere weegfactor hebben dan indicatoren die wel zijn meegewogen.

Tabel 1. Vraagstukken, indicatoren en weegfactoren per wereldbeeld.

Alleen de items die voor de berekening van de duurzaamheidsindex zijn gebruikt. Per wereldbeeld zijn slechts een beperkt aantal indicatoren gebruikt. De scores uit de TNS-NIPO enquête, die voor de berekening op nul zijn gezet, zijn in rood weergegeven. Zie ook paragraaf III.2.5.

Scores					Vraagstuk uit Enquêtes	Indicator
NL	A1	A2	B1	B2		
6.9	6.0	6.4	7.6	7.3	9 Waterveroontreiniging	N-gebruik in kunstmest
6.0	5.1	5.5	7.2	6.0	5 Broeikaseffect	Emissie broeikasgassen
2.9	4.9	3.2	1.8	2.5	19 Staatsschuld	Staatsschuld (%bbp)
3.9	4.6	4.7	2.2	3.7	37 Gezondheidszorg	Levensverwachting (jaren)
5.4	4.3	5.1	5.5	5.5	16 Ozonlaag	Conc pot. Cl en Br
2.6	4.3	3.6	1.3	2.3	49 Criminaliteit	Misdrijven per persoon
2.2	4.1	2.0	2.4	2.0	29 Onderwijs	Hoogopgeleiden HBO/WO (%)
3.1	3.7	3.5	1.9	2.9	26 Energievoorzieningszekerheid	Import energie index
1.3	3.3	1.5	0.8	0.9	18 Concurrentiepositie	Arb.kosten buitenl. tov NL
2.3	3.2	3.3	0.7	1.9	20 Collectieve Lastendruk	Collectieve lastendruk (%bbp)
1.9	2.9	2.4	0.9	1.5	36 Werkloosheid	Werkloosheid (%bev.)
3.4	3.0	4.5	1.9	3.4	44 Pensioenvoorziening	Dekkingsgraad pensioenvoorz.
4.6	3.6	4.2	5.5	5.1	8 Biodiversiteit	Living Planet Index
2.1	2.9	3.4	0.7	1.7	32 Culturele Diversiteit	Aantal Asielverzoeken
2.3	2.7	3.1	1.5	2.2	50 Naleving van regels	Misdrijven per persoon
2.1	1.7	2.9	1.0	2.1	25 Inkomenszekerheid	Werkloosheid (%bev.)
4.4	2.7	3.1	6.9	4.9	30 Honger	Ondervoede mensen (%wer. bev.)
3.5	2.7	2.3	6.0	3.9	46 Mensenrechten	Aantal functionerende democratieën
3.9	2.1	2.7	5.9	4.3	17 Internationale Samenwerking	OECD agr steun
2.8	1.7	1.6	5.5	3.3	48 Armoede	Mensen < \$1 dag (%)
3.4	2.6	2.6	5.1	3.7	15 Drinkwaterbeschikbaarheid	Mensen ztt schoon drinkwater (%)
2.7	1.8	1.4	4.8	3.0	42 Analfabetisme	Analfabeten (%wer. bev.)
3.0	1.6	2.2	4.7	3.5	28 Maatschappelijke Rol Grote bedrijven	Arbeidsinkomensquote NL
3.0	1.8	2.5	4.1	3.2	27 Olie en Gasvoorraden	Energieconsumptie
6.9	3.6	3.2	4.4	3.8	51 Gewapende conflicten en Terrorisme	Aantal terrorisme-incidenten

In Figuur 1 t/m Figuur 3 worden deze 23 verschillende indicatoren weergegeven, uitgesplitst naar domein. Merk op dat een groot aantal van deze indicatoren qua vorm sterk op elkaar lijken. Dat zal ook later in deze notitie nader aan de orde komen. Tevens wordt in deze figuren met de kleuren blauw en rood aangegeven welke data daadwerkelijk beschikbaar waren (blauw) en welke data door middel van interpolatie dan wel extrapolatie verkregen zijn. Het valt op dat voor een aantal indicatoren slechts een beperkte hoeveelheid 'gemeten' data voorhanden is en dat deze indicatoren derhalve een grote mate van onzekerheid kennen. Met groene lijnen is aangegeven wat de onder- en bovengrenzen zijn die in de DV gebruikt zijn. Zie ook voetnoot 2.

### III.2.2 Standaard constructiemethode

Bij het bepalen van de duurzaamheidsindex (*DI*) is voor de DV de volgende procedure gevolgd voor het schalen en wegen van de diverse indicatoren:

Gegeven een indicator  $X(t)$ , voor  $t=1970, 1971, \dots, 2000$ , laat  $X_{min}$  en  $X_{max}$  geven zijn door

$$X_{min} = \min\{X(1970), \dots, X(2000)\} \quad X_{max} = \max\{X(1970), \dots, X(2000)\}$$

Deze indicator  $X$  wordt geschaald tot een variabele  $Y$  op basis van het totale bereik in de periode 1970-2000. Voor de schaling worden 2 gevallen onderscheiden

- 1) Afnemende waarden van  $X$  "verbeteren de duurzaamheid"  $Y(t) = \frac{X(t) - X_{min}}{X_{max} - X_{min}}$
- 2) Oplopende waarden van  $X$  "verbeteren de duurzaamheid"  $Y(t) = 1 - \frac{X(t) - X_{min}}{X_{max} - X_{min}}$

Voor deze afgeleide indicator  $Y$  geldt nu dat  $0 \leq Y(t) \leq 1$  en "afnemende waarden van  $Y$  verbeteren de duurzaamheid", d.w.z.  $Y(t)=0$  is 'goed' en  $Y(t)=1$  is 'slecht'.

Aan elk van de indicatoren is op basis van de enquêtes een weefactor  $w$  (per wereldbeeld verschillend, zie Tabel 1) toegekend. De uiteindelijke duurzaamheidsindex (per wereldbeeld en per jaar  $t$ ) is bepaald middels

$$DI(t) = 100 - 100 \frac{\sum_k w_k Y_k(t)}{\sum_k w_k}$$

Deze formulering zegt in feite dat als alle indicatoren  $Y$  op hun 'beste' waarde (0) staan, dat dan de *DI* gelijk is aan 100 (goed). In het andere uiterste, alle  $Y$ 's op 1, wordt de *DI* gelijk aan 0 (slecht).

### III.2.3 Kanttekeningen

In het hierna volgende passeren een aantal varianten en alternatieve schalingen de revue, en wordt een inschatting gemaakt van de invloed van de diverse indicatoren op de berekening van de samengestelde *DI*. Daarbij worden de volgende kanttekeningen geplaatst:



### III.2.3.1 Periode en indicatorbouw

De bepaling van de samengestelde DI in de DV is geheel vastgebakken aan het tijdframe 1970-2000, waarbij de duurzaamheidscore – variërend tussen 0 en 100 – geheel op die periode gebaseerd zijn. Hiermee is deze index moeilijk bruikbaar voor prognoses naar de toekomst toe, omdat dit kan dit leiden tot negatieve waarden voor de DI. Ook kan hij slecht gebruikt worden voor de onderlinge vergelijking van verschillende landen, omdat de weging en schaling gebaseerd is op de specifieke situatie in een betreffend land (i.c. Nederland).

Een alternatieve aanpak, zie Hoofdstuk 3, maakt gebruik van een zogenaamde *Distance to Targets* benadering, waarbij per indicator een optimale situatie (Target) gedefinieerd wordt en het verschil tussen de waarde van de indicator en de Target vervolgens als maat voor de duurzaamheid wordt gebruikt. Hoewel bij deze methode ook sprake is van een zekere willekeur (keuze van de targets), is het voordeel dat de methode niet afhankelijk is van de periode en tevens voor andere landen gebruikt zou kunnen worden. Aan de andere kant kan men stellen dat er voor verschillende landen en wereldbeelden verschillende targets zouden kunnen gelden. Voor wereldbeelden is dit ook wat men zou verwachten. Verschillen tussen landen maken een vergelijking van de duurzaamheid juist weer twijfelachtig.

Bij TELOS wordt voor de indicatorendefinitie een meer gedifferentieerde indexering gebruikt die niet gekoppeld is aan een tijdframe: het waardenbereik van een indicator wordt in een aantal trajecten/zones opgesplitst, gerelateerd aan (normatief vastgestelde) grens- en streefwaarden (normering) die een mate van wenselijkheid, acceptabiliteit aanduiden, lopend van 0 tot 100 (zie Haarman et al. 2004, Lemmens et al. 2003). De indexering van de indicator vindt over deze trajecten op lineaire wijze plaats, en levert in feite een stuksgewijs-lineaire benadering van niet-lineaire herschaling over het oorspronkelijk waardenbereik, die bovendien een Distance-to-Target karakter heeft (w.b. gebruik van streefwaarden).

Voor de samenstelling van de DI uit de afzonderlijke indicatoren wordt gebruik gemaakt van een lineaire, additieve benadering. Invloeden van synergie, koppelingen, versterkingen en verzwakkingen worden buiten beschouwing gelaten, terwijl de aggregatie bovendien suggereert dat er eenvoudige compensatie/uitruil kan plaatsvinden tussen de indicatoren (hoge waarden van de ene indicator kunnen door lage waarden van anderen gecompenseerd worden). Dit alles kan het zicht ontnemen op mogelijke relevante conflicten en cruciale keuzes rond duurzaamheidsvraagstukken (zie ook Hueting en Reijnders, 2004). Er wordt ook niet gecorrigeerd voor mogelijke dubbeltellingen. Uit het verloop van de indicatoren (Figuur 1 t/m Figuur 3) is op te maken dat meerdere indicatoren een vrijwel identiek verloop hebben.

Toekomstige (partiële) verbeteringen zouden mogelijk zijn door een (niet-lineaire) herschaling van indicatoren waarbij rekening gehouden wordt met het feit dat de impact van relatieve veranderingen in een indicator kunnen afhangen van het niveau waarop hij zit (zie ook Nardo *et al.*, 2005). Bijvoorbeeld, als veranderingen belangrijker zijn bij lage waarden van de indicator dan zou herschaling met een concave functie voor de hand liggen (bijv. log of n-de machts wortel). Als ze meer effect hebben bij hoge waarden, dan een convexe vorm (bijv. exponentieel of machten). In feite probeert men zo als het ware een soort nieuwe, gemeenschappelijke meetschaal te creëren, gericht op het meten van de significantie of impact van de betreffende indicator (voor duurzaamheid). Bedenk hierbij wel dat dit doorgaans maar beperkt mogelijk is, enerzijds omdat er over de onderlinge causaliteit en doorwerking onvoldoende bekend is (complexiteit en onzekerheid), en anderzijds ten gevolge van de subjectieve/normatieve invulling van duurzaamheid. Zie ook Elliott, 1981. Vergelijk ook de hierboven vermelde stuksgewijs-lineaire herschalingsstrategie die door TELOS wordt toegepast (Haarman et al. 2004, Lemmens et al. 2003).

### III.2.3.2 Schaling langs de Y-as en ijkpunt.

Bij deze studie dient men zich te realiseren dat het ijkpunt waarnaar gekeken wordt van belang kan zijn. Indien bijvoorbeeld de invloed van een indicator  $X(t)$  op  $DI(t)$  gedefinieerd wordt als  $DI(t)/X(t)$ , dan levert dit andere resultaten bij bovenstaande definitie van de  $DI$ , dan indien de volgende definitie gebruikt zou worden (100=slecht, 0=goed):

$$DI(t) = 100 \frac{\sum_k w_k Y_k(t)}{\sum_k w_k}$$

In feite kan gesteld worden dat het alleen van belang is hoe de  $DI$  varieert over de tijd en hoe de  $DI$ 's van de vier wereldbeelden ten opzichte van elkaar verlopen. In de hierna volgende analyses is rekening gehouden met dit effect. Om deze reden is de factor 100 in de navolgende

berekeningen niet gebruikt en is in feite gerekend met  $DI(t) = -\frac{\sum_k w_k Y_k(t)}{\sum_k w_k}$ . Voor de meeste

figuren heeft dit geen effect, omdat daar bewust de y-as geen getallen bevat. Voor een enkele figuur en tabel worden echter wel 'harde' getallen gebruikt.

### III.2.3.3 Positieve en negatieve werking

Een tweede probleem is het onderscheid dat gemaakt wordt naar indicatoren die bij oplopende waarden de duurzaamheid verbeteren dan wel verslechteren. Bij de voor de DV gebruikte methode kan middels voorgaande procedure gegarandeerd worden dat de afgeleide indicatoren  $Y$  altijd positief blijven. Bij andere schalingen, bijvoorbeeld rond het gemiddelde, is dit niet meer mogelijk met als gevolg dat indicatoren met zowel positieve als negatieve waarden voorkomen.

Een eenvoudige manier om dit onderscheid (verbeteren/verslechteren) uit de data te halen is door voor indicatoren waarbij oplopende waarden de duurzaamheid verbeteren  $X(t)$  te vervangen door zijn tegengestelde,  $-X(t)$ . Hierdoor kan voor de standaard toegepaste constructiemethode het onderscheid in het bepalen van de afgeleide indicator  $Y$  vervallen:

Als  $X(t)$  vervangen wordt door  $Z(t) = -X(t)$ , dan geldt dat  $X_{max} = -Z_{min}$  en  $X_{min} = -Z_{max}$  en derhalve

$$Y(t) = 1 - \frac{X(t) - X_{min}}{X_{max} - X_{min}} = \frac{X_{max} - X(t)}{X_{max} - X_{min}} = \frac{-Z_{min} + Z(t)}{-Z_{min} + Z_{max}} = \frac{Z(t) - Z_{min}}{Z_{max} - Z_{min}}$$

Het volgt dat nu alle indicatoren op gelijke wijze behandeld kunnen worden. In de analyses is deze herdefinitie consequent gehanteerd.

### III.2.3.4 Weegfactoren

De mate waarin de diverse indicatoren een rol spelen bij de bepaling van de duurzaamheidsindex is gebaseerd op de TNS-NIPO enquête naar maatschappelijke vraagstukken en wereldbeelden. Er zal hier niet verder in op de onderliggende procedure worden ingegaan, maar wel opgemerkt dat de respondenten geen kennis hadden van de indicatoren die gebruikt zijn voor de constructie van de  $DI$ . Zij hebben alleen hun mening kunnen geven over de rol van de diverse maatschappelijke

vraagstukken. De vraag is gerechtvaardigd of zij niet een andere mening/waardering gegeven zouden hebben indien zij niet alleen de vraagstukken, maar ook de bijbehorende indicatoren voor ogen hadden gehad. Stelt de gemiddelde respondent zich bij het vraagstuk Gezondheidszorg niet iets anders voor dan bij de daarvoor gebruikte indicator *Levensverwachting in jaren?* Zou hij dat vraagstuk, wetend dat deze proxy gebruikt zou worden, mogelijk niet veel minder belangrijk kunnen vinden dan nu het geval is geweest?

### III.2.3.5 Keuze indicatoren/proxy's

Dit leidt tevens tot de vraag wat de rol van de keuze van de indicatoren (proxy's) is op de samenstelling en uitkomsten van de DI. Het is op dit moment niet mogelijk om alternatieve proxy's in deze studie op te nemen, omdat daar eenvoudigweg de data voor ontbreken. Voor een aantal indicatoren kan wel een eenvoudig alternatief gebruikt worden. Dit betreft de zes indicatoren die geformuleerd zijn als percentage van de bevolking (Nederland of mondiaal). Deze kunnen ook als absolute aantallen worden meegenomen. In Figuur 4 is het verloop van deze indicatoren weergegeven, zowel in relatieve als absolute termen, teruggeschaald naar [0 1] volgens de min-max schaling. Tevens is in deze figuur het verloop van de duurzaamheidsindex weergegeven, die volgt uit de keuze tussen relatief of absoluut. Het blijkt dat er wel degelijk verschillen zijn in de indicatoren, maar dat de verschillen in de resulterende duurzaamheidsindex klein zijn.

Het verdient aanbeveling om bij toekomstige duurzaamheidsverkenningen in een veel eerder stadium aandacht te besteden aan data-acquisitie. Daarbij zou de vraag aan de orde moeten komen of men de respondenten een beeld zou moeten geven van de mogelijke proxy's die aan een vraagstuk gekoppeld kunnen worden. In feite zou dat niet nodig moeten zijn, als ervoor gezorgd wordt dat een proxy gekozen wordt die het betreffende maatschappelijke vraagstuk goed in kaart brengt. Probleem daarbij is vaak dat er lang niet altijd voldoende (historische) data voorhanden zijn. Zie ook paragraaf III.2.6.

