

**GEVOELIGHEIDSANALYSE VIA METAMODELLEN
EN PROEFOPZETTEN, TOEGEPAST OP EEN
SIMULATIEMODEL VOOR HET BROEIKASPROBLEEM**

J. Rotmans^{*}, G. van Ham^{**}, en J.P.C. Kleijnen^{**}

SAMENVATTING

Op het Rijksinstituut voor Volksgezondheid en Milieuhygiëne (RIVM) is een computersimulatiemodel voor het broeikasprobleem ontwikkeld: the Integrated Model for the Assessment of the Greenhouse Effect (IMAGE). In dit artikel wordt beschreven, hoe op dit model een gevoeligheidsanalyse is toegepast met behulp van de methode van metamodeltering. Daarbij worden afzonderlijke modules van het simulatiemodel benaderd met regressiemodellen ofwel metamodellen. Via de kleinste kwadraten methode worden de coëfficiënten ofwel parameters van het metamodel bepaald. Door een statistische proefopzet worden de experimenten met het simulatiemodel zo efficiënt en effectief mogelijk uitgevoerd. Aldus zijn voor verschillende modules van IMAGE adequate metamodellen ontwikkeld.

* Rijksinstituut voor Volksgezondheid en Milieuhygiëne
Postbus 1, 3720 BA Bilthoven; tel. 030-743323 ; fax 030-250740

** Katholieke Universiteit Brabant
Postbus 90153, 5000 LE Tilburg; tel. 013-662029 ; fax 013-663072

1. INLEIDING

Eén van de belangrijkste mondiale milieuproblemen voor de toekomst is het broeikasprobleem, dat wil zeggen het verschijnsel dat de aarde en de onderste lagen van de atmosfeer opgewarmd worden. Deze opwarming wordt veroorzaakt door de toename van een aantal in de atmosfeer voorkomende broeikasgassen, zoals kooldioxide, methaan, stikstofdioxide, freonen, ozon en waterdamp. De temperatuur op aarde wordt bepaald door de invallende zonnestraling en de warmte-uitstraling: de stralingsbalans. Deze broeikasgassen laten de invallende zonnestraling (kortgolfige straling) door, maar zij absorberen de door de aarde uitgestraalde warmte, en stralen deze weer uit, zowel naar de ruimte als naar het aardoppervlak (tegenstraling). Dit verschijnsel is bekend als het (versterkte) broeikaseffect. Dit door de mens versterkte broeikaseffect kan wereldwijde klimaatveranderingen teweegbrengen, waarbij de regionale verschillen groot kunnen zijn. Naar verwachting zal het temperatuurseffect groter zijn op hoge breedten en minder sterk in de tropen. De verwarming zal mede leiden tot een intensivering van de hydrologische cyclus en een verschuiving van het neerslag- en verdampingspatroon. Een van de mogelijke gevolgen is een zeespiegelstijging, veroorzaakt door drie processen: thermische expansie van de oceaan, afsmelting van gletsjers en smelting van landijs aan de Noord- en Zuidpool. Bovendien neemt de plantengroei toe, door de stijging van de CO_2 -concentratie, de hoeveelheid neerslag en de hogere temperaturen. In het tweede deeladvies van de Gezondheidsraad (1987) wordt een uitgebreid overzicht gegeven van de mogelijke maatschappelijke gevolgen van het broeikaseffect.

Teneinde een kwantitatief overzicht te bieden van deze complexe problematiek, heeft het RIVM een geïntegreerd simulatiemodel voor het broeikaseffect ontwikkeld: the Integrated Model for the Assessment of the Greenhouse Effect ofwel IMAGE. Het broeikasprobleem is gemodelleerd als een dynamisch systeem met discrete tijdstappen van een half jaar en een simulatietijd van 200 jaar, namelijk van het jaar 1900 tot 2100. Het computersimulatiemodel IMAGE bestaat uit afzonderlijke, onafhankelijke modules, die aan elkaar gekoppeld zijn (Rotmans et al., 1990); zie figuur 1. Elke module behandelt een afzonderlijk onderdeel van de problematiek,

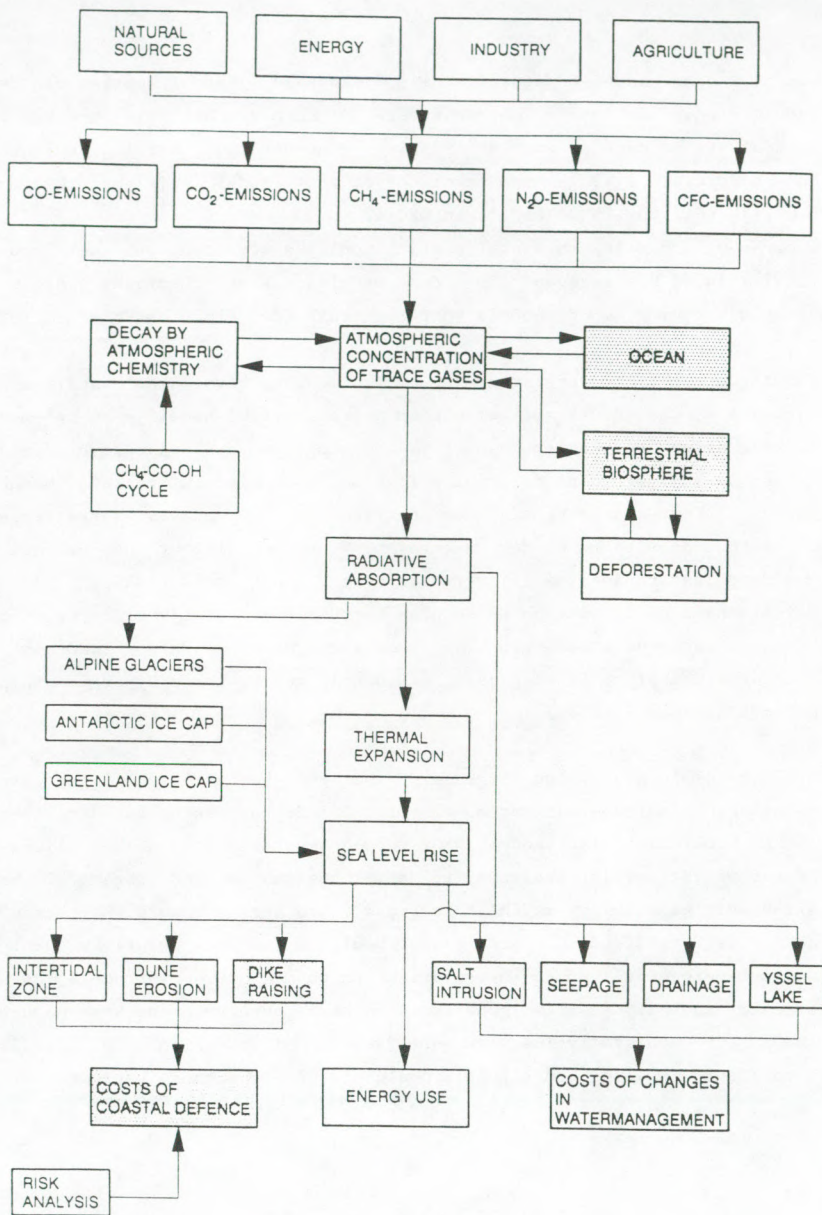
waarbij deze modules zodanig zijn gekoppeld, dat uitkomsten van de ene module dienen als invoer voor de andere. De kern van het geïntegreerde model bestaat uit de modules voor de emissies, concentraties, stralingshuishouding en zeespiegelstijging. Momenteel bevat het model vijf broeikasgassen, namelijk CO_2 , CH_4 , N_2O , CFC-11 en CFC-12.

