

ASPECT-ANALYSE VAN CAUSALE MODELLEN
(PADCOEFFICIENTEN, SIMULTANE VERGELIJKINGEN)

Th. J. Ferrari en J. Mol¹⁾

Inleiding

Er bestaan een aantal methoden, die de mogelijkheid bieden om complexe relaties uit de economie, biologie, sociologie, psychologie enz. te beschrijven en te analyseren; veel gebruikte methoden zijn de aspect-analyse en de padcoefficienten-analyse of de analyse met de simultane vergelijkingen uit de economie. De modellen waarop deze analyses berusten en het gebruik zijn nogal verschillend (Ferrari, 1965). De laatstgenoemde methode wordt voornamelijk gebruikt om min of meer uitgewerkte hypothesen te toetsen en parameters te kwantificeren. De aspect-analyse kan voor dit doel ook, zij het in mindere mate, gebruikt worden; de analyse kan echter goed toegepast worden om hypothesen op te stellen of te verbeteren (Cattell, 1965).

Het model waarop de aspect-analyse is gebaseerd, is weinig gespecificeerd. Het systeem is verder ook niet geïdentificeerd, waardoor de oplossing niet eenduidig is en draaiingen van het assenstelsel mogelijk zijn (Ferrari, 1965). Dit is de oorzaak van vele moeilijkheden, die de interpretatie van de uitkomsten van deze analyses meestal oplevert. De onderzoeker doorziet moeilijk de betekenis van de verkregen resultaten (Liberg en Mol, 1961). De resultaten van de analyses met simultane vergelijkingen zijn uiteraard gemakkelijker te interpreteren.

In verband hiermede is een vergelijking van de uitkomsten van beide analysemethoden, toegepast op dezelfde gegevens, verhelderend. Mol (1966) liet reeds zien, wat het aantal exogene variabelen in een economisch model voor de uitkomsten van een aspect-analyse betekenen. In het volgende zullen wij het gedrag van berekende waarnemingen, die wij verkregen hebben door in een padcoefficienten-model één of meer primaire oorzaken (exogene of onafhankelijke variabelen) kunstmatig te veranderen, bestuderen en met aspect-analyses onderzoeken. De resultaten van deze analyses zullen wij vergelijken met de inhoud van het model waarvan is uitgegaan. Deze vergelijking zal dan de mogelijkheid geven een aantal regels ter formulering van meer gespecificeerde modellen en ter bestudering van de uitkomsten (draaiingen) op te stellen.

De gegevens worden verkregen uit een reeds eerder beschreven model over de invloed van een aantal factoren op het MgO-gehalte van het gras (Ferrari, 1964). Dit model is in fig. 1 weergegeven. Men ziet hierin hoe de variabelen humusgehalte (x_1),

1) Resp. Instituut voor Bodemvruchtbaarheid, Prof. van Hallstraat 3 Groningen en Instituut voor Economisch Onderzoek der Universiteit, Oude Boteringestraat 44, Groningen.

MgO-gehalte (x_3), K_2O -gehalte (x_2) en pH (x_4) van de grond het MgO-gehalte (y_1) van het gras rechtstreeks en via veranderingen in het percentage kruiden (y_3) en in het eiwitgehalte (y_2) van het gras beïnvloeden. Dit model kan met de volgende structuurvergelijkingen worden weergegeven¹⁾.

$$(1) y_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4$$

$$(2) y_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + b_{23}y_3$$

$$(3) y_1 = a_{12}x_2 + a_{13}x_3 + b_{12}y_2 + b_{13}y_3$$

In deze vergelijkingen zijn x_i de primaire oorzaken of exogene variabelen uit de economie. De veranderingen in de effecten y_i onder invloed van veranderingen van deze onafhankelijke variabelen kunnen uit de vergelijkingen in de gereduceerde vorm berekend worden; hierin is telkens een effect als een functie van alleen onafhankelijke variabelen uitgedrukt. Voor de berekeningen gebruikten wij de parameter-waarden uit tabel 1, zoals zijn in een eerder onderzoek gevonden waren (Ferrari, 1963).