### III.2.4 Varianten op de duurzaamheidsindex

Om de invloed van een aantal keuzes te laten zien is de berekening van de DI ook uitgevoerd met een aantal varianten. Hiervoor zijn de volgende opties gekozen:

- a) Met de echte maxima en minima over de periode 1970-2000<sup>2</sup>
- b) Met de echte maxima en minima over de periode 1980-2000.
- c) Met alle (positieve) weegfactoren gelijk gesteld aan 1
- d) Met een extra "importantieweging"

<sup>2</sup> Voor de DV zijn niet de 'echte' maxima en minima gebruikt, maar zijn daar door Dick Nagelhout iets afwijkende waarden ingevuld. Zie de groene lijnen in Figuur 1 t/m Figuur 3. Dit kan van invloed zijn op de gevoeligheid en bijdrage van die indicator. Een voorbeeld hiervan vormt de indicator "*levensverwachting in jaren*". Deze varieert van 74.2 in 1970 tot 78.3 in 2000. Voor de DV zijn echter de waarden 73 en 79 als grenswaarden gebruikt. Dit heeft tot gevolg dat de bijbehorende afgeleide indicator  $Y(t)$  niet varieert tussen 0 en 1, maar tussen 0.200 en 0.883.

Het gevolg is dat lage waarden van deze indicator te zwaar worden meegenomen in de berekening van de duurzaamheidsindex, terwijl hoge waarden juist te weinig meetellen. Een ander voorbeeld is de "*Import energie Index*", waar het daadwerkelijke minimum/maximum gelijk is aan 239.7 respectievelijk 417.4, maar waarbij voor de DV gebruik is gemaakt van 250 en 450. Dit heeft tot gevolg dat deze indicator na schaling varieert tussen -0.051 en 0.837.

Bij deze laatste optie zijn (ter illustratie) aan de indicatoren labels toegekend:

- Minder belangrijk (extra weging 1)      2 indicatoren, Vraagstukken met nrs. 26, 51
- Belangrijk (extra weging 3)              8 indicatoren, nrs. 19, 37, 29, 20, 46, 17, 42, 28
- Zeer belangrijk (extra weging 5)        15 indicatoren, overige vraagstukken

De individuele scores zijn met de extra weging vermenigvuldigd.

De resultaten zijn grafisch weergegeven in Figuur 5. Daarbij vallen de volgende zaken op:

- Ad a) Alleen de index voor wereldbeeld B2 verloopt duidelijk anders en komt 'hoger' te liggen
- Ad b) Voor wereldbeeld B1 is geen sprake meer van een stijging over de periode 1970-1980.
- Ad c) Aanpassing van de weegfactoren heeft alleen een sterk effect op de index voor B2.  
Deze komt vanaf begin jaren 80 veel hoger te liggen
- Ad d) De extra importantieweging lijkt weinig invloed te hebben. Alleen voor B2 is een kleine verandering zichtbaar. Dit is in overeenstemming met het beeld bij c). Vermoedelijk wordt dit veroorzaakt door de grote redundantie in de indicatordata, zie ook paragraaf III.2.6

### **III.2.5 Aantal indicatoren en Index per domein**

In Figuur 1 t/m Figuur 3 is aangegeven tot welk domein (respectievelijk ecologisch, economisch en sociaal-cultureel) de diverse indicatoren die voor de DI gebruikt zijn, behoren. Het valt daarbij op dat het aantal indicatoren per domein nogal verschilt (5 ecologische, 8 economische en 10 sociale indicatoren).

Daarnaast valt uit Tabel 1 op te maken dat het aantal indicatoren dat per wereldbeeld 'meetelt' ook nogal verschilt. Het 'hoe en waarom' van de keuze van het aantal indicatoren per wereldbeeld is niet helemaal helder. Er zijn per wereldbeeld 8 indicatoren genomen en voor sommige zijn daar nog een aantal indicatoren aan toegevoegd<sup>3</sup>, om divergentie tussen de wereldbeelden te verkrijgen. De vraag rijst waar dat aantal van acht op gebaseerd is (bijv. waarom geen 7 of 9 etc.) en in hoeverre het streven naar divergentie legitiem is.

Om te onderzoeken in hoeverre dit van invloed is, is de duurzaamheidsindex ook berekend voor de situatie waarbij het aantal indicatoren per wereldbeeld gelijk is. Omdat voor wereldbeeld B2 de minste indicatoren (8) meewegen, is voor de andere wereldbeelden ook gekozen voor de 8 belangrijkste (hoogste weegfactor) indicatoren. Tevens is de duurzaamheidsindex per domein bepaald.

Deze resultaten zijn weergegeven in Figuur 6, waarbij in elke plaatje tevens is aangegeven hoeveel indicatoren per wereldbeeld van belang zijn.

Het blijkt dat het weglaten van indicatoren een behoorlijke invloed heeft<sup>4</sup>,

<sup>3</sup> In de DV wordt dit verklaard in Bijlage 2 op pagina 207: "Voor de constructie van de duurzaamheidsindex in *figuur 2.8.1* (zie *hoofdstuk 2*) zijn de indicatoren gewogen naar het belang dat in de wereldbeelden aan de onderwerpen wordt toegekend. In de lijst hierna worden de meest belangrijke duurzaamheidsvraagstukken per wereldbeeld aangegeven, inclusief de significant afwijkende onderwerpen. In de kolom is het belang van het onderwerp aangegeven, zoals dat blijkt uit de enquête van TNS-NIPO. Dit belang is gebruikt voor de weging van de verschillende indicatoren tot de samengestelde duurzaamheidsindex uit *figuur 2.8.1*.

<sup>4</sup> De index voor wereldbeeld B2 verandert niet omdat het aantal indicatoren hiervoor gelijk is gebleven.

- De indices krijgen een gladder, minder grillig verloop
- Het verschil tussen wereldbeelden A1 en A2 voor de eerste 10 jaren verdwijnt. Beide krijgen een piekerig gedrag rond 1975, wat veroorzaakt lijkt te worden door het weglaten van de *import energie index*. De negatieve pieken die in de oorspronkelijke DI voor A1 in 1988 en voor A2 in 1994 voorkwamen, worden veroorzaakt door de *collectieve lastendruk*, respectievelijk het *aantal asiolverzoeken*.
- Voor wereldbeeld B1 worden de waarden in de eerste 15 jaar hoger, ten gevolge van het weglaten van het *energieverbruik* en het *%mensen zonder toegang tot schoon drinkwater*.

Indien wordt gekeken naar de verdeling over de domeinen, dan valt op:

- In het ecologische domein geeft de samengestelde index weinig onderscheid tussen de wereldbeelden onderling. Alleen voor B1 daalt de duurzaamheid iets minder snel dan voor de overige wereldbeelden.
- In het economische domein blijkt alleen B2 een duidelijk andere bijdrage te hebben. Dat komt omdat in dit wereldbeeld slechts één economische indicator mee weegt, de *OECD agrarische steun*.
- In het sociaal-culturele domein treedt een sterke toename op voor B1 en B2.

Blijkbaar is het aantal indicatoren dat wel/niet wordt meegenomen in de index sterk bepalend voor het beeld. Ter illustratie zijn tevens berekeningen gedaan waarbij

- 1) alle 25 indicatoren zijn meegenomen, waarbij ook het Nederlands gemiddelde is bepaald<sup>5</sup>.
- 2) de 12 indicatoren die de hoogste score hebben (per wereldbeeld) zijn meegenomen. Voor dit geval is tevens de uitsplitsing naar domein bepaald.

De resultaten van deze berekeningen zijn weergegeven in Figuur 7. Het blijkt dat bij de eerste berekening alleen wereldbeeld B1 echt afwijkt van de overige, terwijl de index voor B2 vrijwel gelijk ligt met de index voor Nederland totaal. Dit is te verklaren uit het feit dat volgens de DV 45% van Nederland dit wereldbeeld aanhangt.

Bij de tweede berekening valt op dat de resultaten voor A1 en A2, respectievelijk B1 en B2 zeer dicht bij elkaar liggen. Verder is het interessant op te merken dat –hoewel voor wereldbeeld B1 in de DV ook 12 indicatoren gebruikt werden, het verloop bij deze laatste berekening duidelijk afwijkt. Dit wordt veroorzaakt doordat er andere indicatoren zijn meegenomen, zoals is op te maken uit Tabel 1.

In dit kader is het interessant om ook de verdeling over de domeinen te kijken die uit de enquête volgt. In Tabel 2 staan de totalen van de scores per domein weergegeven over alle 53 vragen uit de TNS-NIPO Maatschappelijke-Vraagstukken enquête<sup>6</sup>. Hierbij zijn dus de scores voor alle vragen over ecologie samengeteld etc. Indien deze tabel als maatstaf wordt genomen voor het belang dat mensen met een bepaald wereldbeeld hechten aan een bepaald domein dan blijkt dat alleen aanhangers van wereldbeeld A1 een beetje afwijken van het algemene beeld. Verder valt op dat de verdeling over de domeinen bij alle wereldbeelden redelijk gelijk is.

---

<sup>5</sup> Voor de DI zijn slechts 25 indicatoren gebruikt, hoewel de TNS-NIPO enquête uitging van 53 vragen. Dit is het gevolg van de keuze van de acht belangrijkste, aangevuld met significant verschillende. Voor de overige vragen waren bij het schrijven van deze notitie onvoldoende betrouwbare indicatordata voorhanden.

<sup>6</sup> De berekening is gedaan op basis van de data in het bestand “file voor gert eggink voor index.xls”.

Dit beeld wordt niet duidelijk terug gevonden in de berekening van de DI. Tabel 3 presenteert de totalen van de scores van de vragen die voor de bepaling van de DI gebruikt zijn. Het volgt dat de onderlinge verdeling over de domeinen alleen voor wereldbeeld A1 enigszins overeenkomt met de verdeling in Tabel 2. Voor de overige wereldbeelden zijn de verschillen groot. De conclusie is dus dat de weging die voor de DI gebruikt is, niet overeenkomt met de weging voor het geval waarbij alle 53 vragen zouden zijn meegenomen bij de berekening van de duurzaamheidsindex. Economie wordt bij 3 van de 4 wereldbeelden veel minder zwaar meegeteld in de totaalindex en ook het sociaal-cultureel domein blijft onderbelicht. Door de specifieke selectie van de indicatoren (8 belangrijkste etc.) krijgt het ecologisch domein een sterkere invloed, met name omdat het ecologisch domein oververtegenwoordigd is in de hoge scores.

**Tabel 2. Totaal Scores per domein en per wereldbeeld uit de 53 vragen in de TNS-NIPO Waarden en Normen-enquête.**

(afgerond op 1 decimaal).

	NL	A1	A2	B1	B2
<b>Ecologie</b>	40.9	37.0	40.2	42.0	41.8
<b>Economie</b>	34.9	38.5	35.4	34.2	34.3
<b>Sociaal</b>	44.1	44.4	44.4	43.8	43.9
<b>Totaal</b>	120	120	120	120	120

**Tabel 3. Totaal Scores per domein en per wereldbeeld, zoals gebruikt bij de bepaling van de duurzaamheidsindex.**

	A1	A2	B1	B2
<b>Ecologie</b>	15.4	21.2	30.9	23.9
<b>Economie</b>	15.1	11.3	14.7	4.3
<b>Sociaal</b>	15.9	17.7	23.2	12.6
<b>Totaal</b>	46.4	52.2	68.8	40.8

Hieruit kunnen de volgende conclusies getrokken worden:

- Selectie van welke indicatoren wel/niet worden meegenomen in de index kan het beeld sterk bepalen.
- Het beeld van de afzonderlijke duurzaamheidsdomeinen kan sterk verschillen. Enkel kijken naar het totaalbeeld (oorspronkelijke index waarin alle domeinen vertegenwoordigd zijn) verdoezelt deze verschillen voor een groot deel.<sup>7</sup>
- De keuze van indicatoren voor de DI geeft een zeer afwijkend beeld in termen van de verdeling over de domeinen in vergelijking met het geheel aan enquêtevragen.<sup>8</sup>

### Toedeling aan domeinen

Een vraag, die helaas niet meer gemakkelijk te beantwoorden is, betreft hoeveel invloed de keuze van het domein waaraan een indicator c.q. vraagstuk gekoppeld is, heeft op de index (per domein). Voor de meeste indicatoren c.q. vraagstukken is de keuze tot welk domein ze behoren voor de hand liggend, maar er zijn enkele gevallen, waar discussie over mogelijk is. Zo is werkloosheid in de DV beschouwd als een sociaal-culturele indicator, maar het is te

<sup>7</sup> Dit is bij de DV ook onderkend, maar er is bewust voor gekozen.

<sup>8</sup> Dit is mede het gevolg van het streven naar divergentie tussen wereldbeelden..

beargumenteren dat dit eigenlijk ook als een economische indicator beschouwd kan worden. Waterbeschikbaarheid wordt in de DV als een ecologische indicator gedefinieerd, maar zou net als armoede ook als sociaal-cultureel gezien kunnen worden. Het effect van een dergelijke verschuiving is niet goed meer na te gaan, omdat deze zal lijden tot andere uitkomsten in de enquête-scores die de uiteindelijke indicatorkeuze en indexconstructie bepalen. Het is goed mogelijk dat bij een andere domeintoedeling een bepaald vraagstuk bij veel respondenten niet meer in de 'domein-top 5' terecht komt met als gevolg dat de uiteindelijke score voor dit vraagstuk lager zal uitvallen. Andersom kan het goed zijn dat een vraagstuk na herindeling in een ander domein veel vaker tot de top-5 gaat behoren en derhalve juist een hogere score krijgt. In feite zou er een nieuwe enquête nodig zijn om de invloed van domeintoedeling goed te kunnen analyseren. Wel kan gesteld worden dat de index per domein (zie de conclusies hierboven) zeer gevoelig is voor het aantal indicatoren. In dit geval kan het verschuiven van indicatoren tussen domeinen ook van grote invloed zijn op de index per domein.

### III.2.6 Redundantie

Om te onderzoeken in hoeverre er sprake is van redundantie of sterke correlatie tussen de diverse indicatoren, zoals zou kunnen worden afgeleid uit de weergave van de individuele indicatoren in Figuur 1 t/m Figuur 3, worden in deze paragraaf de resultaten weergegeven van een eenvoudige Principale Componenten Analyse (PCA). Deze analyse is uitgevoerd door middel van een Singuliere Waarden Ontbinding van de matrix van indicatorwaarden, zonder inachtneming van de weegfactoren. Om artefacten te voorkomen worden de indicatoren daartoe eerst gecentreerd rond hun gemiddelde en gestandaardiseerd op hun standaarddeviatie.

Indien ervan uitgegaan wordt dat alleen singuliere waarden met een waarde groter dan 5%, of 10% van de grootste singuliere waarde daadwerkelijk iets bijdragen<sup>9</sup>, dan blijkt dat er slechts een beperkt aantal 'onafhankelijke' bijdragen zijn:

Tabel 4. Principale componenten van gecentreerde en gestandaardiseerde indicatoren.

Wereldbeeld	Aantal indicatoren	Aantal onafhankelijke componenten	
		5 %	10 %
A1	11	7	5
A2	12	7	5
B1	12	5	3
B2	8	5	3

Deze tabel kan als volgt geïnterpreteerd worden. De matrix met 11 indicatoren, die voor wereldbeeld A1 gebruikt worden, kan benaderd worden door een matrix met slechts 7 variabelen (combinaties van de 11 originele indicatoren), dusdanig dat de benaderingsfout kleiner is dan 5 %.

Er kan dus gesteld worden dat een aantal indicatoren (deels) hetzelfde effect meten. Dit is natuurlijk het geval (zie Tabel 1), voor de vraagstukken *Criminaliteit* en *Naleving van Regels*, die beide gebruik maken van dezelfde indicator, namelijk *Het aantal ter kennis van politie gekomen*

<sup>9</sup> Dit komt neer op een relatieve kwadratische fout van maximaal 5, respectievelijk 10 procent.