Voor Nederland zijn een aantal maatschappelijke effecten van het probleem ontwikkeld, weergegeven in de modules kustverdediging, de module waterhuishouding, en de module energiegebruik (den Elzen en Rotmans, 1988).

In figuur 1 is aangegeven, op welke modules gevoeligheidsanalyses zijn uitgevoerd, die in dit artikel worden gepresenteerd, namelijk op de modules dijkverhoging en kosten van dijkverhoging, en de koolstofcyclus module. Alleen de analyse van de koolstofcyclus zal hier uitgebreid behandeld worden. Korte tijdshalve zal de analyse op de modules betreffende de dijkverhoging slechts worden samengevat. Voor een uitgebreide beschrijving van deze laatste analyse zij verwezen naar Van Ham et al. (1990).

Vermeldenswaard is, dat op de modules betreffende de dijkverhoging, en op het oceaangedeelte van de koolstofcyclus module eveneens een onzekerheidsanalyse is toegepast, met behulp van de Latin Hypercube Sampling methode (Lammerts, 1989).

Dit artikel is als volgt ingedeeld. Na een korte beschrijving van de gehanteerde methode van metamodelering zal de toepassing van deze techniek op het broeikassimulatiemodel IMAGE worden besproken. Gezien het tijdrovende aspect van dergelijke analyses, is de methode van metamodelering slechts op afzonderlijke modules van IMAGE toegepast. Als eerste module is gekozen voor de module betreffende de harde kustverdediging, omdat a priori verwacht werd dat het verband tussen de invoervariabelen en de uitvoervariabele nagenoeg lineair zou zijn. Vervolgens is de methode gebruikt om de kern module van IMAGE, de koolstofcyclus, te analyseren. Hierna zijn de resultaten gevalideerd en wordt een algehele evaluatie van de methode gepresenteerd.



Figuur 1. Schematische weergave van het model IMAGE. Op de gemarkeerde modules is een gevoeligheidsanalyse uitgevoerd.

2. METAMODELLERING

Overeenkomstig Kleijnen (1987) wordt een metamodel geïnterpreteerd als een regressiemodel van het feitelijke simulatiemodel. De relatie tussen de simulatie output Y , en k onderzochte inputs ofwel factoren z_k kan worden gerepresenteerd door

$$Y = f_1(z_1, z_2, \dots, z_k) \quad (1)$$

waarbij $z_1, \dots, z_k$ de gekozen factoren uit het simulatiemodel aanduiden. Gebruiken we de Taylor benadering, dan kan (1) als volgt benaderd worden door:

$$\tilde{Y} = \sum_{j=0}^n a_j w_j \approx Y \quad (2)$$

waarbij n een willekeurig natuurlijk getal is, a_j constanten zijn, $j \in \{0, \dots, n\}$, en w_j , $j \in \{0, \dots, n\}$ zijn geschikt gekozen functies van $z_1, \dots, z_k$ met $w_0 = 1$. Een eerste orde benadering levert $n = k$ en $w_j = z_j$, met $j \in \{0, \dots, k\}$, waar a_j de hoofdeffecten weergeeft. Daarentegen zou een benadering met interacties leiden tot $n = k + (k*(k-1))/2$, waarbij $w_0, \dots, w_k$ identiek zijn aan de factoren in het eerste orde geval, en $w_{k+1}, \dots, w_{k + (k*(k-1))/2}$ de eerste orde interacties reflecteren. Vanuit theoretisch oogpunt gezien is er geen enkele beperking ten aanzien van de functionele relatie tussen de w_k 's en de z_k 's; deze mag bijvoorbeeld logaritmisch van aard zijn: $w_k = \log(z_k)$.