Tabel 1. Waarden van de parameters uit de vergelijkingen 1, 2 en 3

$a_{31} = 1,67$	$a_{21} = -0,74$	$a_{12} = -0,0038$
$a_{32} = -0,23$	$a_{22} = 0,11$	$a_{13} = 0,0004$
$a_{33} = -0,031$	$a_{23} = 0,011$	$b_{13} = 0,0041$
$a_{34} = 5,26$	$b_{23} = 0,20$	$b_{12} = 0,0083$

Vervolgens onderzoeken wij op welke wijze de effecten in dit model veranderen, wanneer één oorzaak (x_3) veranderd wordt en de andere oorzaken constant blijven. Wij gaan dit ook na bij veranderingen van twee onafhankelijke variabelen (x_3 en x_2) met constante waarden van x_1 en x_4 . Tenslotte zullen wij nagaan, wat veranderingen in drie en meer onafhankelijke variabelen betekenen. Wij simuleren op deze wijze de reactie van de natuur op bepaalde veranderingen. Op de zo verkregen gegevens zullen aspect-analyses toegepast worden, waarvan de resultaten met de inhoud van het uitgangsmodel vergeleken worden. Uit deze vergelijkingen zullen enkele conclusies getrokken kunnen worden.

Reactie van de afhankelijke variabelen op veranderingen van een of meer onafhankelijke variabelen

Wij geven de variabele x_3 een aantal (20) waarden, die volgens toeval gekozen zijn en binnen het variatiegebied van het oorspronkelijke onderzoek liggen. De waarden van y_3 , y_2 en y_1 , die bij deze 20 waarden van x_3 horen, moeten vervolgens

1) In deze en volgende vergelijkingen wordt het niveau (constante term) buiten beschouwing gelaten.

berekend worden. Daar wij in dit geval alleen met veranderingen in x_3 te maken hebben, is een systeem van vereenvoudigde vergelijkingen nodig, n.l.:

$$(4) y_3 = a_{33}x_3$$

$$(5) y_2 = a_{23}x_3 + b_{23}y_3$$

$$(6) y_1 = a_{13}x_3 + b_{12}y_2 + b_{13}y_3$$

Hierin heeft vergelijking 4 reeds de gereduceerde vorm. Deze is voor de andere vergelijkingen als volgt:

$$(5a) y_2 = (a_{23} + a_{33}b_{23}) x_3$$

$$(6a) y_1 = (a_{13} + a_{23}b_{12} + a_{33}b_{12} + a_{33}b_{13}) x_3.$$

Met de aanname dat de natuur volgens de parameterwaarden uit tabel 1 zal reageren, worden deze vergelijkingen:

$$(4b) y_3 = -0,031 x_3$$

$$(5b) y_2 = 0,0048 x_3$$

$$(6b) y_1 = 0,0003127 x_3$$

Wij kunnen nu met behulp van deze vergelijkingen de waarden van y_3 , y_2 en y_1 bij de 20 verschillende waarden van x_3 uitrekenen. Tabel 2 geeft enkele uitkomsten.

Tabel 2. De met de vergelijkingen 4b, 5b en 6b berekende waarden van y_3 , y_2 en y_1 bij enkele waarden van x_3

waarneming	1	2	3	-----	20
variabelen					
x_3	167	122	50	-----	185
y_3	-5,18	-3,78	-1,55	-----	-5,74
y_2	0,80	0,59	0,24	-----	0,88
y_1	0,052	0,038	0,016	-----	0,058

Een onderzoeker zal over het algemeen onbekend met de achter deze getallen liggende samenhang zijn. Het is dan zijn opgave om deze samenhang te achterhalen. Wij echter kennen deze samenhang. Een eenvoudige bewerking leert ons dan, dat de getallen in de tweede rij $-0,031$ maal de overeenkomstige getallen uit de eerste rij zijn; de getallen in de rij van y_2 zijn weer $-0,155$ maal de getallen uit de tweede rij enz. De rijen (en kolommen) uit deze tabel zijn dus van elkaar afhankelijk; indien de getallen uit een rij gegeven zijn, kunnen de getallen in de andere rijen met behulp van deze verhoudingsgetallen berekend worden. Dit betekent wiskundig, dat de rang van deze matrix van getallen gelijk aan 1 is.