*misdrijven per persoon*. Merk op dat beide vraagstukken meegewogen worden voor wereldbeeld A2, met als effect dat deze indicator in feite het zwaarste meeweegt in dit wereldbeeld. Hetzelfde geldt voor de vraagstukken *Werkloosheid* en *Inkomenszekerheid*, die beide gebruik maken van de indicator *%werkelozen*, zij het dat deze vraagstukken in geen enkel wereldbeeld allebei meetellen. In het rapport van Saltelli *et al.* (2002) wordt aangegeven dat er geen eenduidig antwoord is op de vraag hoe men met dit soort situaties dient om te gaan. Enerzijds kan men stellen (Saisana *et al.*, 2005), dat deze sterke samenhang intrinsiek te wijten is aan onderliggende mechanismen van het systeem, die ervoor zorgen dat correlaties in de indicatorwaarden onvermijdelijk zijn ("System is Leading"). Een meer statistisch standpunt zegt juist dat deze correlaties verwijderd moeten worden of dat er andere indicatoren beschouwd zouden moeten worden ("Data is Leading").<sup>10</sup>

In dit kader is het natuurlijk zeer interessant om de keuze van proxy's nader tegen het licht te houden. Voor de diverse vragen/onderwerpen uit de enquêtes zijn meerdere indicatoren mogelijk en andere keuzes zouden tot een ander verloop van de duurzaamheidsindex kunnen leiden, zie de rapportage (Nagelhout, 2006) over de indicatoren. (bijvoorbeeld voor armoede is nu als indicator gekozen voor het aantal mensen met minder dan 1\$ per dag. In (Nagelhout, 2006) wordt een reeks aan alternatieven voorgesteld, zoals het verleggen van de grens naar 2\$ per dag of de Nederlandse uitgaven aan ontwikkelingshulp. Voor werkloosheid zou een alternatief zijn om naar langdurige werkloosheid te kijken. Voor mensenrechten wordt het aantal functionerende democratieën gebruikt, maar het was ook mogelijk geweest dit nader uit te diepen naar het aantal inwoners, etc.). Hierbij speelt vooral de vraag in hoeverre een indicator/proxy als representatief kan worden gezien voor het betreffende maatschappelijk vraagstuk.

---

<sup>10</sup> Het bleek niet mogelijk om met de beschikbare software een ook een factoranalyse uit te voeren op de data, omdat de data te sterk gecorreleerd zijn.



### III.3 Schaling en versterking

#### III.3.1 Alternatieve schalingsmethoden

Een van de mogelijke kritiekpunten op de DI, zoals gepresenteerd in de DV is de onderliggende methodiek van schaling, gebaseerd op het bereik van de onderliggende indicatoren. Deze methode heeft als gevolg dat indicatoren die slechts weinig variëren in feite worden 'opgeblazen', terwijl indicatoren met een groot bereik juist worden 'verzwakt'<sup>11</sup>. In dit hoofdstuk worden een tweetal alternatieven voor deze schalingsmethodiek beschouwd, op basis van

- het gemiddelde over de periode 1970-2000
- de standaard deviatie over de periode 1970-2000

**Ad a)** Hierbij geldt dat een indicator  $X(t)$  vervangen wordt door de afgeleide indicator  $Y(t)$ , met

$$Y(t) = \frac{X(t) - M}{|M|} \text{ met } M = \frac{1}{T} \sum_{t=1970}^{2000} X(t), \quad T = 31$$

Hoewel deze methode in de praktijk niet vaak gebruikt wordt, hebben is deze hier toch toegepast omdat bij deze methode een van de belangrijke kritiekpunten op de min-max methode van de DV, te weten het 'opblaas effect' teniet wordt gedaan<sup>12</sup>. Indien een indicator  $X_1$  varieert tussen 99 en 101, en een indicator  $X_2$  tussen 0 en 2, dan worden deze variaties in de DV als even belangrijk beschouwd, terwijl ze bij de schaling op basis van het gemiddelde heel anders gewogen worden, zoals uit onderstaande Tabel 5 volgt:

**Tabel 5. Voorbeeld van verschil in versterking tussen min-max schaling en schaling op basis van gemiddelden.**

Variatie in indicator	Variatie in afgeleide indicator	
	Op basis van min-max schaling	Op basis van gemiddelden
99 - 101	0 - 1	-0.01 - 0.01
0 - 2	0 - 1	-1 - 1

**Ad b)** Hierbij wordt  $X(t)$  vervangen door

$$Y(t) = \frac{X(t) - M}{S} \text{ met } S^2 = \frac{1}{T-1} \sum_{t=1970}^{2000} (X(t) - M)^2$$

<sup>11</sup> Een van de argumenten om deze kritiek te weerleggen is dat indicatoren alleen gebruikt worden indien de veranderingen belangrijk zijn. Zo is bijvoorbeeld een temperatuurverschil van 1 graad weliswaar klein, maar zeker significant met betrekking tot klimaatverandering. Probleem bij de min-max schaling blijft echter dat een verschil over het tijdtraject van 1 graad, na normalisatie, even zwaar weegt als een verschil van 3 graden.

<sup>12</sup> Ook op deze methode is van alles af te dingen. De waarde van de indicator wordt gemeten, relatief t.o.v. het gemiddelde en hierdoor wordt de uiteindelijke waarde sterk afhankelijk van de definitie van de indicator. Bijvoorbeeld, indien in plaats van *levensverwachting in jaren* gekeken wordt naar *stijging van de levensverwachting sinds 1950*, dan zal bij de min-max schaling geen verschil optreden, maar bij de schaling op basis van het gemiddelde wel. De versterking bij de eerste keuze is veel kleiner dan bij het alternatief.

Aan het einde van dit hoofdstuk wordt ook nog in het kort een mogelijke alternatieve aanpak besproken, aangeduid als de *Distance to Targets* benadering.

In Figuur 8 worden de duurzaamheidsindices voor de standaardmethode (met echte minima en maxima) en de schalingen op basis van gemiddelde en standaarddeviatie weergegeven. Het valt daarbij op dat het beeld bij de schaling op basis van de standaard deviatie zeer sterk lijkt op dat van de standaardmethode. Het enige verschil is dat de index voor A1 over de hele linie wat hoger komt te liggen.

De methode gebaseerd op het gemiddelde geeft een veel gladder verloop voor alle indicatoren. Opvallend is dat de negatieve piek voor A2 in 1994, veroorzaakt door het grote aantal asielerzoeken in dat jaar, bij deze methode nog sterker doorwerkt dan bij de standaardmethode, ten gevolge van de grote uitschieter. Verder daalt de index voor A1 in de periode 1985-1995 veel minder dan bij de standaardmethode en is de stijging van de index B1 in de periode 1985-1990 voor sterker dan bij de standaardmethode. Voor B2 valt op dat de index minder verloop kent over de hele periode; ook zijn de verschillen tussen A1 en B2 veel geprononceerder. De grafiek lijkt als het ware gekanteld rond midden jaren '80.

### III.3.2 Versterking

Een van de kritiekpunten die gegeven kan worden op de standaard methodiek, gebruikt in de DV, is dat indicatoren met een klein bereik ( $X_{max} - X_{min}$ ) 'opgeblazen' worden, terwijl indicatoren met een groot bereik 'ingekrompen' worden, terwijl er –afgezien van de weegfactoren- geen mechanisme is dat de ernst van de veranderingen onderling weegt.

In Tabel 6 wordt voor de gebruikte indicatoren geprobeerd hier een beeld van te scheppen. In feite geldt voor elk van de schalingsmethodieken dat de afgeleide indicator te schrijven is als  $Y(t)=AX(t)+B$ . In de tabel worden de factoren A en B voor de schalingsmethodieken a), c) en d) weergegeven. Merk op dat er weinig grote verschillen op te merken zijn tussen de onderlinge versterking op basis van de standaardmethode en die op basis van de standaarddeviatie, wat te verwachten is indien de data bij benadering normaal verdeeld zijn.

Het quotiënt van de A-factoren, om precies te zijn

$$\frac{\text{A – factor alternatieve methode}}{\text{A – factor standaard methode}}$$

is voor de twee alternatieven weergegeven in Figuur 9. Het is duidelijk dat de A-factor voor de methode gebaseerd op de standaard deviatie ongeveer 3 keer zo groot is als de A-factor voor de standaardmethode, zonder echt grote uitschieters. Dit effect is ook terug te zien in Figuur 8, waaruit blijkt dat de duurzaamheidsindices voor beide methoden zeer sterk op elkaar lijken. Voor de methode gebaseerd op het gemiddelde is er veel meer variatie, met als gevolg dat de duurzaamheidsindices meer afwijken.

Uit de tabel volgt dat het onderlinge verschil tussen de versterking per indicator voor alle schalingsmethoden groot is, waarbij dit verschil bij de min-max schaling en de schaling op basis van de standaarddeviatie groter is dan bij de schaling op basis van het gemiddelde. Dit is grafisch weergegeven in Figuur 10 waar de versterkingsfactoren (A-coëfficiënten) voor de min-max schaling en de schaling op basis van gemiddelden zijn uitgezet, waarbij de data genormaliseerd zijn t.o.v. hun minimale waarde. De figuur gebruikt een logaritmische y-as. Bij de methode op basis van

gemiddelden is dus minder sprake van opblaas/inkrimp-effecten. Tevens valt op dat er bij alle schalingsmethoden zeer grote verschillen zijn in versterking<sup>13</sup>.

**Tabel 6. Transformatie per indicator X voor verschillende schalingsmethoden.**

**Afgeleide indicator Y voldoet aan  $Y(t)=AX(t)+B$ . Getallen A en B x 1000**

Indicator	Als in DV		Mean		Std	
	A	B	A	B	A	B
N-gebruik in kunstmest	19	-600	16	-1000	60	-3856
Emissie broeikasgassen	73	-1874	30	-1000	269	-8939
Staatschuld (%bbp)	26	-1045	16	-1000	69	-4217
Levensverwachting (jaren)	244	19098	13	1000	835	63983
Conc pot. Cl en Br	440	-894	295	-1000	1252	-4248
Misdrijven per persoon	15495	-318	16211	-1000	45590	-2812
Hoogopgeleiden HBO/WO (%)	67	1467	71	1000	204	2889
Import energie index	5.6	-1349	3.2	-1000	22	-6940
Arb.kosten buitenl. tov NL	22	2959	8.85	1000	64	7192
Collectieve lastendruk (%bbp)	98	-3425	24	-1000	441	-18410
Werkloosheid (%bev.)	122	-115	187	-1000	397	-2122
Dekkingsgraad pensioenvoorz.	11	1552	10	1000	38	3673
Living Planet Index	2500	2525	1135	1000	7588	6685
Aantal Asielverzoeken	0.02	-1.91	0.07	-1000	0.06	-832
Misdrijven per persoon	15495	-318	16211	-1000	45590	-2812
Werkloosheid (%bev.)	122	-115	187	-1000	397	-2122
Ondervoede mensen (%wer. bev.)	77	-1014	54	-1000	241	-4500
Aantal functionerende democratieën	21	1702	20	1000	60	3004
OECD agr steun	345	-448	378	-1000	1064	-2819
Mensen < \$1 dag (%)	35	-754	29	-1000	105	-3575
Mensen ztt schoon drinkwater (%)	31	-563	32	-1000	87	-2747
Analfabeten (%wer. bev.)	61	-1256	36	-1000	197	-5543
Arbeidsinkomensquote NL	70	6593	12	1000	204	17556
Energieconsumptie	20	-1617	9.01	-1000	80	-8845
Aantal terrorisme-incidenten	3.5	-314	4.3	-1000	14	-3410

### III.3.3 Distance to Targets

Een ander alternatief voor schaling van indicatoren is gebaseerd op het (subjectief) vast stellen van normen of doelen ('Targets') en het meten van de afstand tot dit target (vergelijk het gebruik van de Millenium Development Goals). Ook bij deze methode zal geschaald moeten worden, om de ernst van de afstand te beoordelen. Deze schaling (of normering) is dus ook weer een subjectieve keuze. Voordeel van deze aanpak is echter wel dat de keuze bij voorbaat duidelijk is, en niet afhankelijk is van de waarden over een tijdsperiode, zoals bij de tot nu toe besproken schalingsmethoden. Tevens is een dergelijke aanpak goed te gebruiken voor het bepalen en

<sup>13</sup> Ook hier dient men te onderkennen dat de gebruikte maat en ijkpunt voor een indicator niet altijd waarde vrij is. Zo kan men bijvoorbeeld temperatuur zowel in graden Kelvin, Fahrenheit als Celcius weergeven. Voor de schalingsmethodiek van de DV en de methode gebaseerd op de standaarddeviatie maakt deze keuze geen verschil, voor de methode gebaseerd op het gemiddelde echter wel. Indien een indicator  $X(t)$  vervangen wordt door  $aX(t)+b$ , dan wordt de afgeleide indicator  $Y(t)$  vervangen door  $cY(t)$ , met  $c=aM/(aM+b)$ , waarbij  $M$  het gemiddelde van de indicator  $X(t)$  over de periode 1970-2000 weergeeft.

vergelijken van duurzaamheid in meerdere landen. Aan de andere kant kan men stellen dat er voor verschillende landen en wereldbeelden verschillende targets zouden kunnen gelden. Voor wereldbeelden is dit ook wat men zou verwachten. Verschillen in targets tussen landen maken een onderlinge vergelijking van de duurzaamheid juist weer twijfelachtig.

Indien een indicator 'beter' is dan de doelwaarde, dan krijgt de afgeleide indicator de waarde nul. Indien de duurzaamheidsindex op deze wijze bepaald wordt dan krijgt deze dus altijd waarden groter of gelijk aan nul, waarbij een waarde van nul aangeeft dat de duurzaamheid 'perfect' is.

Ter illustratie zijn vier indicatoren genomen, en daarbij zijn targets en schaalfactoren gekozen, zie Tabel 7. Met deze 4 indicatoren is vervolgens een duurzaamheidsindex bepaald volgens de standaardmethode uit de DV, volgens de methode gebaseerd op het gemiddelde en volgens de Distance to Targets benadering. Daarbij zijn geen weegfactoren gebruikt. Om de invloed van de schaling bij de Distance to Target methode te illustreren zijn 2 situaties doorgerekend met verschillende schaalfactoren. Deze zijn aangeduid in de tabel met A en B.

Zie Figuur 11 voor de resultaten. Om de verschillende methoden te kunnen vergelijken zijn de grafieken in deze figuur dusdanig verschoven en geschaald dat de waarden in 1970 en 2000 overeenkomen. Het blijkt dat de schaling naar verwachting ook bij deze methode een grote rol speelt.

**Tabel 7. Indicatoren gebruikt voor illustratie van Distance to Targets benadering**

Indicator	Target	Schaalfactor	
		A	B
Emissie broeikasgassen (pg, CO2 eq.)	30	5	6
Collectieve lastendruk (%bbp)	40	5	5
Werkloosheid (%bev.)	3	4	3
Ondervoede mensen (%wer. bev.)	10	10	10

## III.4 Invloed van individuele indicatoren

Zoals uit Tabel 1 blijkt, zijn per wereldbeeld slechts een beperkt aantal indicatoren van invloed. Voor elk van de wereldbeelden is bepaald wat de bijdrage is aan de duurzaamheidsindex van de individuele indicatoren en wat de gevoeligheid is van de DI voor veranderingen in deze indicatoren. Bij dit laatste is alleen gekeken naar veranderingen per jaar. Met andere woorden, er is onderzocht in hoeverre de duurzaamheidsindex in jaar  $t$  verandert indien een bijdragende indicator in dat jaar verandert, terwijl de waarden voor de overige jaren gelijk blijven. In Hoofdstuk 5 wordt nader ingegaan op veranderingen over de hele periode heen. In dit hoofdstuk wordt steeds de situatie met 23 indicatoren beschouwd. De weegfactoren zijn dusdanig aangepast dat de totale berekening wel correct verloopt.

### III.4.1 Bijdrage per indicator

Voor het bepalen van de bijdrage  $B_k(t)$  van een indicator  $X_k(t)$ , met afgeleide indicator  $Y_k(t)$ , is de volgende definitie gehanteerd,

$$B_k(t) = - \frac{w_k Y_k(t)}{\sum_k w_k}.$$

Het volgt dat, afgezien van een constante factor, de duurzaamheidsindex de som is van alle bijdragen. Het verloop van de bijdragen wordt grafisch weergegeven in Figuur 12 voor de standaard methodiek en in Figuur 13 voor de schaling op basis van het gemiddelde.

Figuur 12 bevestigt dat in de standaardmethode alle indicatoren min of meer gelijk meetellen, waarbij het onderscheid veroorzaakt wordt door de weegfactoren. Wederom valt op dat de bijdragen voor een aantal gevallen zeer sterk op elkaar lijken.

Uit Figuur 13 blijkt duidelijk dat, bij de schaling op basis van het gemiddelde, de indicatoren niet allemaal even sterk meetellen. Het opvallendst is wel de grote bijdrage voor wereldbeeld A2 van het aantal asielverzoeken.