Nu dienen de coëfficiënten $a_0, \dots, a_n$ in het metamodel te worden geschat, met behulp van de kleinste kwadraten methode, en dient te worden getoetst of het metamodel valide is. Zo niet, dan dient het metamodel te worden verbeterd, door bv. meer termen aan (2) toe te voegen. Stel Y_i is de uitvoer van run i waarbij de niveaus van de factoren $z_{i1}, \dots, z_{ik}$ leiden tot $w_{i1}, \dots, w_{in}$. Stel m is het aantal runs, $m \geq n + 1$, en W is de matrix:

$$W = \begin{matrix} & 1 & w_{11} & \dots & w_{1n} \\ \begin{matrix} 1 \\ \vdots \\ m \end{matrix} & \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} & \begin{matrix} w_{i1} \\ \vdots \\ w_{in} \end{matrix} & \dots & \begin{matrix} w_{in} \\ \vdots \\ w_{in} \end{matrix} \\ & 1 & w_{m1} & \dots & w_{mn} \end{matrix} \quad (3)$$

De kleinste kwadraten schatter voor de coëfficiënten $(a_0, a_1, \dots, a_n)$ is:

$$\hat{\beta} = (W' W)^{-1} W' \tilde{Y} \quad (4)$$

met $\tilde{Y} = (Y_1, \dots, Y_m)$. Vanzelfsprekend wordt in (4) verondersteld dat $(W'W)$ een niet-singuliere matrix is. (4) levert de volgende schatter (aangeduid met $\hat{Y}$) van $\tilde{Y}$ in (2):

$$\hat{Y} = W (W'W)^{-1} W' \tilde{Y}, \quad (5)$$

waarbij w de vector is van de functies $(w_0, w_1, \dots, w_n)$.

Deze wiskundige analyse kan worden uitgebreid met een statistisch model voor de aanpassingsfouten $e = Y - \hat{Y}$. Indien wordt aangenomen dat deze fouten normaal en onafhankelijk verdeeld zijn, met constante variantie, dan is $\hat{\beta}$ de beste zuivere schatter en kan de significantie van $\hat{\beta}$ getoetst worden, zie Kleijnen (1990). Recentelijk is door een groep Amerikaanse statistici een algemeen model voorgesteld: de e is niet onafhankelijk verdeeld, maar vormt een stochastisch proces met een bepaalde auto-correlatie-functie, zie Sachs et al. (1989).

3. METAMODELLEN VOOR DE MODULES VOOR DIJKVERHOOGING EN KOSTEN VAN DIJKVERHOOGING

3.1. Inleiding

De techniek van metamodeltering is reeds eerder toegepast op onderdelen van het computersimulatiemodel IMAGE (Rotmans e.a., 1988, Rotmans en Vrieze, 1989). Thans is deze analyse eerst toegepast op de modules voor de harde kustverdediging. In deze modules wordt de dijkverhoging respectievelijk de kosten van zo'n dijkverhoging berekend aan de hand van de gesimuleerde zeespiegelstijging. Daarbij worden ondermeer de schadeverwachting bij een dijkdoorbraak en de overschrijdingsfrequenties van de verschillende zeeniveaus in beschouwing genomen (den Elzen en Rotmans, 1988). Hier zullen

slechts de belangrijkste bevindingen die zijn opgedaan bij deze analyse worden gepresenteerd. Technische details worden pas besproken in de volgende paragraaf.

3.2. Resultaten

In deze analyse worden de cumulatieve kosten van dijkverhoging (in miljarden guldens) bij een dijkverhogingsstrategie van gelijkblijvende veiligheidsnorm gekozen als uitvoervariabele van het simulatiemodel. Als eindtijdstip waarop deze dijkverhogingskosten betrekking hebben, werd het jaar 2100 genomen, het eindtijdstip van de simulatie. Er worden 11 factoren geselecteerd, waaronder de initiële kosten van dijkverhoging, de kosten om een dijk één meter te verhogen, de minimale dijk aanpassing, etc. Er zijn uiteindelijk negen verschillende metamodelen gespecificeerd en getoetst, welke zijn doorerekend met negen verschillende 2^{k-p} proefopzetten. Deze negen metamodelen zijn derhalve geen overlappende deelmodellen die aan de hand van één dataset (gegenereerd door één proefopzet) worden vergeleken. Enige belangrijke resultaten van deze analyse zijn:

- * In eerste instantie blijkt het niet mogelijk een valide metamodel op te stellen. De geschatte effecten hebben niet de juiste tekens, waardoor zo'n model geen verklarende en voorspellende werking heeft. Na grondige analyse van het simulatiemodel blijkt, dat dit niet goed functioneerde, en dat de simulatie diende te worden opgesplitst. Eerst moet de dijkverhoging worden uitgerekend, waarna deze separaat moet worden ingelezen in de module die de kosten van dijkverhoging berekent. In eerste instantie werden deze dijkverhogingskosten echter berekend, zonder dat de berekende dijkverhoging steeds werd aangepast en ingelezen. We zien hier dus een bijzonder nuttige toepassing van metamodelering. Doordat deze methode een diepgaande verkenning van het simulatiemodel vereist, kunnen op deze wijze onjuistheden of onvolkomenheden binnen het simulatiemodel aan het licht komen. Bovendien kan een afwijking tussen het simulatie- en het metamodel duiden op verkeerde combinaties van invoerwaarden: het simulatiemodel kan fout werken buiten een bepaald "experimenteer-gebied", zie 4.3.2 (inkrimping van het gebied).

* Het effect van standaardisatie van de factoren is nagegaan. Stel dat het oorspronkelijke metamodel geformuleerd kan worden als:

$$\tilde{Y} = \sum_{j=0}^k \beta_j z_j \quad (6)$$

Via de volgende eenvoudige lineaire transformatie kunnen de oorspronkelijke factoren worden genormeerd tot factoren met ondergrens -1 en bovengrens +1 (Kleijnen, 1987, pag. 341-345)

$$z_j = c_j x_j + d_j \quad (7)$$

met $c_j = (H_j - L_j)/2$ en $d_j = (H_j + L_j)/2$; H_j en L_j zijn de hoogste respectievelijk laagste waarde van de oorspronkelijke factor z_j . Dan wordt vergelijking (6) omgezet in:

$$\tilde{Y} = \sum_{j=0}^k \alpha_j x_j \quad (8)$$

Een alternatief is het "gecentreerde" model:

$$\tilde{Y} = \sum_{j=0}^k \delta_j (z_j - \bar{z}_j) \quad (9)$$

met $\bar{z}_j = (H_j + L_j)/2$

Bettonvil en Kleijnen (1988) geven het algemene verband tussen α , β , en δ . Voor alle drie de metamodellen, weergegeven in (6), (8), en (9) is de volgorde van de factoren, gerangschikt naar afnemend belang, vergeleken met de volgorde bij een analyse, waarbij telkens slechts één factor wordt gevarieerd. Hieruit blijkt, dat het gestandaardiseerde model de juiste volgorde van de factoren aangeeft, in tegenstelling tot het oorspronkelijke of gecentreerde model: het gestandaardiseerde effect geeft aan hoe de response verandert indien de betreffende factor stijgt van zijn minimum naar zijn maximum. Dit resultaat stemt overeen met de algemene conclusie in Kleijnen (1987) en Bettonvil en Kleijnen (1988). In de

volgende analyse wordt derhalve alleen nog met gestandaardiseerde modellen gewerkt.