Voor een goed begrip van de resultaten van een aspectanalyse is de meetkundige weergave van het begrip rang van belang. Uitgangspunt is hierbij de weergave van de correlaties tussen de 4 variabelen in een 20-dimensionaal loodrecht assenstelsel, waarin de 20 waarnemingen als assen fungeren.

De variabelen zelf worden hierin als vectoren afgebeeld. In fig. 2 zijn de assen van alleen 3 waarnemingen uit tabel 2 getekend; in werkelijkheid liggen de 4 vectoren in een 20-dimensionale ruimte. Het blijkt nu, dat de 4 vectoren in dit geval steeds dezelfde of de tegengestelde richting hebben; de richtingen vallen samen. Dit betekent, dat de correlaties absoluut zijn en + 1 of - 1 zijn. De ligging van deze 4 vectoren kan dus in een een-dimensionale ruimte worden beschreven. Deze een-dimensionale vectorenconfiguratie komt overeen met de een-dimensionale stippenfiguur in een normale correlatie-weergave, waarin de variabelen als assen figuren.

Het feit, dat de variatie in de drie afhankelijke variabelen door één onafhankelijke variabele veroorzaakt wordt, heeft dus tot gevolg gehad dat de rang van de matrix van gegevens 1 en de vectorenconfiguratie een-dimensionaal is. De ruimte van het gemeenschappelijk bewegingspatroon (Mol, 1964) heeft een dimensie van 1. Hierbij is op te merken, dat het feitelijk gedrag van de variabelen met 3 vergelijkingen moet worden weergegeven. De ruimte van de vectorenconfiguratie geeft informatie over dit aantal vergelijkingen, niet over de aard van deze vergelijkingen. De onderzoeker, die met een aspect-analyse dit aantal heeft bepaald, heeft de vrijheid verschillende interpretaties aan dit bewegingspatroon te geven.

Wij laten vervolgens de variabelen x_3 en x_2 uit het model onafhankelijk van elkaar veranderen. De variabelen x_1 en x_4 blijven dan constant en kunnen in de vergelijkingen 1, 2 en 3 buiten beschouwing gelaten worden. De gereduceerde vergelijkingen zijn als volgt:

$$(7a) y_3 = a_{32}x_2 + a_{33}x_3$$

$$(8a) y_2 = (a_{22} + a_{32}b_{23})x_2 + (a_{23} + a_{33}b_{23})x_3$$

$$(9a) y_1 = (a_{12} + a_{22}b_{12} + a_{32}b_{23}b_{12} + a_{32}b_{13})x_2 + (a_{13} + a_{23}b_{12} + a_{33}b_{23}b_{12} + a_{33}b_{13})x_3$$

Wij nemen weer aan dat het model reageert volgens de gegevens uit tabel 2 en krijgen de volgende vergelijkingen:

$$(7b) y_3 = -0,230 x_2 - 0,031 x_3$$

$$(8b) y_2 = 0,0640 x_2 + 0,0048 x_3$$

$$(9b) y_1 = -0,00421 x_2 + 0,00031 x_3$$

Met behulp van deze vergelijkingen worden de waarden van y_3 , y_2 en y_1 voor 20 waarden van x_3 en van x_2 uitgerekend. Tabel 3 geeft enkele uitkomsten.