### III.4.2 Gevoeligheid per indicator

Een ander punt van interesse is de invloed van veranderingen van de diverse indicatoren op de  $DI$ , oftewel de gevoeligheid van de duurzaamheidsindex voor kleine variaties in de indicatorwaarden over de jaren heen. Voor het bepalen van deze gevoeligheid zou het het gemakkelijkste zijn indien de  $DI$  een (bijna) lineaire functie van de diverse indicatoren zou zijn als in

$$DI(t) = \beta_0(t) + \beta_1(t)X_1(t) + \dots + \beta_N(t)X_N(t),$$

omdat dan bij de gevoeligheidsanalyse van eenvoudige lineaire regressietechnieken gebruik gemaakt kan worden.

Voor de in de DV gebruikte methode is dit voor veel jaren ook het geval, behalve voor de jaren waarbij voor een van de indicatoren geldt dat  $X_{min}=X(t)$  of  $X_{max}=X(t)$ , omdat dan schalings-

effecten expliciet een rol spelen, via het feit dat  $Y(t) = (X - X_{min}) / (X_{max} - X_{min})$ . Ook voor de alternatieve schalingsmethoden is geen sprake meer van een exacte lineaire relatie, omdat de schalingsfactor afhankelijk is van de indicatorwaarden. Toch wordt geprobeerd of een redelijke inschatting gemaakt kan worden van de bijdragen van de diverse indicatoren door de samengestelde index te benaderen met een lineair model,

$$DI(t) = \beta_0(t) + \beta_1(t)X_1(t) + \dots + \beta_N(t)X_N(t) + E(t),$$

waarbij  $E(t)$  de fout in de approximatie weergeeft.

Voor de praktische uitvoering is daartoe gebruikt gemaakt van een eenvoudige regressieanalyse, waarbij voor elk jaar  $t$  de volgende procedure gevolgd is<sup>14</sup>:

- Ken aan elke indicator een onzekerheidsband van 10% toe<sup>15</sup>. Dit leidt tot 23 parameters<sup>16</sup> met een onzekerheidsband.
- Trek met behulp van Latin Hypercube Sampling (LHS) (Saltelli *et al.*, 2002) 1000 samples uit een uniforme verdeling.
- Bepaal voor elk van de trekkingen de  $DI(t)$ , door de waarden in het desbetreffende jaar  $t$  te vervangen door de gesampled waarden.
- Voer een lineaire regressieanalyse uit om de waarden van de gevoeligheidscoëfficiënten  $\beta_k(t)$  te bepalen

<sup>14</sup> Voor een aantal gevallen kan  $\beta_k$  ook theoretisch bepaald worden.

A. Voor de schalingsmethoden a), b) en e) geldt –onder de voorwaarde dat  $X_{k,min} < X_k(t) < X_{k,max}$  – dat

$$DI(t) = 100 - 100 \frac{\sum_k w_k \frac{X_k(t) - X_{k,min}}{X_{k,max} - X_{k,min}}}{\sum_k w_k} \quad \text{en dus} \quad \beta_k = \frac{-100}{(X_{k,max} - X_{k,min})} \frac{w_k}{\sum_k w_k}$$

B. Voor de duurzaamheidsindex c) gebaseerd op het gemiddelde geldt (onder de veronderstelling dat het gemiddelde niet van teken verandert) de volgende eerste orde benadering:

$$DI(t) = \frac{\sum_k w_k \frac{X_k(t) - M_k}{\pm M_k}}{\sum_k w_k} \approx -1 + \frac{\sum_k w_k \left[ \frac{TM_k - X_k(t)}{\pm TM_k^2} \right] X_k(t)}{\sum_k w_k}, \quad \beta_k = \left[ \frac{TM_k - X_k(t)}{\pm TM_k^2} \right] \frac{w_k}{\sum_k w_k}.$$

C. Voor de duurzaamheidsindex d) gebaseerd op de standaarddeviatie kan afgeleid worden dat

$$DI(t) = \frac{\sum_k w_k \frac{X_k(t) - M_k}{S_k}}{\sum_k w_k} \approx \frac{\sum_k w_k \left[ \frac{T-1}{TS_k} \right] X_k(t)}{\sum_k w_k} \quad \text{en dus} \quad \beta_k = \left[ \frac{T-1}{TS_k} \right] \frac{w_k}{\sum_k w_k}.$$

<sup>15</sup> De grootte van het percentage afwijking is hierbij van redelijk grote invloed. Hoe kleiner het percentage, des te beter zal een lineair model de constructie kunnen weergeven. Hier is voor 10% procent gekozen omdat dit –gemiddeld gezien– een zeer redelijke veronderstelling lijkt over de onzekerheid in de data.

<sup>16</sup> Om singulariteitsproblemen te vermijden is het hierbij noodzakelijk dat de vraagstukken *Werkloosheid* en *Inkomenszekerheid* tot één vraagstuk gecombineerd worden, hierna aan te duiden als *Werkloosheid*. Hetzelfde geldt voor de vraagstukken *Criminaliteit* en *Naleving van regels*, hierna aangeduid als *Criminaliteit*.

In Figuur 14 worden voor drie schalingsmethoden de resulterende approximaties naast de werkelijke duurzaamheidsindex uitgezet. De approximaties zijn weergegeven middels stippellijnen. Het is evident dat in alle gevallen de benaderingsfout klein is. De afwijkingen zijn voor de schaling op basis van het gemiddelde met het blote oog zelfs niet waar te nemen.

Voor het beoordelen van de gevoeligheid van de duurzaamheidsindex zijn vervolgens de grootheden  $\beta_i(t)X_i(t)$  beschouwd. Immers, indien de variabele  $X_i(t)$  met  $P$  procent verandert, dan leidt dit tot een absolute toename van  $Y(t)$  met  $0.01 * P * \beta_i(t)X_i(t)$ .

In Tabel 8 wordt een idee gegeven van de gemiddelde procentuele (over de jaren heen) bijdrage per indicator aan het dynamische deel van de duurzaamheidsindex is. Deze grootheid<sup>17</sup>

$$B_i = \sum_{t=1970}^{2000} \frac{|\beta_i(t)X_i(t)|}{|\beta_1(t)X_1(t)| + \dots + |\beta_N(t)X_N(t)|} \times 100$$

blijkt relatief weinig onderlinge verschillen te vertonen voor de methode gebaseerd op het gemiddelde; de verschillen zijn dusdanig dat van geen indicator gesteld kan worden dat deze overheerst, noch dat deze verwaarloosd kan worden.

Er blijken echter wel zeer grote verschillen op te treden voor zowel de standaard methodiek als de methode gebaseerd op de standaard deviatie. De indicator *Levensverwachting in jaren* heeft een onverwacht groot aandeel van 35% en 50% voor de wereldbeelden A1 en A2

Voor wereldbeeld B1 springen met name de *Emissie van Broeikasgassen* en de *Arbeidsinkomensquote* er uit, voor B2 zijn de *Emissie van Broeikasgassen* en de *Living Planet Index* het meest belangrijke, maar zijn *Stikstofgebruik in kunstmest*, het *%ondervoede mensen* en de *Concentratie atmosferisch chloor en broom* ook nog aanzienlijk.

Tabel 8. Gemiddelde procentuele bijdrage per jaar, per indicator

Indicator	Als in DV				Mean				Std			
	A1	A2	B1	B2	A1	A2	B1	B2	A1	A2	B1	B2
N-gebruik in kunstmest	7	7	8	15	13	13	11	18	5	5	7	14
Emissie broeikasgassen	12	12	15	25	11	11	11	15	10	10	16	27
Staatsschuld (%bbp)	7	0	0	0	11	0	0	0	4	0	0	0
Levensverwachting (jaren)	35	36	0	0	10	10	0	0	49	48	0	0
Conc pot. Cl en Br	5	6	6	13	9	10	8	13	4	4	6	12
Misdrijven per persoon	4	6	0	0	9	13	0	0	3	4	0	0
Hoogopgeleiden HBO/WO (%)	4	0	0	0	9	0	0	0	2	0	0	0
Import energie index	6	6	0	0	8	7	0	0	5	5	0	0
Arb.kosten buitenl. tov NL	7	0	0	0	7	0	0	0	5	0	0	0
Collectieve lastendruk (%bbp)	12	12	0	0	7	7	0	0	12	12	0	0
Werkloosheid (%bev.)	2	2	0	0	6	6	0	0	1	1	0	0
Dekkingsgraad pensioenvoorz.	0	4	0	0	0	9	0	0	0	3	0	0
Living Planet Index	0	8	9	17	0	9	8	13	0	6	9	17
Aantal Asielverzoeken	0	1	0	0	0	5	0	0	0	1	0	0
Ondervoede mensen (%wer. bev.)	0	0	8	12	0	0	10	12	0	0	8	11
Aantal functionerende democratieën	0	0	5	7	0	0	9	9	0	0	5	6
OECD agr steun	0	0	5	7	0	0	8	11	0	0	4	6
Mensen < \$1 dag (%)	0	0	6	0	0	0	8	0	0	0	5	0
Mensen ztt schoon drinkwater (%)	0	0	4	0	0	0	7	0	0	0	4	0
Analfabeten (% wer. bev.)	0	0	7	0	0	0	7	0	0	0	7	0
Arbeidsinkomensquote NL	0	0	19	0	0	0	7	0	0	0	20	0
Energieconsumptie	0	0	8	0	0	0	6	0	0	0	9	0
Aantal terrorisme-incidenten	0	0	0	5	0	0	0	9	0	0	0	6

<sup>17</sup> Omdat de grootheid  $\beta_k(t)X_k(t)$  zowel positieve als negatieve waarden kan aannemen, is hier gekozen voor een maat waarbij naar de absolute waarden gekeken wordt. De hier gedefinieerde variabele  $B_i$  varieert tussen 0 en 100.

De reden van de grote invloed van de *Levensverwachting* ligt in de constructie van de duurzaamheidsindex. Uit voetnoot 14 volgt de waarde van  $\beta_k(t)$  voor de standaardmethode (onder de veronderstelling dat de indicatorwaarden  $X_{k,min} < X_k(t) < X_{k,max}$ ). Het volgt dat de grootte  $\beta_k(t) X_k(t)$  groot zal zijn voor indicatoren met een klein bereik en grote nominale waarde. Er zijn wel degelijk indicatoren met een kleiner bereik dan *Levensverwachting*, bijvoorbeeld het *Aantal misdrijven per persoon*, maar deze laatste heeft een veel lagere nominale waarde.

Bij de methode gebaseerd op het gemiddelde volgt (even aangenomen dat het gemiddelde positief is) uit de eerste orde benadering dat

$$\beta_k(t) X_k(t) = \left[ \frac{TM_k X_k(t) - X_k^2(t)}{TM_k^2} \right] \frac{w_k}{\sum w_k} = \left[ \frac{X_k(t)}{M_k} - \frac{X_k^2(t)}{TM_k^2} \right] \frac{w_k}{\sum w_k}$$

Zij  $a, b$  dusdanig dat  $X_{min} = aM_k$  en  $X_{max} = bM_k$ , dan varieert<sup>18</sup> deze grootte van  $a \left[ 1 - \frac{a}{T} \right] \frac{w_k}{\sum w_k}$  tot  $b \left[ 1 - \frac{b}{T} \right] \frac{w_k}{\sum w_k}$ . Indien  $a$  en  $b$  dicht bij elkaar liggen (klein bereik), zal de grootte dus weinig variëren. Bij een groot bereik (relatief t.o.v. het gemiddelde  $M_k$ ) zal de bijdrage juist wel variëren.

In Figuur 15 wordt voor de indicatoren *Stikstofgebruik in kunstmest*, *Levensverwachting* en *Aantal misdrijven per persoon* het verloop van de grootte  $\beta_k(t) X_k(t)$  weergegeven voor de standaardmethode en de methode gebaseerd op het gemiddelde. Het is bij de standaardmethode duidelijk waar te nemen dat aan de uiteinden de grootte voor *Levensverwachting* beïnvloed wordt doordat daar het minimum, respectievelijk het maximum wordt aangenomen. De indicatoren *Levensverwachting* en *Aantal misdrijven per persoon* hebben beiden een klein bereik, maar bij de standaardmethode is de gevoeligheid voor de levensverwachting veel groter door de hogere nominale waarde. Voor de methode gebaseerd op het gemiddelde is waar te nemen dat de indicator *Aantal misdrijven per persoon* veel meer varieert. Dit komt omdat deze indicator een grote variatie kent ten opzichte van het gemiddelde. De onderlinge verschillen in de gevoeligheden van de drie getoonde indicatoren zijn bij de laatste methode wel een stuk kleiner dan bij de standaardmethode.

<sup>18</sup> Deze eigenschap geldt alleen indien  $T/2$  niet in het interval  $[a, b]$  ligt, hetgeen voor alle indicatoren van toepassing is.



### III.5 Onzekerheidsanalyse

Dit laatste hoofdstuk is gewijd aan een onzekerheidsanalyse van de berekening van de duurzaamheidsindex. De analyse is beperkt tot twee schalingsmethoden, te weten de standaardmethodiek van de DV, maar dan wel gebaseerd op echte ranges, en de methode gebaseerd op het gemiddelde. Het basisdoel is om een beeld te scheppen van de variatie die optreedt in de duurzaamheidsindex indien geaccepteerd wordt dat de onderliggende data behept zijn met onzekerheid. Daarbij worden zowel variaties in de waarden van de indicatoren als in de weegfactoren beschouwd. De vraag die hiermee beantwoord moet worden is of de gepresenteerde duurzaamheidsindex een betrouwbaar beeld oplevert, indien wordt uitgegaan van onzekerheden en of het resultaat van die vraag anders uitvalt bij een alternatieve constructiemethodiek. Voor dit alternatief is gekozen voor de methode gebaseerd op het gemiddelde, omdat deze het meest afwijkt van de standaardmethode.

Er is voor dit doel gebruik gemaakt van een Monte-Carlo analyse, waarbij door middel van Latin-Hypercube sampling aan elk van de variabelen een onzekerheidsband is toegekend waaruit samples getrokken zijn op basis van een uniforme verdeling. Hierbij is aan elk van de weegfactoren een onzekerheid toegekend op basis van een onzekerheidsanalyse, uitgevoerd door Hans Visser (Visser *et al.*, 2005; Aalbers *et al.*, 2006). Daarbij worden twee vormen van onzekerheid beschouwd en uitgedrukt in de vorm van standaarddeviaties. Het betreft de onzekerheid t.g.v. van variabiliteit bij de respondenten ( $\sigma_1 = S_{resp}$ ), en de onzekerheid t.g.v. de scoringsmethodiek

( $\sigma_2 = S_{meth}$ ). De totale standaarddeviatie  $\sigma_t = S_{tot}$  is vervolgens bepaald d.m.v.  $\sigma_t = \sqrt{\sigma_1^2 + \sigma_2^2}$ .

Deze analyse is in eerste instantie uitgevoerd zonder onderscheid te maken naar wereldbeeld (Visser *et al.*, 2005). In een later stadium is tevens onderzocht wat de effecten zijn indien er wel onderscheid gemaakt wordt naar wereldbeelden (Aalbers *et al.*, 2006). De resulterende standaarddeviaties per vraag en per wereldbeeld zijn vermeld in Tabel 9. Voor elke weegfactor  $w$  is uniform gesampled uit het interval  $[w-2\sigma, w+2\sigma]$ , waarbij de afzonderlijke weegfactoren onafhankelijk verondersteld zijn.

Aan de indicatorwaarden is een naar het verleden toenemende relatieve onzekerheid toegekend, lineair oplopend van  $\pm A\%$  in 2000 tot  $\pm B\%$  in 1970. Per indicator zijn tentatief verschillende waarden voor A en B gekozen, omdat het niet realistisch is om aan elke indicator dezelfde mate van onzekerheid toe te kennen. Zie Tabel 10 voor de waarden van A en B per gebruikte indicator<sup>19</sup>. Het resulterende beeld verandert niet indien andere waarden gekozen zouden worden<sup>20</sup>.

De onzekerheden in de indicatoren zijn onderling, d.w.z. wat betreft indicator X en indicator Z, onafhankelijk verondersteld. Echter, omdat de waarden van een afzonderlijke indicator van jaar tot jaar natuurlijk niet als onafhankelijk te beschouwen zijn, is er uitgegaan van een correlatie van 90% tussen de waarden  $X(t)$  en  $X(t+1)$  in opeenvolgende jaren.

Er is gebruik gemaakt van 25 indicatoren, 31 jaren en per indicator vier verschillende weegfactoren. Dit leidt tot een totaal van  $25 \times (31+4) = 875$  variabelen. Er zijn 10 keer zoveel, dus 8750, samples getrokken. Alle trekkingen worden uit uniforme verdelingen gedaan, omdat er geen goede onderbouwing te geven was om uit andere verdelingen te trekken.

<sup>19</sup> Deze (tentatieve) waarden zijn opgesteld in overleg met D. Nagelhout (MNP).