- * Het uiteindelijke metamodel bevat negen significante factoren, alsmede één significante twee-factor interactie. Hierbij blijkt de keuze van de grenzen van de geselecteerde factoren van doorslaggevend belang te zijn. Bij het definitieve metamodel zijn deze grenzen verengd ten opzichte van de oorspronkelijke waarden. De kostenfactor om een dijk één meter te verhogen blijkt veruit de belangrijkste factor te zijn, wat overeenkomstig de verwachting is. De volgorde van de resterende acht significante factoren, alsmede de interactie, bevat daarentegen enige verrassende aspecten; dit heeft mede geleid tot een vergroot inzicht in de werking van het simulatiemodel.
- * Teneinde niet alleen het verklarende maar ook het voorspellende effect van het metamodel te toetsen, is het metamodel op verschillende manieren gevalideerd. Via achtereenvolgens kruisvalidatie, de uitvoering van extra runs, en tenslotte via het uitvoeren van twee extra runs met gekozen punten binnen het experimenteer-gebied, is het metamodel getoetst. In alle gevallen blijken de relatieve residuen (het procentuele verschil tussen het berekende effect en het geschatte effect) niet groter dan 10 % te zijn (van Ham e.a., 1990).
- * De resultaten van deze analyse met metamodellen stemmen grotendeels overeen met de resultaten volgens de Latin Hypercube Sampling methode, uitgevoerd door Lammerts (1989).

4. METAMODELLEN VOOR DE KOOLSTOF CYCLUS MODULE

4.1. Inleiding

Nadat de modules voor dijkverhoging en kosten van dijkverhoging zijn geanalyseerd, is een gevoeligheidsanalyse uitgevoerd op de koolstofcyclus module, welke het verband aangeeft tussen de CO₂-concentratie in de

atmosfeer, de terrestrische biosfeer en de oceanen, zie figuur 1. De koolstofcyclus module bestaat uit twee afzonderlijke delen, een module voor de uitwisseling van CO_2 tussen de oceaan en de atmosfeer, en een voor de uitwisseling van CO_2 tussen de terrestrische biota en de atmosfeer. Op beide modules is metamodeltering toegepast. Via de metamodeltering wordt getracht een verband aan te geven tussen enerzijds de CO_2 -concentratie in de atmosfeer, en anderzijds de belangrijkste parameters uit de twee onderhavige modules. Als uitvoervariabele wordt gekozen de CO_2 -concentratie in de atmosfeer (in ppm, deeltjes per miljoen luchtdeeltjes). Als tijdstip waarop deze CO_2 -concentratie betrekking heeft, wordt wederom het jaar 2100 gekozen. De analyse is gesplitst in een op te stellen metamodel voor de oceaan module en een voor de terrestrische biosfeer.

4.2. Metamodellen voor de oceaan module

4.2.1. Keuze van de factoren en hun grenzen

In de oceaan module van IMAGE zijn de oceanen op eenvoudige wijze geschematiseerd, en opgedeeld in de volgende lagen: op de lagere breedtegraden een menglaag van 75 meter met daaronder een laag van circa 325 meter en op de hogere breedtegraden een laag van ongeveer 400 meter, en onder deze lagen negen lagen van ieder circa 378 meter. In de oceanen vinden met betrekking tot CO_2 drie processen plaats (Goudriaan en Ketner, 1984):

1. Water wordt van de polen via de diepe oceaanalagen naar de evenaar gepompt. Via de bovenste laag stroomt het water weer terug naar de polen.
2. Tussen de verschillende lagen treedt diffusie op.
3. Onopgelost CO_2 slaat neer op de bodem van de oceaan.

Voor deze drie primaire processen zijn de volgende tien factoren met hun grenzen onderzocht voor het metamodel, zie tabel 4.1:

Tabel 4.1 Factoren met hun grenzen

factor	omschrijving	minimum waarde	basis waarde	maximum waarde	eenheid
z_1	neerslag van onopgelost CO_2 op de bodem van de oceaan	7	8	9	GtC/jr
z_2	diffusiecoëfficiënt	3716	4000	5984	cm^2/sec
z_3	dikte van de warmere menglaag	65	70	80	m
z_4	dikte van de koudere menglaag	300	400	500	m
z_5	dikte van de negen onderlagen	3000	3400	4000	m
z_6	verblijftijd van CO_2 in de warmere menglaag	0.05	1	2	jaar
z_7	verblijftijd van CO_2 in de koudere menglaag	15	20	25	jaar
z_8	massastroom van water	$2.1 \cdot 10^{15}$	$2.3 \cdot 10^{15}$	$2.5 \cdot 10^{15}$	m^3/yr
z_9	oppervlakte van de oceaan	$0.32 \cdot 10^{12}$	$0.36 \cdot 10^{12}$	$0.40 \cdot 10^{12}$	m^2
z_{10}	atmosferische coëfficiënt	0.469	0.471	0.472	ppm/GtC