Tabel 3. De met de vergelijkingen 7b, 8b en 9b berekende waarden van y_3 , y_2 en y_1 bij enkele waarden van x_3 en x_2

waarneming	1	2	3	-----	20
variabelen					
x_3	167	122	50	-----	185
x_2	21	15	21	-----	6
y_3	-10,01	-7,23	-6,38	-----	-7,12
y_2	2,15	1,55	1,58		1,27
y_1	-0,037	-0,025	-0,073	-----	0,032

Het feit dat deze gegevens zijn ontstaan door veranderingen in 2 onafhankelijke variabelen, maakt dat de rang van de matrix van getallen 2 is. Men kan de getallen van elke rij berekenen indien de getallen van 2 rijen gegeven zijn. De correlaties tussen de getallen in de rijen (variabelen) kunnen nu ook met de hierboven beschreven methode bestudeerd worden. Deze samenhangen zijn in figuur 3 getekend, waarin de drie waarnemingsassen om tekentechnische redenen weggelaten zijn. De vectoren (variabelen) blijken nu niet in een een-dimensionale ruimte te liggen, maar in een twee-dimensionaal vlak uit te waaiëren. De variabelen x_3 en x_2 zijn niet geïmpliceerd en de richtingen van deze vectoren zijn dan ook onderling loodrecht. De variabele y_2 is met de variabelen x_2 en x_3 positief gecorreleerd, y_3 daarentegen negatief. Tenslotte is y_1 negatief met x_2 en positief met x_3 gecorreleerd.

Terwijl de samenhangen tussen de variabelen (bewegingspatroon) uit het eerste voorbeeld met 1 onafhankelijke variabele met een een-dimensionale ruimte beschreven konden worden, hebben wij nu een twee-dimensionaal vlak nodig om de samenhangen onder invloed van 2 onafhankelijke variabelen te beschrijven. De rang 2 komt met een twee-dimensionaal bewegingspatroon overeen. De onderzoeker, onbekend met het ontstaansmodel, kan nu de conclusie trekken, dat het aantal onafhankelijke variabelen 2 is geweest en dat het aantal vergelijkingen van het model gelijk aan het totaal aantal variabelen minus het aantal onafhankelijke variabelen of dimensies van de vectorenconfiguratie (rang) is geweest. Ook nu weer geldt dat hij verschillende interpretaties aan deze vergelijkingen en samenhangen kan geven.

Het is in analogie met het bovenstaande gemakkelijk in te zien, dat men een drie-dimensionaal bewegingspatroon krijgt bij veranderingen van 3 onafhankelijke variabelen enz. Men kan daarom de volgende regel opstellen: het aantal vergelijkingen in een systeem is gelijk aan het aantal variabelen minus de dimensie van de vectorenconfiguratie; deze dimensie is gelijk aan het aantal onafhankelijke variabelen.

Met behulp van deze regel kan een onderzoeker nagaan of het door hem opgestelde model, dat hij wil toetsen, in overeenstemming met de werkelijkheid is; van de andere kant kan hij de kennis over de dimensie van het bewegingspatroon en over de variabelen die hierin mee variëren gebruiken om een meer gespecificeerd model als hypothese op te stellen. Deze regel kan ook gebruikt worden voor een betere interpretatie van de uitkomsten van de aspect-analyse.

Aspect-analyse en draaiing

De aspectanalyse beoogt het aantal dimensies van de vectorenconfiguratie vast te stellen en aan te geven in welke mate de variabelen in een bepaald bewegingspatroon met elkaar samenhangen. Door de analyse wordt de vectorenconfiguratie als een projectie op een loodrecht assenstelsel (referentiestelsel), met dezelfde oorsprong als de vectoren, weergegeven. De ligging van dit referentiestelsel ten opzichte van de vectorenconfiguratie is van de analyse-methode^{x)} afhankelijk (Harman, 1960, Seal 1964). Het model van de aspect-analyse is niet-geïdentificeerd (Ferrari, 1964), zodat de verkregen oplossing niet principieel is. Door draaiing van het referentiestelsel is elke oplossing te verkrijgen, waarbij de onderlinge

x) Wij hebben steeds de principal factor methode van Hotelling toegepast.