<sup>20</sup> Alleen indien verondersteld wordt dat de waarden voor een indicator over alle jaren heen met eenzelfde factor vermenigvuldigd worden (correlatie van 1), dan zal bij de standaardmethode geen onzekerheid bestaan in de duurzaamheidsindex. Dit is echter een zeer onrealistische veronderstelling.

Om de resultaten van de twee schalingsmethoden op een eerlijke wijze te kunnen vergelijken zijn de indices die met de methode op basis van het gemiddelde zijn bepaald dusdanig geschaald en verschoven dat de eindwaarden (1970 en 2000) van de (zonder onzekerheid bepaalde) index voor wereldbeeld A1 voor beide schalingsmethoden samenvallen.

Van het resulterende sample van indices zijn het 5 en het 95 percentiel bepaald. Deze zijn weergegeven in Figuur 16. Tevens zijn daar de maximale en minimale waarden weergegeven. Het is duidelijk waar te nemen dat de methode gebaseerd op het gemiddelde veel minder gevoelig is voor de onzekerheden dan de standaardmethode.

Om de vraag te kunnen beantwoorden of er nog wel voldoende onderscheid tussen de diverse wereldbeelden is terug te vinden binnen de resulterende onzekerheid, is tevens het verschil tussen de indices van een aantal wereldbeelden bepaald in deze analyse. In Figuur 17 is de resulterende variatie in de verschillen tussen wereldbeelden A1-A2, A1-B2 en B1-B2 uitgezet.

Het blijkt dat er wel zeker sprake is van onderscheidend vermogen tussen wereldbeelden, bij beide schalingsmethoden.

Deze resultaten zijn van belang indien er in toekomstige analyses gebruik gemaakt zou worden van prognoses, omdat in dat geval de onderliggende scenariogegevens met wellicht grotere onzekerheden dan hier toegepast, gebruikt zullen worden. De analyse leidt tot de conclusie dat de methode gebaseerd op gemiddelden beter geschikt is indien onzekerheden in beschouwing worden genomen.

**Tabel 9. Standaarddeviaties voor weegfactoren bij vraagstukken.**

Bepaald op basis van variabiliteit in respondenten en scoringsmethodiek, zie Visser *et al.* (2005) en Aalbers *et al.* (2006). De in rood gedrukte waarden zijn niet gebruikt bij de onzekerheidsanalyse.

Vraagstuk uit Enquêtes	Stot				
	NL	A1	A2	B1	B2
9 Waterverontreiniging	0.25	0.25	0.21	0.26	0.26
5 Broeikaseffect	0.33	0.30	0.32	0.38	0.28
19 Staatsschuld	0.14	0.14	0.15	0.12	0.12
37 Gezondheidszorg	0.18	0.30	0.16	0.16	0.14
16 Ozonlaag	0.18	0.18	0.15	0.17	0.15
49 Criminaliteit	0.17	0.15	0.13	0.18	0.15
29 Onderwijs	0.18	0.18	0.14	0.22	0.18
26 Energievoorzieningszekerheid	0.26	0.26	0.27	0.28	0.26
18 Concurrentiepositie	0.13	0.13	0.11	0.09	0.11
20 Collectieve Lastendruk	0.17	0.19	0.15	0.17	0.14
36 Werkloosheid	0.15	0.19	0.14	0.16	0.14
44 Pensioenvoorziening	0.16	0.19	0.15	0.15	0.13
8 Biodiversiteit	0.15	0.15	0.10	0.19	0.11
32 Culturele Diversiteit	0.18	0.13	0.14	0.20	0.10
50 Naleving van regels	0.17	0.22	0.15	0.18	0.16
25 Inkomenszekerheid	0.11	0.12	0.10	0.09	0.09
30 Honger	0.22	0.20	0.15	0.36	0.22
46 Mensenrechten	0.15	0.13	0.11	0.34	0.14
17 Internationale Samenwerking	0.27	0.13	0.18	0.46	0.31
48 Armoede	0.14	0.14	0.11	0.22	0.13
15 Drinkwaterbeschikbaarheid	0.11	0.16	0.06	0.16	0.08
42 Analfabetisme	0.13	0.10	0.10	0.21	0.14
28 Maatschappelijke Rol Grote bedrijven	0.24	0.09	0.17	0.40	0.27
27 Olie en Gasvoorraden	0.26	0.11	0.20	0.34	0.27
51 Gewapende conflicten en Terrorisme	0.25	0.14	0.14	0.27	0.31

**Tabel 10. Relatieve onzekerheidsgrenzen , gebruikt voor Monte Carlo analyse (in %).****Data gegeven voor 1970 en 2000. Voor tussenliggende jaren is lineair geïnterpoleerd**

<b>Indicator</b>	<b>1970</b>	<b>2000</b>
N-gebruik in kunstmest	6	3
Emissie broeikasgassen	10	5
Staatsschuld (%bbp)	1	1
Levensverwachting (jaren)	2	1
Conc pot. Cl en Br	15	3
Misdrijven per persoon	1	1
Hoogopgeleiden HBO/WO (%)	1	1
Import energie index	1	1
Arb.kosten buitenl. tov NL	4	2
Collectieve lastendruk (%bbp)	3	1
Werkloosheid (%bev.)	1	1
Dekkingsgraad pensioenvorz.	1	1
Living Planet Index	20	5
Aantal Asielverzoeken	1	1
Ondervoede mensen (% wer. bev.)	20	5
Aantal functionerende democratieën	2	1
OECD agr steun	5	2
Mensen < \$1 dag (%)	10	5
Mensen ztt schoon drinkwater (%)	20	5
Analfabeten (%wer. bev.)	20	5
Arbeidsinkomensquote NL	1	1
Energieconsumptie	8	4
Aantal terrorisme-incidenten	1	1



## III.6 Conclusies

In dit hoofdstuk worden in het kort de belangrijkste conclusies van de voorgaande gevoeligheids- en onzekerheidsanalyses besproken:

- De indexconstructie in de DV is - via selectie van indicatoren en weging - met name sterk op de enquêteresultaten gebaseerd (vanuit de maatschappij; wereldbeelden), en niet zozeer op een achterliggend systeemraamwerk (vanuit de wetenschap; bijv. via DPSIR-raamwerk).
- De indexconstructie komt via het gewogen sommeren van afzonderlijke herschaalde indicatoren tot stand, d.w.z. betreft een lineaire additieve aanpak. Hierbij blijven de (vaak niet-lineaire) invloeden van synergie, koppelingen, versterkingen en verzwakkingen buiten beschouwing, terwijl de sommatie bovendien suggereert dat er eenvoudige compensatie/uitruil kan plaatsvinden tussen de indicatoren (hoge waarden van de ene indicator kunnen door lage waarden van anderen gecompenseerd worden). Dit alles kan het zicht ontnemen op mogelijke relevante conflicten en cruciale keuzes rond duurzaamheidsvraagstukken.
- De index is o.a. gevoelig voor:
  - **Indicatorsselectie**: de keuze van aantal (Figuur 6 en Figuur 7) en soort (proxy-karakter; Figuur 4) indicatoren per wereldbeeld, en per domein kan het verloop van de index in de diverse wereldbeelden sterk bepalen, met name op de afzonderlijke duurzaamheidsdomeinen. In de praktijk zijn databeschikbaarheid en datakwaliteit vaak sterk bepalend voor de uiteindelijke keuze van de proxy (Figuur 1-Figuur 3).
  - **Schaling** van indicatoren: de huidige schaling is sterk gekoppeld aan een specifiek tijdframe (1970-2000) en de hierdoor veroorzaakte afzwakkings- of versterkingseffecten kunnen groot zijn (Tabel 6). Alternatieven verdienen aandacht en kunnen ook een duidelijk ander beeld geven (Figuur 8 en Figuur 11) bijv. Distance-to-targets; Schaling op basis van gemiddelde, of het gebruik van een meer gedifferentieerde indexering à la Telos, die specifiek gericht is op streefwaarden/wensbeeld (d.w.z. een beetje Distance-to-targets achtig; zie Lemmens et al. 2003).
  - **Weging** van indicatoren: het lijkt alsof de index niet al te gevoelig is voor de specifieke weging die gehanteerd wordt (vergelijk bijv. Figuur 5); mogelijke oorzaak hiervan is de 'redundantie' (vergelijkbare dynamiek) van afzonderlijke indicatoren.
- Totaal-index of domein-index?: De domein-index verdient de voorkeur (Figuur 6 en Figuur 7); de totaal-index verdoezelt voor een deel het verschillende gedrag op de afzonderlijke domeinen, ten gevolge van compensatie/aggregatie, en suggereert bovendien dat eenvoudige uitruil *tussen* domeinen duurzaamheid zou kunnen verbeteren.
- Bijdragen van individuele indicatoren aan index: Ook dit kan sterk afhangen van indexconstructie, vergelijk Figuur 12 en Figuur 13 met respectievelijk schaling volgens standaardmethode DV (min-max schaling) en schaling op gemiddelden.

- Onzekerheid van index: Deze kan groot zijn (per wereldbeeld), met name voor standaardmethode (Figuur 16), maar de verschillen tussen de wereldbeelden blijven ondanks onzekerheden significant zichtbaar (Figuur 17).
- Keuze van proxy's: Centrale vraag blijft of de gekozen proxy's een goed beeld geven van de maatschappelijke vraagstukken (Tabel 1). Hierbij speelt mee of er alternatieve proxy's beschikbaar zijn waarvoor voldoende (adequate) data te verkrijgen is, en of men zich vanuit een specifiek wereldbeeld kan vinden in de keuze van de proxy. Daarnaast kan men zich de vraag stellen of het verantwoord is om meerdere vraagstukken met dezelfde proxy te meten? Dit heeft er in de DV bijv. toe geleid dat voor wereldbeeld A2 de indicator *het aantal bij de politie bekende misdrijven per persoon* als belangrijkste indicator werd beoordeeld, zonder dat de bijbehorende maatschappelijke vraagstukken als zeer belangrijk werden gescoord. In dit licht verdienen deze aspecten extra aandacht als men in de toekomst een vergelijkbare exercitie wil uitvoeren.
- Uitvoering. Indien gekozen wordt voor een bepaalde schalingsmethodiek bij de indexconstructie, dient ervoor gezorgd te worden dat de methode ook consistent juist wordt toegepast. In de DV zijn niet altijd de 'echte' minimale en maximale waarden van de indicatoren gebruikt, hetgeen kan leiden tot onder/over-waardering van het corresponderende vraagstuk. (Zie de groene lijnen in Figuur 1-Figuur 3 en voetnoot 2)
- Correlatie en Redundantie van indicatordata. Bij het bepalen van een samengestelde index op basis van indicatoren dient een duidelijke keuze gemaakt en gemotiveerd te worden over hoe er met samenhang en redundantie in de data wordt omgegaan: neemt men deze voor lief, of bepaalt men een (deel)selectie van indicatoren die qua dynamiek ten opzichte van elkaar duidelijk verschillen. (Paragraaf III.2.6)

### III.7 Referenties

Aalbers, T.G., Vringer K., Visser, H.(2006) Achtergronden Enquête. Bijlage I bij: Petersen, A. *et al.* (2006), *Methoderapport Duurzaamheidsverkenning*. MNP Rapport 550031001, Milieu- en Natuurplanbureau, Bilthoven.

Elliott M.L. 1981. Pulling the pieces together: amalgamation in environmental impact assessment. *Environmental Impact Assessment Review* 2(1): 11-38.

Hueting, R. and L. Reijnders, (2004). Broad sustainability contra sustainability: the proper construction of sustainability indicators. *Ecological Economics* 50(4): 249-260

Haarmann, W., Hermans, F., Overeem, I., (2004). *Monitoring van provinciale duurzame ontwikkeling. De duurzaamheidbalans getoetst in vier provincies*. Telos, <http://spitswww.uvt.nl/telos/documents/rapportmonitoringdzhb.pdf>

Lemmens, L., Hermans, F., Tabibian, N., Westerlaken, M., ten Caten, H. (2003). *De Duurzaamheidbalans van Noord-Brabant 2001: De Verantwoording*. Telos <http://spitswww.uvt.nl/telos/documents/duurzaam.pdf>

Nagelhout, D. (2006) Indicatoren en Duurzaamheidsindex, Bijlage II bij: Petersen, A. *et al.* (2006), *Methoderapport Duurzaamheidsverkenning*. MNP Rapport 550031001, Milieu- en Natuurplanbureau, Bilthoven.

Nardo, M., Saisana, M., Saltelli A., Tarantola, S. (2005) *Tools for Composite Indicators Building*. European Commission. Directorate General. Joint Research Center. ISPRA. EUR 21682 EN <http://farmweb.jrc.cec.eu.int/ci/Document/EUR%2021682%20EN.pdf>

MNP-RIVM (2004), *Kwaliteit en Toekomst, verkenning van duurzaamheid*. RIVM rapport 500013009, RIVM, Bilthoven.

Saisana, M. and Tarantola, S. (2002) *State-of-the-art report on current methodologies and practices for composite indicator development*, EUR 20408 EN, European Commission-JRC: Italy.

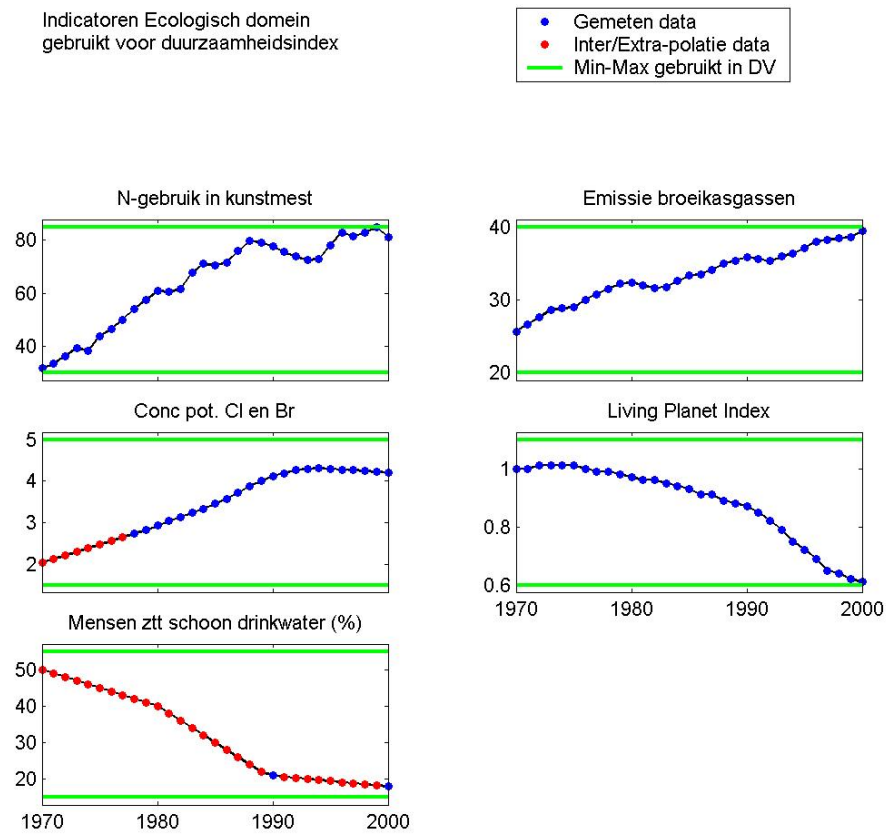
Saisana M., Tarantola S., Saltelli A. (2005) Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators, *Journal of the Royal Statistical Society A*, 168(2): 1-17.

Saltelli, A., Chan, K. and Scott, M. (2000) *Sensitivity analysis, Probability and Statistics series*, New York: John Wiley & Sons.

Visser, H., Aalbers, T.G., Vringer, K. (2005), *Maatschappelijke prioriteiten in Nederland. Statistische analyse van de NIPO-Veldkamp-enquêtes 2003/2005*. MNP Rapport 771404001, Milieu- en Natuurplanbureau, Bilthoven.