4.2.2. Specificatie van het metamodel

Eerst worden de oorspronkelijke factoren z_j getransformeerd tot genormeerde factoren x_j , zie (7). Elf potentieel belangrijke interacties zijn gespecificeerd, namelijk die tussen de factoren x_1 en x_3 , x_1 en x_5 , x_2 en x_3 , x_2 en x_5 , x_3 en x_6 , x_3 en x_7 , x_5 en x_6 , x_5 en x_7 , x_5 en x_8 , x_6 en x_8 en x_7 en x_8 . Het simulatiemodel voor de oceaan kan derhalve benaderd worden door het volgende metamodel:

$$\begin{aligned}
 Y_m(i) = & \alpha_0 + \sum_{j=1}^{10} \alpha_j x_j(i) + \alpha_{1,3} x_1(i) x_3(i) + \alpha_{1,5} x_1(i) x_5(i) + \alpha_{2,3} x_2(i) x_3(i) \\
 & + \alpha_{2,5} x_2(i) x_5(i) + \alpha_{3,6} x_3(i) x_6(i) + \alpha_{3,7} x_3(i) x_7(i) + \alpha_{5,6} x_5(i) x_6(i) \\
 & + \alpha_{5,7} x_5(i) x_7(i) + \alpha_{5,8} x_5(i) x_8(i) + \alpha_{6,8} x_6(i) x_8(i) + \alpha_{7,8} x_7(i) x_8(i) \\
 & (10)
 \end{aligned}$$

met:

- $Y_m(i)$ = de benaderde CO_2 -concentratie in de atmosfeer in het jaar 2100 in de i -de combinatie van factor-niveaus;
 $x_j(i)$ = de waarde van factor x_j in het i -de experiment;
 $x_j(i)x_1(i)$ = de waarde van interactie tussen factoren x_j en x_1 in i -de run
 α_j = het hoofdeffect van factor x_j ;
 $\alpha_{j,1}$ = de interactie tussen factor x_j en x_1 , $j \neq 1$

4.2.3. De proefopzet voor het eerste metamodel

Een volledige factoriële proefopzet ofwel een 2^k proefopzet, waarbij k het aantal factoren voorstelt, zou 2^{10} runs vereisen. Dat eist te veel computertijd, vandaar dat voor een andere proefopzet wordt gekozen: een 2^{k-p} proefopzet; zie Kleijnen (1987). Rekening houdend met de elf specifieke interacties tussen de tien factoren is het aantal te schatten effecten (q) dus $1 + 10 + 11 = 22$. Voor de kleinste kwadraten schatter geldt $n = 2^{k-p} > q$; dus moeten er $n = 2^{10-5} = 32$ runs worden uitgevoerd om de parameters in het metamodel te schatten. Van de tien factoren worden vijf factoren gekozen die relatief weinig voorkomen in de belangrijk geachte interacties: x_1 , x_2 , x_4 , x_9 en x_{10} . Voor deze vijf factoren wordt een volledige factoriële proefopzet opgesteld, wat 2^5 runs vereist. De overige vijf factoren (x_3 , x_5 , x_6 , x_7 , x_8) worden als combinaties van de vijf hoofdfactoren geschreven, hetgeen leidt tot de volgende generatoren:

$$3 = 9 \ 10 \quad 5 = 4 \ 10 \quad 6 = 1 \ 9 \quad 7 = 2 \ 9 \quad 8 = 1 \ 2 \ 4 \quad (11)$$

Waarbij de notatie $3 = 9 \ 10$ betekent $x_3(i) = x_9(i) x_{10}(i)$, d.w.z. de waarde van factor 3 in combinatie i is gelijk aan de waarde van factor 9 in combinatie i , vermenigvuldigd met de waarde van factor 10, etc. Daardoor is het hoofdeffect van factor 3 gestrengeld met de interactie tussen factoren 9 en 10, met $\hat{\alpha}_3 = \hat{\alpha}_{910}$ en $E(\hat{\alpha}_3) = \alpha_3 + \alpha_{910}$, zie Kleijnen, 1987, pag. 295-300).

4.2.4. Resultaten

De geschatte CO_2 -concentratie ($\hat{Y}_m$) kan nu worden bepaald en vergeleken met de gesimuleerde CO_2 -concentratie (Y_s). De relatieve residuen $100 \cdot (Y_s - \hat{Y}_m) / Y_s$ blijken echter in acht runs groter dan 10 % te zijn, hetgeen onacceptabel geacht wordt. Eén van de mogelijkheden om een niet-valide metamodel te herzien, is de grenzen van de factoren aan te passen. Daarom is een volgende reeks metamodellen opgesteld, waarbij de grenzen van x_3 , x_4 , x_5 en x_{10} zijn gevarieerd. De relatieve residuen blijken echter alleen maar groter te worden. Toen is besloten om een nieuw metamodel op te stellen, waarmee ook kwadratische effecten kunnen worden getoetst. Een nieuwe serie experimenten is uitgevoerd, met behulp van een centraal samengestelde proefopzet. Dit soort proefopzet maakt het mogelijk ook de kwadratische effecten van de factoren te schatten evenals sommige interacties, afhankelijk van het soort proefopzet (Kleijnen, 1987, pag. 200). Een nieuw metamodel wordt gespecificeerd met alleen de factoren x_3 , x_4 , x_5 , x_7 en x_{10} . Uitgangspunt is een 2^{5-1} ontwerp met niveaus -1 en +1, met als generator:

$$4 = 3 \ 7 \ 10$$

(12)

Voor de centraal samengestelde proefopzet wordt dit 2^{5-1} ontwerp aangevuld met runs waarin naast -1 en +1 ook het centrale punt 0 en twee andere waarden zijn gekozen, in dit geval -2 en +2; hierdoor kunnen de factoren vijf waarden aannemen. Het aantal runs komt daardoor op 27 ($16 + 2 \cdot 5 + 1$). Het aantal te schatten effecten is 15 (6 hoofdeffecten, 5 kwadraten, en 4 interacties). Dit leidt tot tabel 4.2, waarin alleen de factoren staan weergegeven, die significant zijn bij $\alpha = 0.05$, weergegeven met een *. Indien geen statistisch submodel voor de aanpassingsfouten e wordt aangenomen, dan kan de laatste kolom verwaarloosd worden; de tabel geeft dan niet de "significante" effecten, maar de "relatief belangrijke".