richting der vectoren en de dimensie van het bewegingspatroon bestaan blijven. Deze onderidentificatie en de daarmee samenhangende oplossingsmogelijkheden is oorzaak van vele interpretatiemoeilijkheden. Een vergelijking tussen de inhoud van het gegeven model en de resultaten van een aspect-analyse toegepast op uit dit model berekende gegevens wijst op de mogelijkheid om een identificatie door een bepaalde draaiing te verkrijgen. Deze vergelijking is tevens nuttig om de mogelijkheden en beperkingen van de aspect-analyse aan te geven.

Wij beginnen met een bewerking van de gegevens, die verkregen zijn door veranderingen van 2 onafhankelijke variabelen x_2 en x_3 (tabel 3); de veranderingen van x_2 en x_3 zelf zijn onafhankelijk van elkaar. De onderzoeker wordt verondersteld onbekend met deze achtergrond te zijn en tracht met een aspect-analyse informatie over het ontstaan van de samenhang tussen de variabelen te krijgen. Hiertoe worden de correlatie-coëfficiënten tussen de variabelen berekend (tabel 4).

Tabel 4. Correlatiecoëfficiënten, berekend uit de volledige tabel 3

	x_2	x_3	y_3	y_2	y_1
x_2	1,00	0	-0,74	+0,89	-0,89
x_3		1,00	-0,68	0,47	0,44
y_3			1,00	-0,96	0,35
y_2				1,00	-0,59
y_1					1,00

Deze correlatiematrix vraagt weinig commentaar, de achtergrond is ons immers bekend; het ontbreken van correlaties tussen de onafhankelijke variabelen x_2 en x_3 was te verwachten, wij hadden ze immers onafhankelijk van elkaar laten variëren. Wij zijn meer geïnteresseerd in de resultaten van een op deze matrix toegepaste aspect-analyse (tabel 5).

Tabel 5. Correlatiecoëfficiënten a_{im} tussen aspect en variabelen.

Aspect	F_1	F_2	Som a_{im}^2
Variabele			
x_2	0,94	-0,35	1,01
x_3	0,35	0,94	1,01
y_3	-0,92	-0,38	1,00
y_2	0,99	0,12	0,99
y_1	-0,68	0,73	1,00
Aspectbijdrage (eigenwaarde)	3,30	1,70	5,00

De coëfficiënten a_{im} zijn een maat voor de sterkte van de correlatie tussen aspect en variabelen en lopen van + 1,00 tot - 1,00. Het kwadraat van een coëfficiënt maal honderd geeft aan welk gedeelte van de totale variantie van de variabele met het betreffende aspect samenhangt; de som van deze kwadraten is daarom 1 (totale variantie).

De oorspronkelijke correlatie-coëfficiënt tussen 2 variabelen kan uit het inwendig product der betreffende rijen terug berekend worden, dus de correlatiecoëfficiënt r_{23} tussen de variabelen x_2 en x_3 is gelijk aan: $0,94 \times 0,35 + (-0,35 \times 0,94) = 0$ enz. De eigenwaarde tenslotte is een maat voor de totale variantie die met een aspect samenhangt.

Een vergelijking tussen het model van figuur 1 zonder de variabelen x_1 en x_4 en het model van de aspect-analyse (figuur 4) geeft aan, dat men uit de aspect-analyse nooit de oorspronkelijke informatie terug verwachten mag. Het oorspronkelijke model is gekenmerkt door kettingreacties; deze ontbreken in het model van de aspect analyse. Het model met de padcoëfficiënten bevat informatie over de sterkte van de afzonderlijke invloeden, de mogelijkheid om deze terug te berekenen ontbreekt in het model van de aspect-analyse. De aspect-analyse geeft wel informatie over het gemeenschappelijk variëren van de variabelen, verdeeld over de verschillende aspecten. Men krijgt hiervoor een maat in percentages van de totale variantie: 88 % van de variantie van x_2 gaat samen met 12 % van de variantie van x_3 , met 85 % van de variantie in y_3 enz. Hetzelfde geldt voor de gegevens in aspect 2; hierin wordt de van aspect 1 onafhankelijke variantie van de variabelen beschreven.