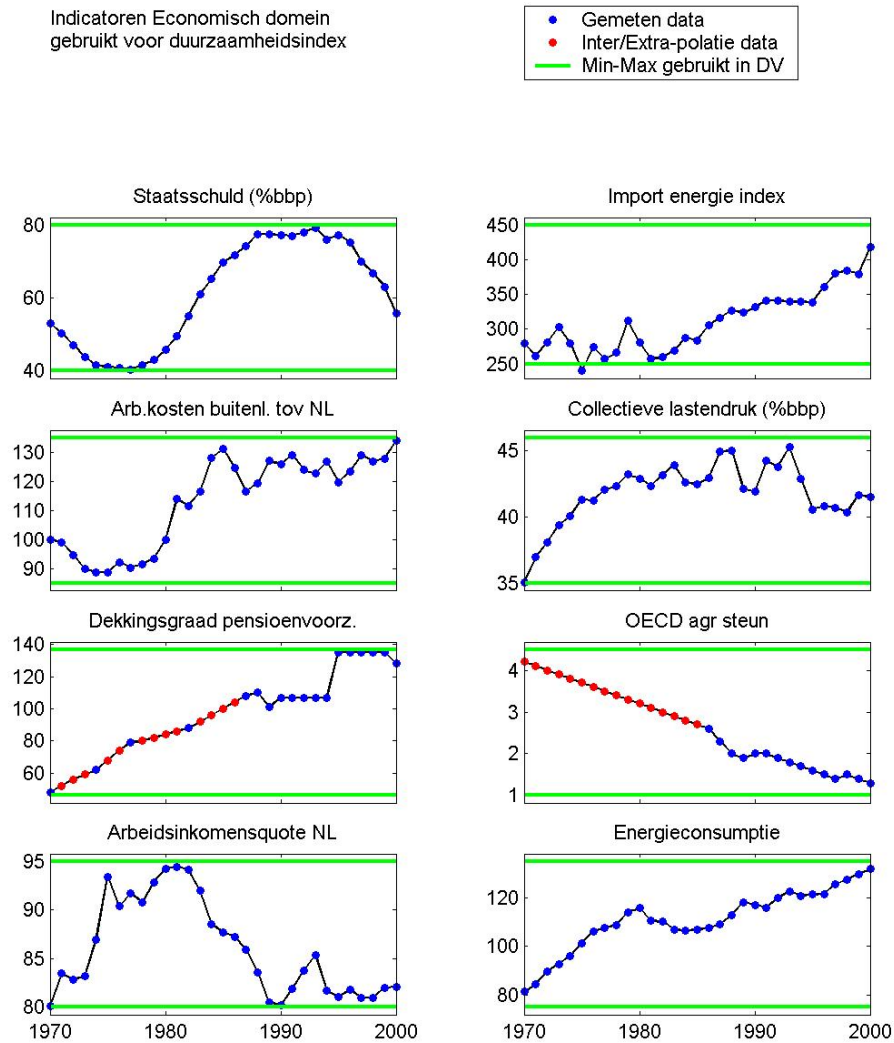






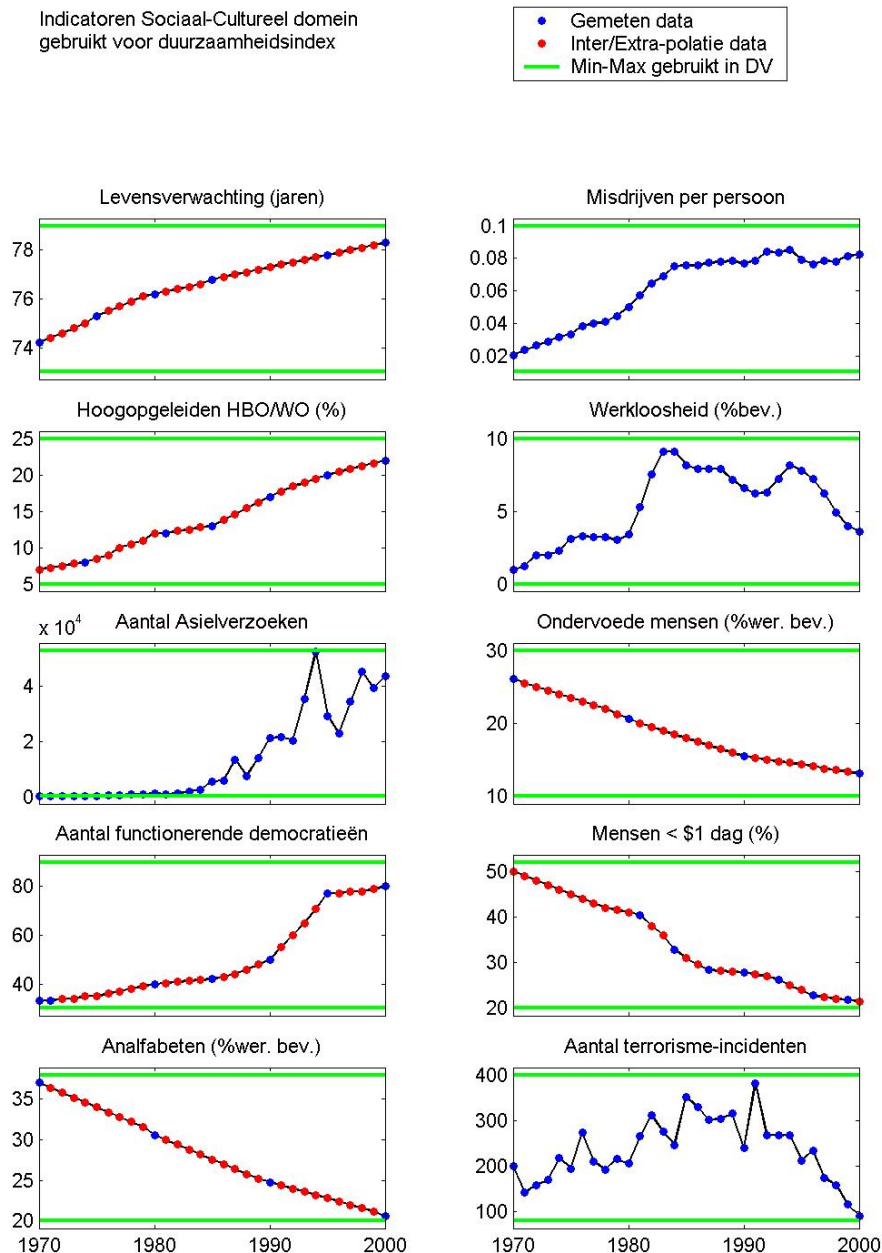
**Figuur 1. Indicatoren uit het Ecologisch Domein, gebruikt voor duurzaamheidsindex.**

**In blauw worden gemeten data weergegeven, in rood de data verkregen d.m.v. interpolatie en extrapolatie. De groene lijnen representeren de minimale en maximale waarden, zoals die in de DV gehanteerd zijn.**



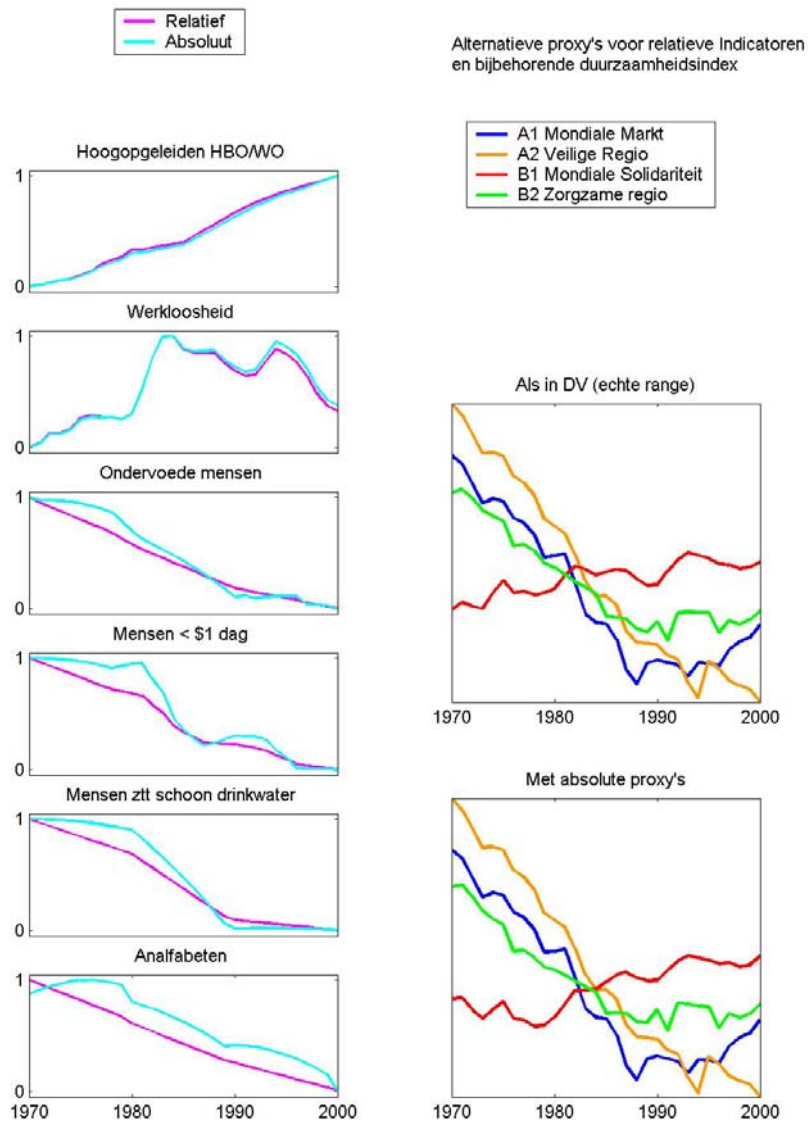
**Figuur 2. Indicatoren uit het Economisch Domein, gebruikt voor duurzaamheidsindex.**

**In blauw worden gemeten data weergegeven, in rood de data verkregen d.m.v. interpolatie en extrapolatie. De groene lijnen representeren de minimale en maximale waarden, zoals die in de DV gehanteerd zijn.**



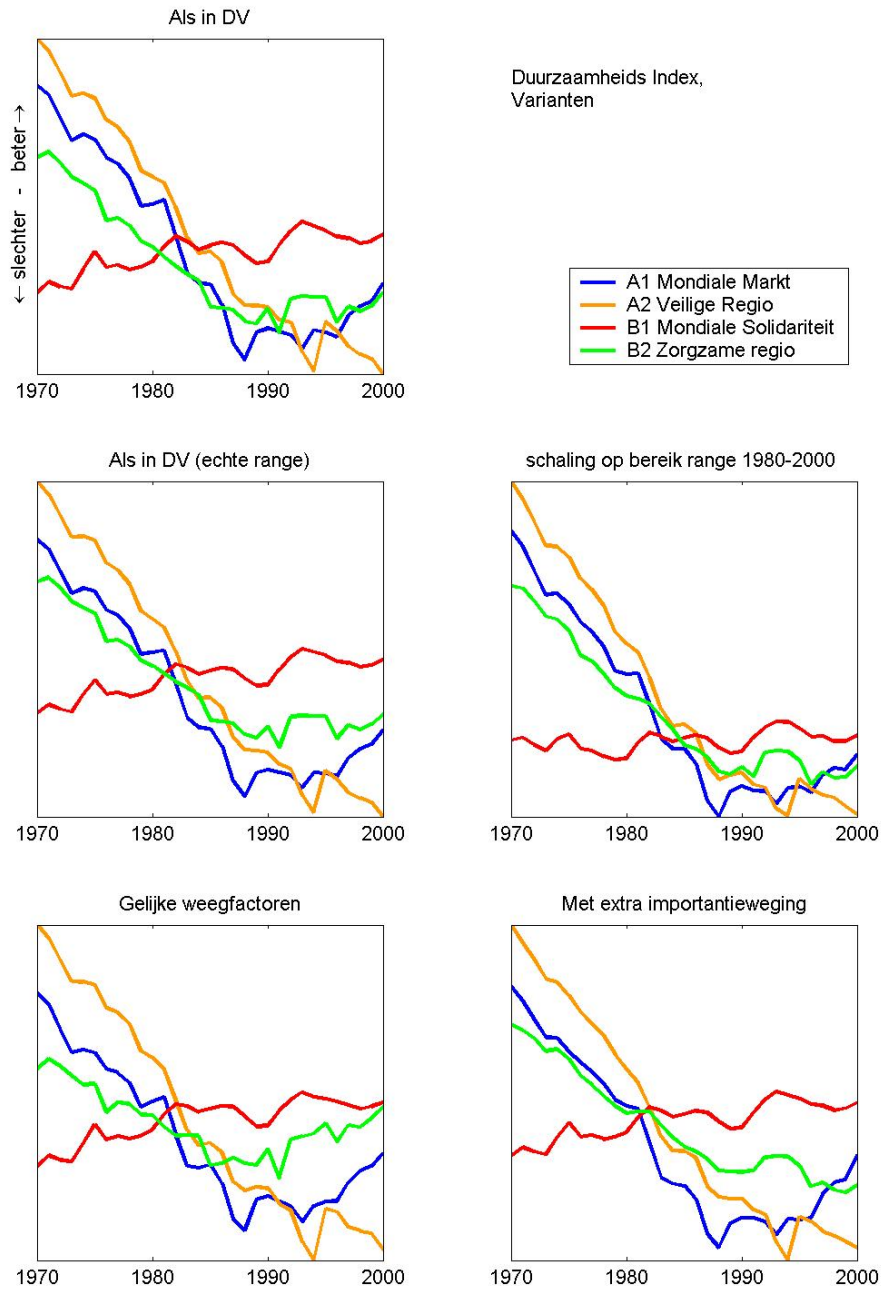
**Figuur 3. Indicatoren uit het Sociaal-Culturele Domein, gebruikt voor duurzaamheidsindex.**

**In blauw worden gemeten data weergegeven, in rood de data verkregen d.m.v. interpolatie en extrapolatie. De groene lijnen representeren de minimale en maximale waarden, zoals die in de DV gehanteerd zijn.**