Tabel 4.2 Factoren met geschatte significante effecten en t-waarden

factor	geschat effect	Student t_{12}
x_0	1074.66	154.08
x_3	51.77	32.07
x_4	158.37	98.10
x_5	-224.95	-139.35
x_7	18.78	11.63
x_5^2	27.78	14.05
$x_3 x_5$	-8.35	-4.22
$x_4 x_5$	-17.65	-8.93
$x_4 x_7$	-12.26	-6.20

Dit model wordt wederom op diverse manieren gevalideerd. De relatieve residuen van de reeds uitgevoerde runs staan in de tabellen 4.3 en 4.4 vermeld.

Tabel 4.3 De geschatte en berekende CO₂-concentraties en relatieve residuen

run	geschatte CO ₂ -concentratie ($\hat{Y}_m$)	berekende CO ₂ -concentratie (Y_s)	relatieve residuen $100 * ((Y_s - \hat{Y}_m) / Y_s)$ (in %)
1	1065	1074	0.85
2	1567	1565	-0.15
3	722	719	-0.40
4	1119	1114	-0.50
5	740	741	0.16
6	1171	1162	-0.83
7	959	963	0.40
8	1428	1432	0.31
9	803	803	0.06
10	1234	1238	0.29
11	973	974	0.16
12	1441	1435	-0.46
13	1053	1053	-0.01
14	1555	1550	-0.31
15	660	669	1.42
16	1058	1054	-0.37
17	1073	1074	0.07
18	1074	1074	-0.02
19	1107	1108	-0.01
20	1032	1032	-0.05
21	1176	1177	0.08
22	969	969	-0.05
23	736	724	-1.62
24	1636	1648	0.74
25	1376	1373	-0.20
26	743	746	0.46
27	1075	1074	-0.06

Tevens is kruisvalidatie uitgevoerd; een bepaalde run (combinatie) is weggelaten, de effecten worden geschat uit de overige 26 runs, en de uitkomst van de weggelaten runs wordt geschat. Zes runs zijn willekeurig weggelaten, zie tabel 4.4. Opgemerkt zij dat de berekening van de residuen mogelijk is zonder het kleinste kwadraten algoritme telkens opnieuw toe te passen op de overige (n-1) waarnemingen. Na weglaten van waarneming i worden de nieuwe residuen (zeg) r_{-i} :

$$r_{-i} = r_i / (1 - h_{ii}), \quad (13)$$

de "leverage factor", waarbij h_{ii} het i-de element is van de matrix

$$H = X (X'X)^{-1} X' \quad (14)$$

Tabel 4.4 Kruisvalidatie

weggelaten run	Y_s	$\hat{Y}_m$	$100 * (Y_s - \hat{Y}_m) / Y_s$
3	718.58	724.49	-0.80
4	1113.79	1125.15	-1.00
6	1161.79	1183.17	-1.84
11	974.41	971.06	0.34
15	669.42	650.12	2.88
16	1053.87	1061.79	-0.75

Tevens is een serie experimenten uitgevoerd volgens een random proefopzet, d.w.z. met willekeurige combinaties van waarden van de invoervariabelen. Voor de resultaten hiervan wordt verwezen naar Van Ham e.a. (1990). De relatieve residuen zijn voor geen enkel experiment groter dan 3 %, hetgeen acceptabel is. Tenslotte is een extra validatierun voor waarden binnen het "experimenteergebied" uitgevoerd: die levert een $\hat{Y}_m$ van 1043.58; de bijbehorende Y_s is 1055.45, zodat het relatieve residu 1.12 % is.

Tenslotte is de regressie nog eens uitgevoerd, waarbij de niet-significante factoren x_{10}^2 , x_3^2 , x_4^2 , x_7^2 , x_{10}^2 en $x_3 x_4$ worden weggelaten. Dit geeft geen grote veranderingen in de puntschattingen, terwijl alle factoren significant blijven.

4.2.5. Conclusie

De relatieve residuen van het uiteindelijke metamodel voor de oceaanmodule zijn nergens groter dan 3%. Dit model wordt derhalve goedgekeurd. De de factoren die de CO_2 -concentratie in de atmosfeer beïnvloeden, zijn in volgorde van afnemend belang:

- x_5 : de totale dikte van de negen onderlagen;
- x_4 : de dikte van de koudere menglaag;
- x_3 : de dikte van de warmere menglaag;
- x_7 : de verblijftijd van CO_2 in de koudere menglaag;
- x_5^2 : de totale dikte van de negen onderlagen, in het kwadraat;
- en de interacties tussen x_4 en x_5 , x_4 en x_7 , en x_3 en x_5 .

De overige zes hoofdeffecten en hogere orde effecten zijn te verwaarlozen.

4.3. Metamodellen voor de terrestrische biosfeer module van IMAGE

4.3.1. Inleiding

Als laatste wordt metamodeltering toegepast op de module voor de terrestrische biosfeer. De terrestrische biosfeer is een belangrijk reservoir binnen de koolstofcyclus. Tussen de atmosfeer en de terrestrische biosfeer vindt uitwisseling van enorme hoeveelheden kooldioxide plaats. De biosfeer is horizontaal in zes ecosystemen verdeeld: tropisch bos, gematigd bos, grasland, landbouwgrond, menselijk gebied en woestijnachtig gebied. Verticaal zijn onderscheiden: bladeren, takken, stammen, wortels, humus en houtskool. Binnen een ecosysteem vindt een stroom van koolstof plaats (op natuurlijke wijze, via bladeren naar houtskool, in uitwisseling met de atmosfeer). Door menselijk toedoen vinden ook verschuivingen plaats tussen de ecosystemen onderling, waarbij koolstofstromen plaatsvinden. Bovendien hebben menselijke activiteiten als zwerflandbouw en verbranding grote invloed op de hoeveelheid koolstof in een ecosysteem (Goudriaan en Ketner, 1984).

4.3.2. Specificatie van metamodellen voor de terrestrische biosfeermodule

Voor de beschreven processen zijn van de ongeveer 200 factoren van deze module 62 factoren met de bijbehorende grenzen onderzocht:

- z_1 : de fractie houtskool uit de decompositie van humus;
- z_2 : de relatieve groei van de bevolking;
- z_3 : de biotische groeifactor;
- z_4, z_5 en z_6 : de fractie houtskool door de verbranding van bladeren, voor resp. tropisch bos, gematigd bos, en grasland;
- $z_7, \dots, z_{12}$: de verblijftijd van houtskool in resp. ecosysteem 1, ..., 6;
- $z_{13}, \dots, z_{18}$: de verblijftijd van humus in resp. ecosysteem 1, ..., 6;
- $z_{19}, \dots, z_{24}$: de verblijftijd van strooisel in resp. ecosysteem 1, ..., 6;
- $z_{25}, \dots, z_{30}$: de humificatiefactor van resp. ecosysteem 1, ..., 6;
- $z_{31}, \dots, z_{36}$: zwerflandbouw van resp. ecosysteem 1, ..., 6;
- z_{37} : de verschuiving van tropisch bos naar grasland;
- z_{38} : de verschuiving van tropisch bos naar landbouwgrond.
- $z_{39}, \dots, z_{44}$: de levensduur van bladeren in resp. ecosysteem 1, ..., 6;
- $z_{45}, \dots, z_{50}$: de levensduur van takken in resp. ecosysteem 1, ..., 6;
- $z_{51}, \dots, z_{56}$: de levensduur van stammen in resp. ecosysteem 1, ..., 6; en
- $z_{57}, \dots, z_{62}$: de levensduur van wortels in resp. ecosysteem 1, ..., 6.

De oorspronkelijke factoren z_j worden weer genormaliseerd tot factoren x_j , zie (7). Vanwege de complexiteit en het vele benodigde rekenwerk zijn in eerste instantie geen interacties meegenomen in het metamodel. De 62 hoofdeffecten van dit model worden in ($2^{62-55} =$) 128 runs geschat. Bij dit model is het hoofdeffect van factor x_{62} significant positief, terwijl verwacht wordt dat deze parameter negatief is. Dit metamodel kan dan ook niet geaccepteerd worden.

Hierna is, in overleg met de modelbouwers, het domein van de invoervariabelen verkleind. De basiswaarden en de minimum- en maximumwaarden van de factoren staan in tabel 4.5. Naast de 62 hoofdeffecten worden nu tevens 26 belangrijk geachte twee-factor interacties meegenomen, namelijk die tussen de factoren x_1 en x_{37} , x_2 en x_{37} , x_3 en x_{37} , x_1 en x_{38} , x_2 en x_{38} , x_3 en x_{38} , x_2 en x_{31} t/m x_{36} , x_3 en x_{31} t/m x_{36} , x_2 en x_{44} , x_3 en x_{44} , x_5 en x_{44} , x_2 en x_{60} , x_3 en x_{60} , x_{37} en x_{60} , x_{38} en x_{60} en x_{44} en x_{60} .

De "defining relation" (Kleijnen, 1987, pag. 295-301) omvat 55 termen (zie Van Ham et al., 1990)

Tabel 4.5 Factoren met hun grenzen.

factor	minimum waarde (-)	basis waarde	maximum waarde (+)	eenheid
x_1	0.03	0.05	0.07	
x_2	0.018	0.020	0.022	
x_3	0.25	0.5	0.75	
x_4	0.03	0.05	0.07	
x_5	0.05	0.1	0.15	
x_6	0.1	0.2	0.3	
x_7 t/m x_{12}	250	500	750	jaar
x_{13}	5	10	15	jaar
x_{14}	25	50	75	jaar
x_{15}	20	40	60	jaar
x_{16}	15	25	35	jaar
x_{17} en x_{18}	25	50	75	jaar
x_{19}	0.5	1	1.5	jaar
x_{20} en x_{21}	1	2	3	jaar
x_{22}	0.5	1	1.5	jaar
x_{23} en x_{24}	1	2	3	jaar
x_{25}	0.2	0.4	0.6	
x_{26} en x_{27}	0.3	0.6	0.9	
x_{28}	0.1	0.2	0.3	
x_{29}	0.25	0.5	0.75	
x_{30}	0.3	0.6	0.9	
x_{31}	5	15	25	mha/jaar
x_{32}	1	2	3	mha/jaar
x_{33} en x_{34}	200	400	600	mha/jaar
x_{35} en x_{36}	0	0	1	mha/jaar
x_{37} en x_{38}	3	6	9	mha
x_{39}	0.5	1	1.5	jaar
x_{40}	1	2	3	jaar
x_{41} t/m x_{44}	0.5	1	1.5	jaar
x_{45} t/m x_{50}	5	10	15	jaar
x_{51}	25	30	35	jaar
x_{52}	30	60	90	jaar
x_{53} t/m x_{56}	25	50	75	jaar
x_{57} en x_{58}	5	10	15	jaar
x_{59} en x_{60}	0.5	1	1.5	jaar
x_{61}	5	10	15	jaar
x_{62}	1	2	3	jaar

4.3.3. Resultaten

De 62 hoofdeffecten en 26 interacties zijn in (2^{62-55}) 128 runs geschat. De resultaten staan in tabel 4.6. Deze tabel bevat alleen de significante factoren, aangeduid met een *. De volledige tabel staat in van Ham e.a. (1990).

Tabel 4.6 Factoren met geschatte effecten en t-waarden

factor	geschat effect	Student t_{39}
x_0	1333.08	477.21*
x_3	-100.02	-35.81*
x_5	-11.92	-4.27*
x_6	-11.78	-4.22*
x_{15}	-25.65	-9.18*
x_{16}	-7.95	-2.85*
x_{25}	7.27	2.60*
x_{27}	-38.88	-13.92*
x_{28}	-9.63	-3.45*
x_{30}	-6.11	-2.19*
x_{37}	-11.26	-4.03*
x_{38}	18.53	6.63*
x_{40}	-12.08	-4.33*
x_{52}	-6.89	-2.47*

Alle effecten hebben nu het juiste teken. Het model wordt op twee manieren gevalideerd, via de relatieve residuen van de reeds uitgevoerde runs, alsmede via 12 extra uitgevoerde runs (Van Ham et al., 1990). Korthedshalve worden hier slechts de resultaten vermeld. De relatieve residuen van de 128 uitgevoerde runs blijken tussen de 0.2 en de 7 % te liggen, terwijl de relatieve residuen van de 12 extra uitgevoerde runs tussen de 0.10 en de 5.5 % liggen.