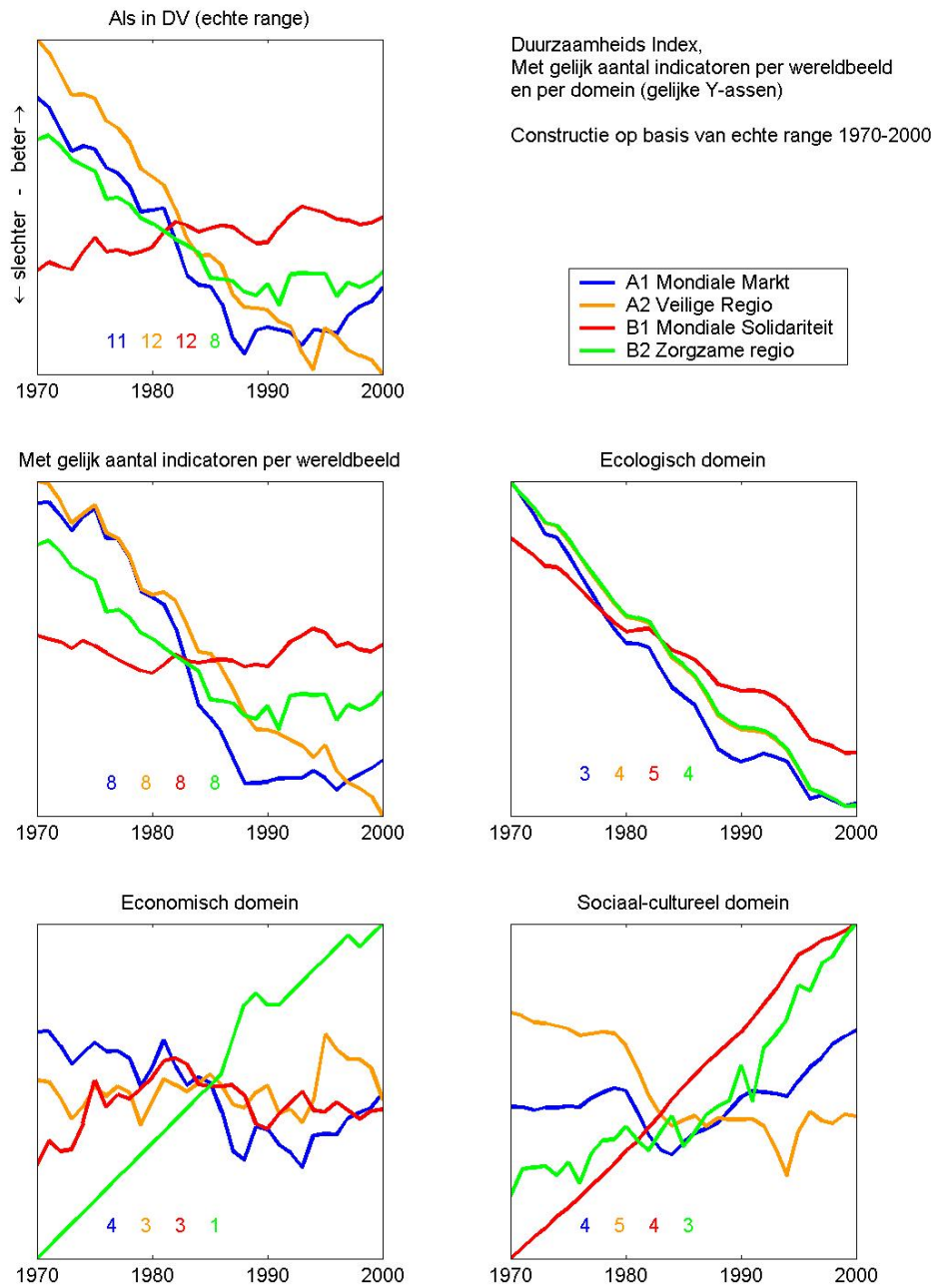


**Figuur 4. Verschillen in bevolkingsgerelateerde indicatoren en de corresponderende duurzaamheidsindex.**

**Verschillen ten gevolge van formulering in relatieve of absolute termen**

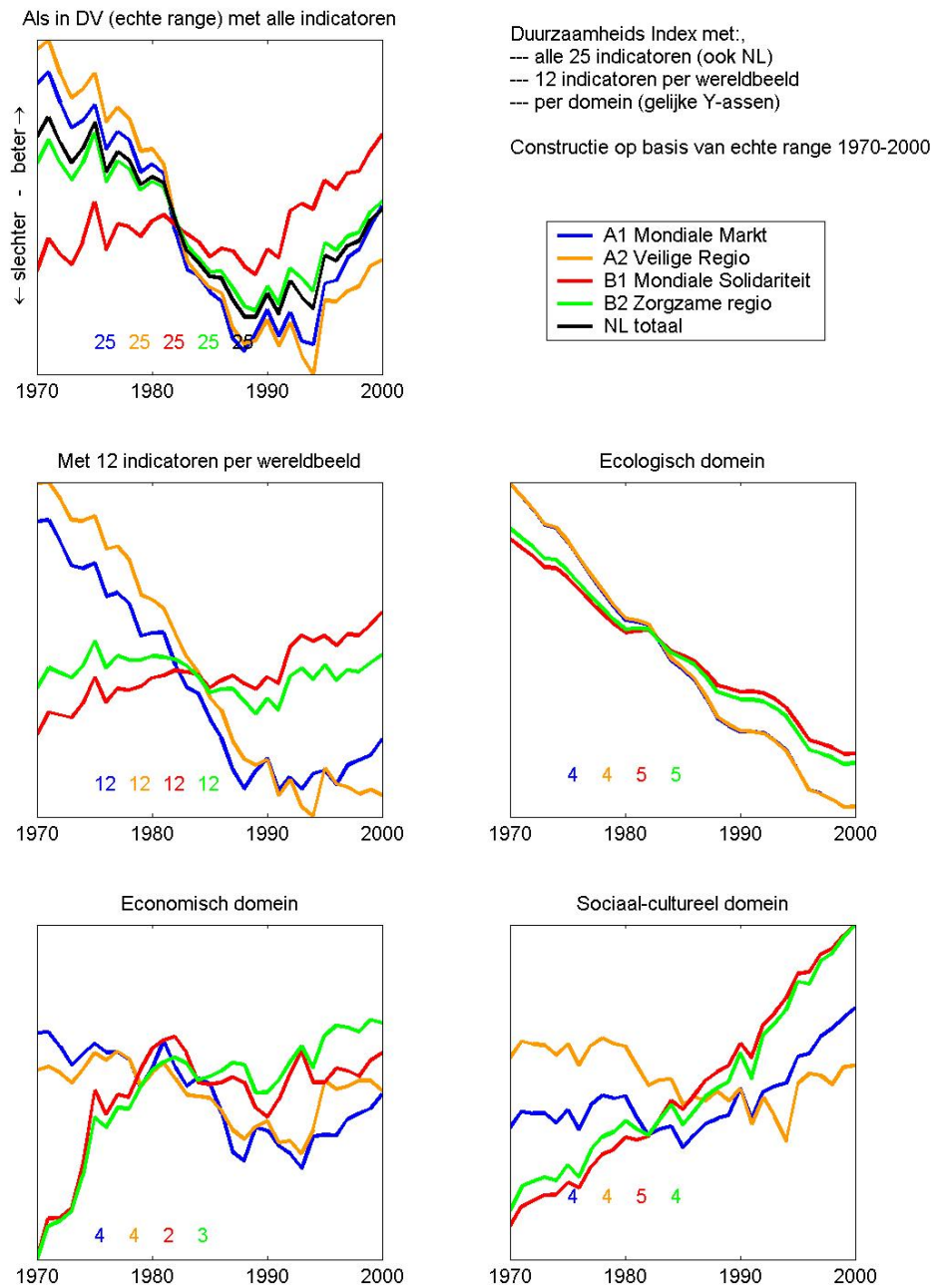


**Figuur 5. Varianten op de berekening van de duurzaamheidsindex.**



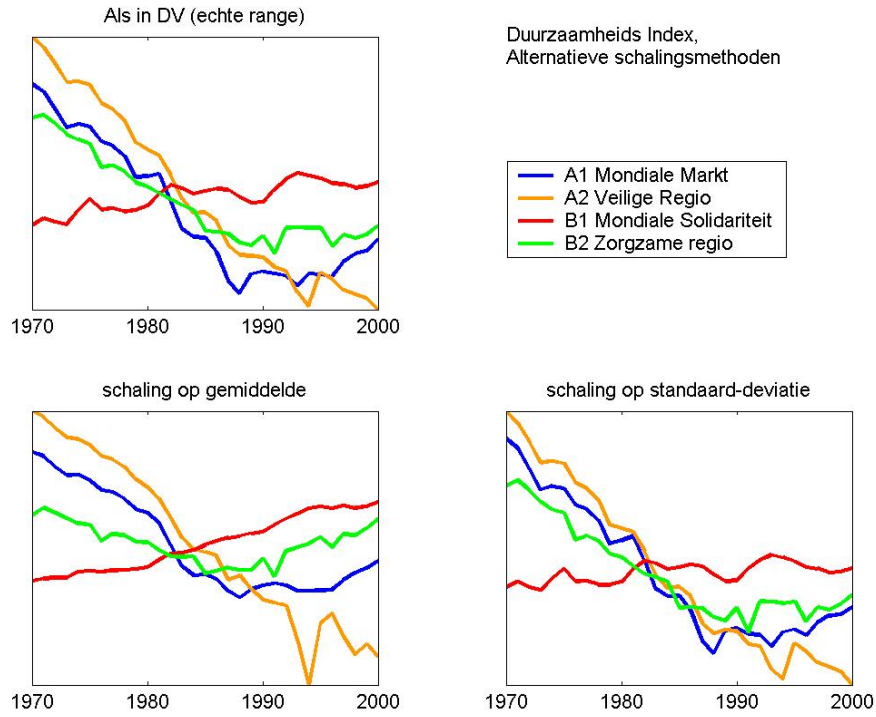
**Figuur 6. Duurzaamheidsindex met gelijk aantal (8) indicatoren per wereldbeeld en de duurzaamheidsindex uitgesplitst naar domein.**

**Bij de domeinen zijn gelijke y-assen gebruikt.**

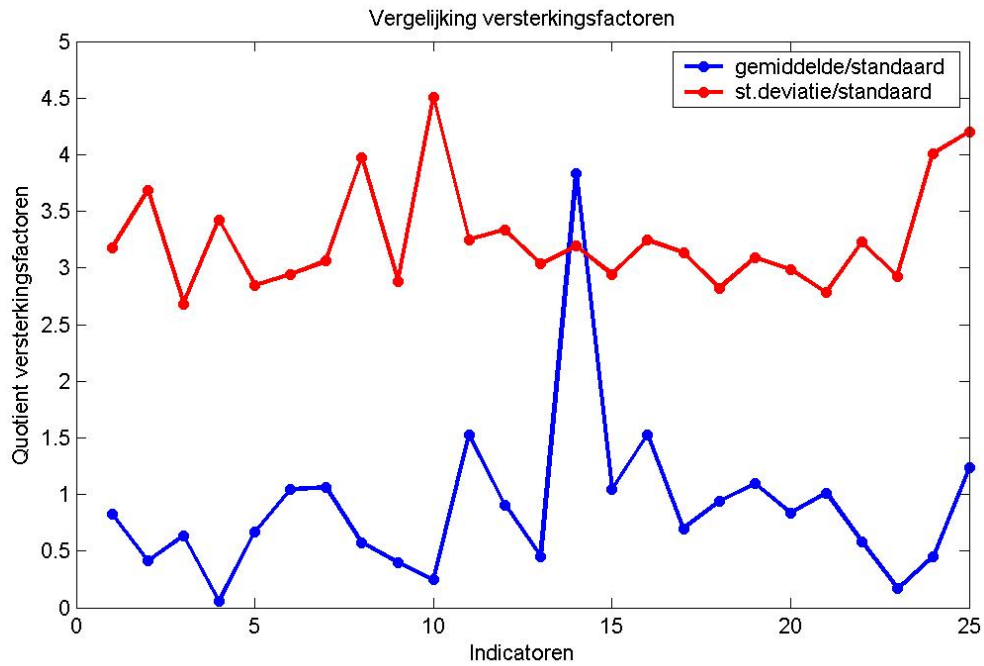


**Figuur 7. Duurzaamheidsindex voor de situatie dat, (1) alle 25 indicatoren worden meegenomen, (2) voor elke wereldbeeld de belangrijkste 12 indicatoren worden meegenomen en de uitsplitsing naar domein in dit laatste geval.**

**Bij de domeinen zijn gelijke y-assen gebruikt**



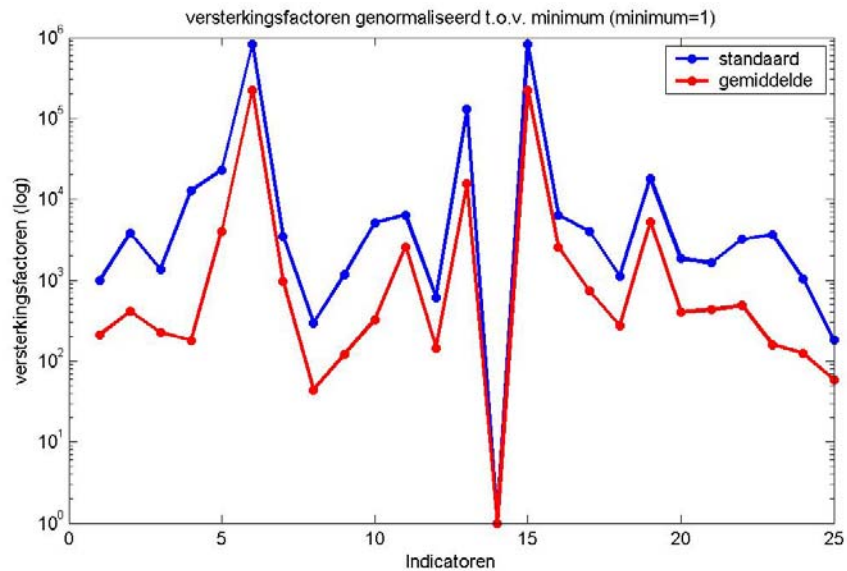
Figuur 8. Duurzaamheidsindices voor drie verschillende schalingsmethodieken



Figuur 9. Quotiënt van versterkingsfactoren voor verschillende schalingsmethoden.

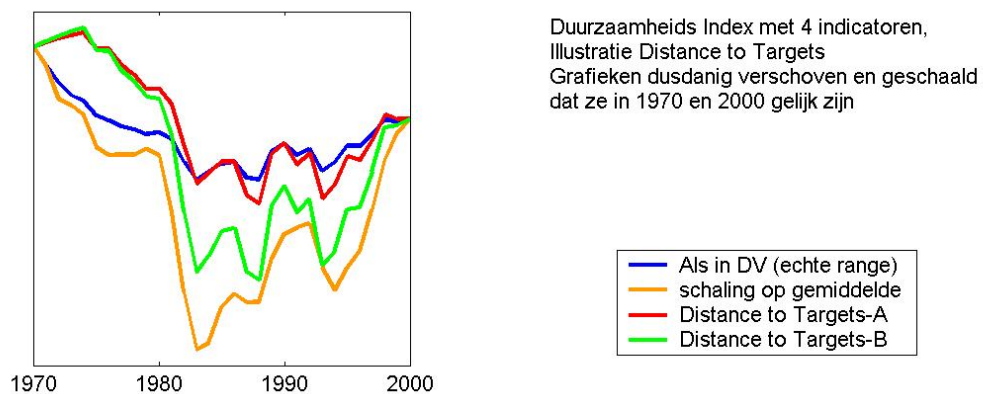
Uitgezet zijn quotiënten van de A-factoren uit Tabel 6.





**Figuur 10. Versterkingsfactoren voor de standaard methode en de methode gebaseerd op gemiddelden.**

**Data zijn geschaald t.o.v hun minimum.**

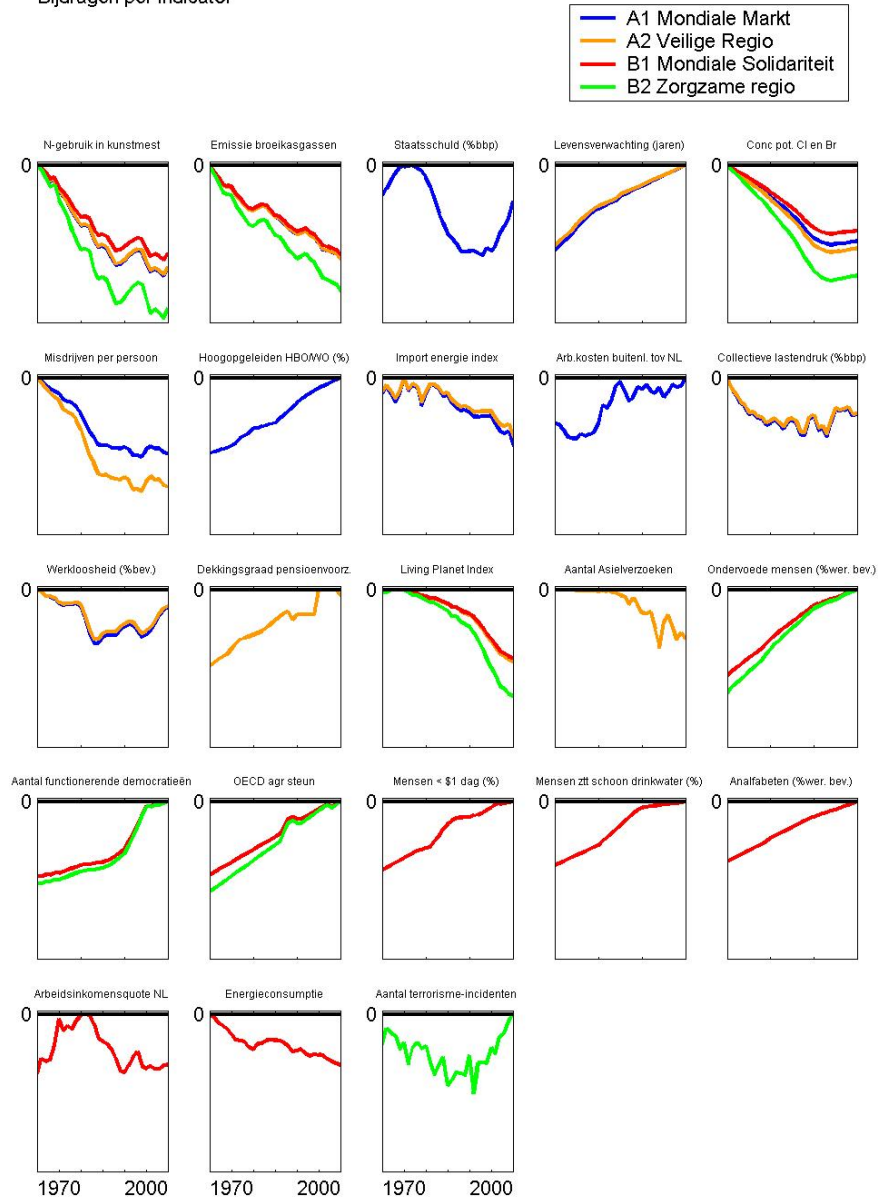


Duurzaamheids Index met 4 indicatoren,  
Illustratie Distance to Targets  
Grafieken dusdanig verschoven en geschaald  
dat ze in 1970 en 2000 gelijk zijn