4.3.4. Conclusie

De relatieve residuen blijken nergens groter dan 7 % te zijn. Dit model wordt derhalve goedgekeurd. Eén factor blijkt erg dominant te zijn, namelijk de biotische groeifactor (x_3). Voorts springen er nog drie factoren uit, nl. de humificatiefactor van grasland (x_{27}), de verblijftijd van humus in grasland (x_{15}), en de verschuiving van tropisch bos naar grasland (x_{38}). Er zijn bij dit metamodel geen significante interacties.

5. CONCLUSIES

Toepassing van metamodellen bij de analyse van een simulatiemodel vereist de nodige kennis omtrent het betreffende simulatiemodel alsmede inventiviteit bij de bepaling van de juiste proefopzet. In deze studie zijn voor verschillende modules van het simulatiemodel IMAGE adequate metamodellen opgesteld, met behulp waarvan deze modules zijn geanalyseerd. Tot dusver zijn weinig toepassingen van metamodellen van een dergelijke omvang bekend, afgezien van de zg. 'screening designs', die in de allereerste fase van de analyse kunnen worden gebruikt, als er nog honderden interessante factoren zijn, zie Bettonvil (1990).

Het specificeren van geldige metamodellen blijkt een tijdrovende bezigheid te zijn. Toch weegt de meerwaarde die een analyse met behulp van metamodellen biedt zeer zeker op tegen de benodigde hoeveelheid tijd en mankracht. Enkele verrassende conclusies komen aan het licht aan de hand van de door ons uitgevoerde gevoeligheidsanalyse. Zo blijkt de module betreffende de harde kustverdediging niet optimaal te functioneren. Na de metamodelanalyse is deze module opgesplitst in twee onafhankelijk opererende modules.

De analyse van de complexe koolstofkringloop module levert enkele (onvermoede) interessante kwadratische effecten als ook interacties op. Tevens blijkt de module voor de terrestrische biosfeer gedomineerd te worden door de biotische groeifactor, welke de CO_2 -bemesting aangeeft. Van de overige twaalf significante factoren zijn vooral belangrijk: de humificatiefactor van grasland, de verblijftijd van humus in grasland en de

verschuiving van tropisch bos naar landbouwgrond. Er zijn bij dit metamodel geen significante interacties gevonden.

De resultaten van deze analyse met behulp van metamodellen zijn van groot belang, omdat deze gebruikt gaan worden bij een te ontwikkelen interactieve versie van het simulatiemodel IMAGE. In algemene zin blijkt uit het onderzoek dat metamodellen en experimentele proefopzetten bruikbare technieken zijn om een gevoeligheidsanalyse uit te voeren op computersimulatiemodellen, ook wanneer deze complex van aard zijn. Dit bevestigt dat metamodellen onderschat worden als hulpmiddel bij de analyse van simulatiemodellen.

6. LITERATUUR

- Bettonvil, B. en Kleijnen J.P.C. (1988): 'Measurement Scales and Resolution IV Designs: a Note', verschijnt in the American Journal of Mathematical and Management Sciences, Catholic University Brabant, Tilburg.
- Elzen, M.G.J. den, and Rotmans, J. (1988). 'Simulatiemodel voor een aantal maatschappelijke gevolgen van het broeikaseffect voor Nederland, RIVM-rapport, rapport nr 758471008, Bilthoven
- Gezondheidsraad (1987): '2e Deeladvies inzake de CO₂-problematiek: Wetenschappelijke Inzichten en Maatschappelijke Gevolgen', Staatsdrukkerij, 's Gravenhage.
- Goudriaan J. en Ketner P. (1984): 'A Simulation Study for the Global Carbon Cycle, Including Man's Impact on the Biosphere', Climatic Change 6, 167-192.
- Ham, G.van, Rotmans, J., and Kleijnen, J.P.C. (1990): 'Een Gevoeligheidsanalyse via Metamodellen en Experimentele Proefopzetten, toegepast op een Simulatiemodel voor het Broeikaseffect', RIVM-rapport, rapport nr. 758471010, Bilthoven
- Kleijnen, J.P.C. (1987): 'Statistical Tools for Simulation Practitioners', Marcel Dekker, Inc., New York.
- Kleijnen, J.P.C. (1990): 'Statistics and Deterministic Simulation Models: Why Not?', KUBrabant, Report nr. FEW 435, Tilburg.
- Lammerts, I. (1989): 'Onzekerheidsanalyse toegepast op het broeikassimulatiemodel IMAGE', afstudeerverslag, Universiteit Utrecht

- Rotmans, J., Vrieze O.J., Peek, G.H.J.C., Veraart, W.N.G.M. (1988): 'Experimenteel Ontwerp ten behoeve van Metamodellering van een CO₂-Simulatie-Model', RIVM-rapport, rapport nr. 758471003, Bilthoven.
- Rotmans, J., Vrieze, O.J. (1989): 'Metamodelling and Experimental Design: Case Study of the Greenhouse Effect', RULimburg, Report M 88-03, Maastricht.
- Rotmans, J., de Boois, H., and Swart, R.J. (1990): 'An Integrated Model for the Assessment of the Greenhouse Effect: the Dutch Approach', Climatic Change 16, 331-356
- Sachs, J., Welch, W.J., Mitchell, T.J., and Wynn, H.P. (1989): 'Design and Analysis of Computer Experiments', Statistical Science 4, no.4, 409-435

Ontvangen: 11-12-1989

Geaccepteerd: 19-09-1990