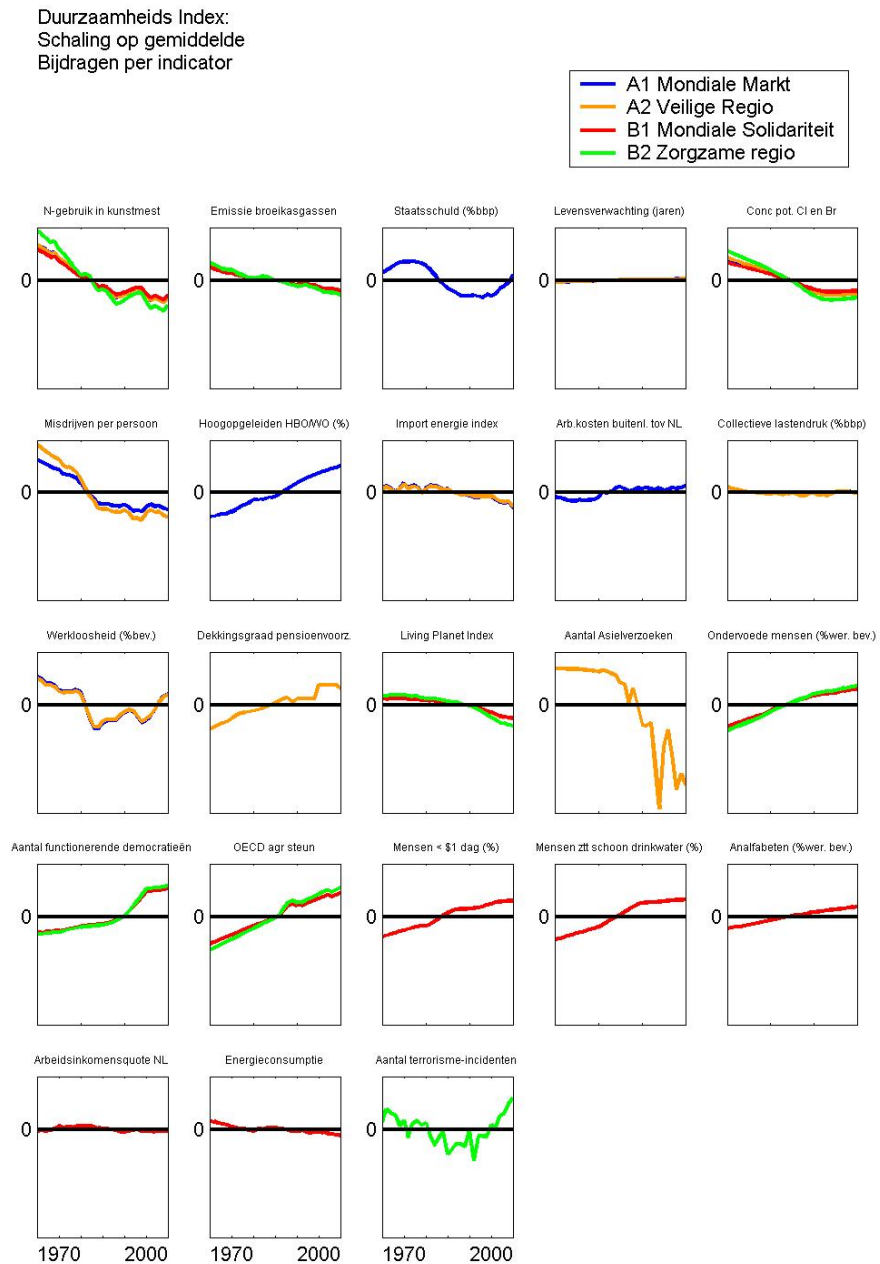
**Figuur 11. Illustratie van distance to targets benadering in vergelijking met standaardmethode en schaling op basis van gemiddelde.**

**De grafieken zijn dusdanig verschoven en geschaald dat de waarden in 1970 en 2000 overeenkomen. De gebruikte indicatoren en waarden van de targets/schalingsfactoren zijn vermeld in Tabel 7.**

Duurzaamheids Index:  
 Als in DV (echte range)  
 Bijdragen per indicator

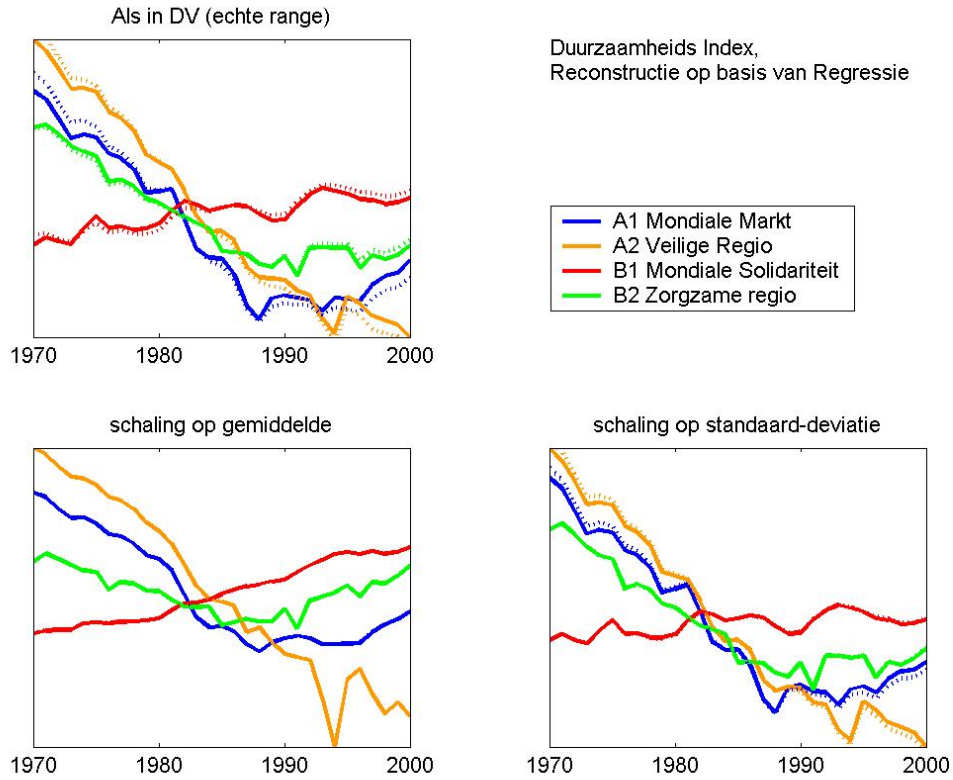


**Figuur 12. Bijdrage individuele indicatoren aan duurzaamheidsindex met schaling zoals in DV.  
 De verticale assen zijn voor alle plaatjes gelijk.**

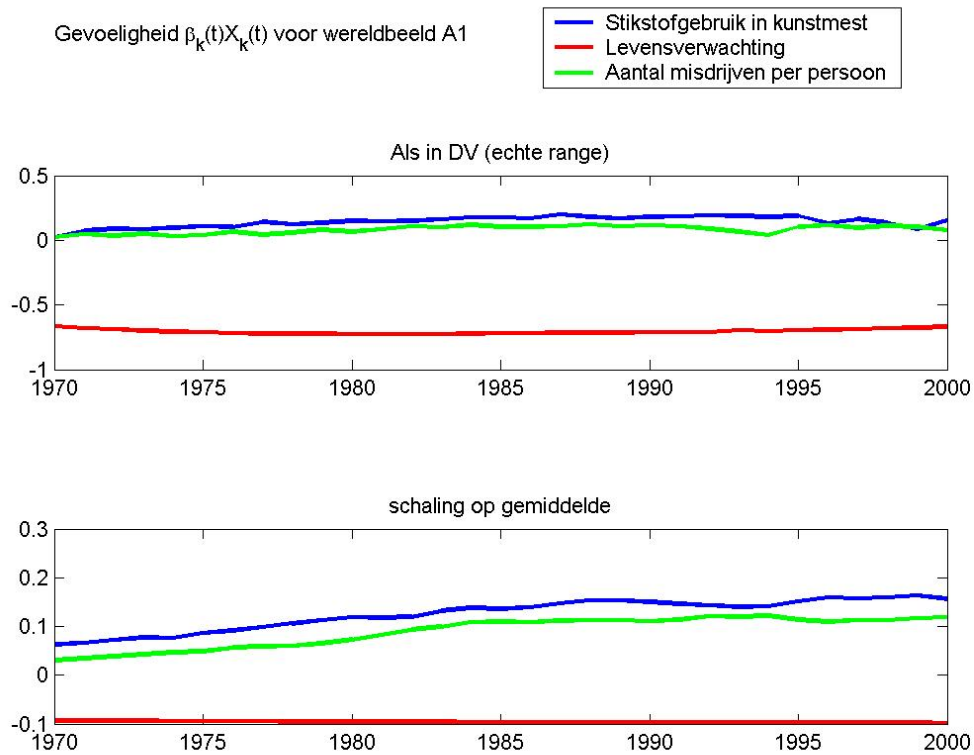


**Figuur 13. Bijdrage individuele indicatoren aan duurzaamheidsindex met schaling op basis van het gemiddelde.**

**De verticale assen zijn voor alle plaatjes gelijk.**



**Figuur 14. Duurzaamheidsindices voor drie verschillende schalingsmethodieken en de lineaire benadering van deze indexen op basis van lineaire regressie (gestippelde lijnen). Regressie gebaseerd op Latin Hypercube sampling met onzekerheidsband van  $\pm 10\%$  op de ruwe indicatordata.**

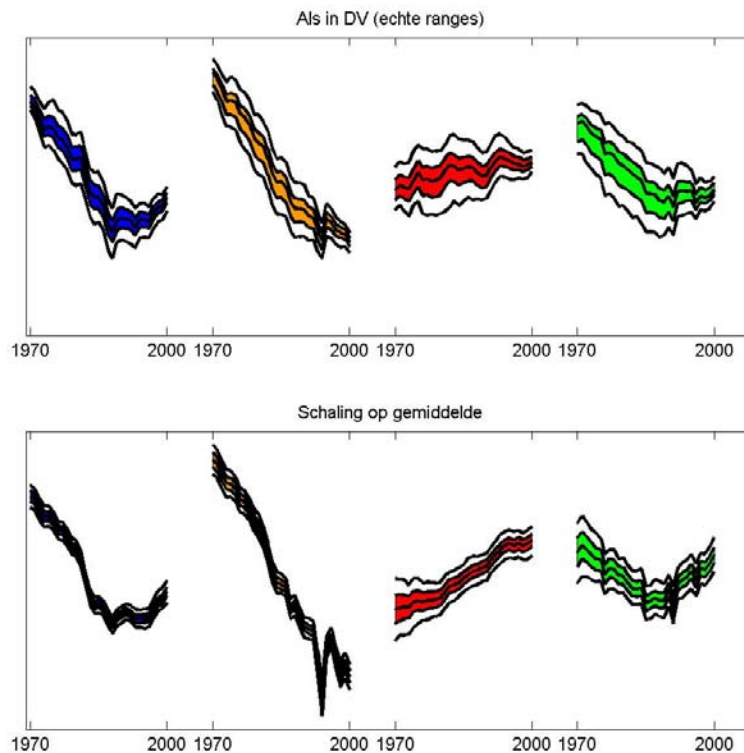


**Figuur 15.** Bijdragen in gevoeligheid voor de indicatoren *Stikstofgebruik in kunstmest*, *Levensverwachting* en *Aantal misdrijven per persoon*.

Weergegeven zijn de grootheden  $\beta_k(t)X_k(t)$ , voor wereldbeeld A1, waarbij  $\beta_k(t)$  de lineaire regressiecoëfficiënt voor jaar  $t$  weergeeft en  $X_k(t)$  de indicatorwaarde voor jaar  $t$ .

Duurzaamheids Index, onzekerheidsanalyse  
 Indicatoren van  $\pm A\%$  in 1970 tot  $B\%$  in 2000 (zie onderschrift voor A en B)  
 Weegfactoren verschillend gesampled per wereldbeeld  
 LHS sampling, met 90% correlatie  
 Sample grootte 8750 ( 875 variabelen)  
 Geschaald zodat indices voor A1  
 samenvallen in 1970 en 2000  
 Onzekerheidsbanden : 5 en 95 percentiel

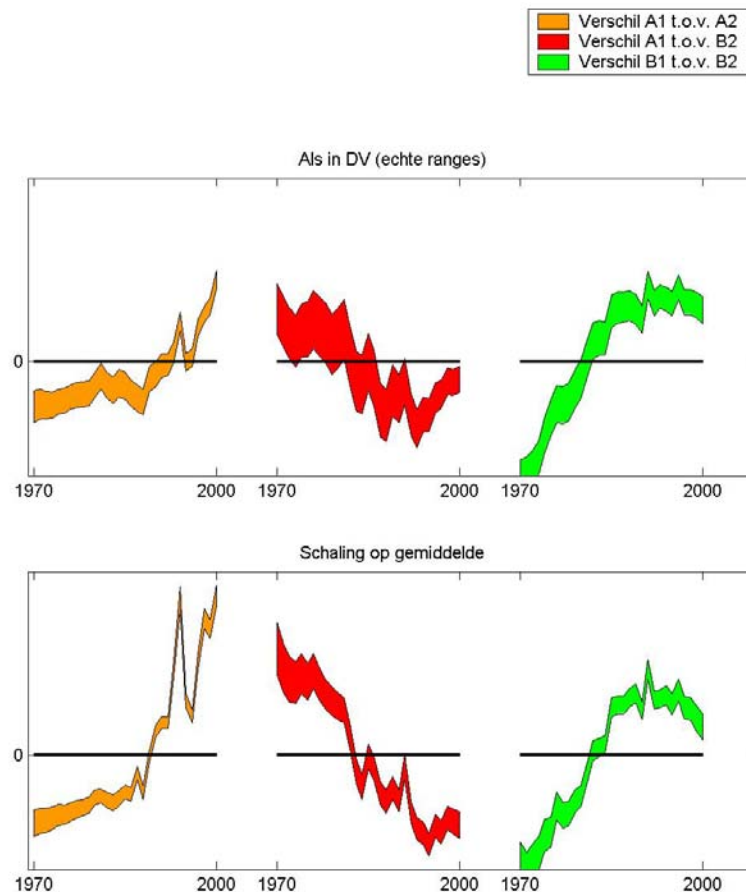
■ A1 Mondiale Markt  
■ A2 Veilige Regio  
■ B1 Mondiale Solidariteit  
■ B2 Zorgzame regio



**Figuur 16. Onzekerheidsmarges in duurzaamheidsindex voor standaardmethode en de methode gebaseerd op gemiddelden.**

Uitgaande van onzekerheden in indicatordata, variërend van  $\pm A\%$  in 1970 tot  $\pm B\%$  in 2000 (zie Tabel 10 voor waarden van A en B) en  $\pm 2\sigma$  onzekerheid in de weegfactoren (zie Tabel 9). Onzekerheid per indicator is van jaar tot jaar gecorreleerd met een correlatie van 90%. De indices zijn dusdanig geschaald dat de waarden in de indices zonder onzekerheid voor wereldbeeld A1 in 1970 en 2000 overeenkomen. Afgebeeld zijn de onzekerheidsbanden tussen het 5 en het 95 percentiel, gebaseerd op een Monte Carlo analyse met 8750 samples. Tevens zijn de minimale en maximale (0 en 100 percentiel) afgebeeld.

Onderlinge Verschillen Duurzaamheids Index, onzekerheidsanalyse  
 Indicatoren van  $\pm A\%$  in 1970 tot  $B\%$  in 2000 (zie onderschrift voor A en B)  
 Weegfactoren verschillend gesampled per wereldbeeld  
 LHS sampling, met 90% correlatie  
 Sample grootte 8750 ( 875 variabelen)  
 Geschaald zodat indices voor A1  
 samenvallen in 1970 en 2000  
 Onzekerheidsbanden : 5 en 95 percentiel



**Figuur 17. Onzekerheidsmarges in verschillen tussen duurzaamheidsindices per wereldbeeld voor standaardmethode en de methode gebaseerd op gemiddelden.**

Uitgaande van onzekerheden in indicatordata, variërend van  $\pm A\%$  in 1970 tot  $\pm B\%$  in 2000 (zie Tabel 10 voor waarden van A en B) en  $\pm 2\sigma$  onzekerheid in de weegfactoren (zie Tabel 9). Onzekerheid per indicator is van jaar tot jaar gecorreleerd met een correlatie van 90%. De indices zijn dusdanig geschaald dat de waarden in de indices zonder onzekerheid voor wereldbeeld A1 in 1970 en 2000 overeenkomen. Afgebeeld zijn de onzekerheidsbanden tussen het 5 en het 95 percentiel van de verschillend tussen wereldbeeld A1 en A2, A1 en B1, en tussen B1 en B2, gebaseerd op een Monte Carlo analyse met 8750 samples.