De gegevens van tabel 5 laten zien, dat de variantie van alle variabelen volledig met de beide aspecten beschreven wordt; de som van de kwadraten is steeds gelijk aan 1. Dit is geheel volgens de verwachtingen, volgend uit de eerder gegeven regel: 2 onafhankelijke variabelen moeten 2 aspecten of bewegingspatronen geven.

Niet volgens de verwachting is de aanwezigheid van correlaties tussen een aspect en de onafhankelijke variabelen x_2 en x_3 . Wij hebben immers deze variabelen onafhankelijk van elkaar laten variëren, hetgeen in correlatiecoëfficiënten tussen beide variabelen gelijk aan nul tot uiting komt. Men mag daarom voor een causale interpretatie niet de situatie hebben, dat beide variabelen binnen een aspect toch tegelijk variëren. De oorzaak is in het feit gelegen, dat de principal factor methode elke as van het referentiestelsel zó legt, dat de totale met deze as samenhangende variatie steeds zo groot mogelijk is. Men krijgt dan een oplossing van figuur 5, waarin de referentie-assen niet met de x -vectoren uit figuur 3 samenvallen. Een x -vector moet dan met 2 referentie-assen beschreven worden.

Dit probleem kan opgelost worden door toepassing van een draaiing; wij hebben immers gezien dat een draaiing steeds geoorloofd is. Men denkt hierbij direct aan een zodanige draaiing, dat de assen en de x -vectoren in dit geval samenvallen. Hierdoor verdwijnen de correlaties tussen aspect en de niet relevante x -variabelen. Men bereikte hierdoor, dat de aspecten volledig geïdentificeerd worden. Men heeft dan een eenduidige oplossing verkregen, waarvan de causale interpretatie geen moeilijkheden oplevert. Het resultaat van deze draaiing ziet men in tabel 6.

Tabel 6. Correlatiecoëfficiënten a_{im} tussen aspect en variabelen, na draaiing

Aspect	F_1	F_2	Som a_{im}^2
Variabele			
x_2	1,00	0	1,00
x_3	0	1,00	1,00
y_3	-0,73	-0,68	1,00
y_2	0,89	0,46	1,00
y_1	-0,89	0,44	0,99

Een causale interpretatie van de na deze draaiing verkregen resultaten is nu gemakkelijk te geven, zij het steeds met de reeds eerder besproken beperkingen. De betekenis van deze draaiing komt nog duidelijker tot uiting, indien men met meer-dimensionale assenstelsels te maken heeft. Wij zullen dit demonstreren met een voorbeeld, waarin 4 onafhankelijke variabelen x_1, x_2, x_3 en x_4 opgenomen waren en het assenstelsel vier-dimensionaal is. De aspect-analyse leverde de gegevens van tabel 7 op.

Tabel 7. Correlatiecoëfficiënten a_{im} tussen aspect en variabelen; gegevens berekend uit de vergelijkingen 1, 2 en 3.

Aspect	F_1	F_2	F_3	F_4	Som a_{im}^2
Variabele					
x_1	0,83	-0,39	-0,04	0,40	1,00
x_2	-0,54	-0,61	0,18	0,55	1,00
x_3	-0,07	0,62	-0,30	0,72	1,00
x_4	0,12	0,31	0,93	0,14	1,00
y_3	0,93	-0,23	0,29	0,02	1,00
y_2	-0,91	0,31	0,26	0,01	0,99
y_1	0,68	0,73	-0,04	0,01	1,00
Aspectbijdrage (eigenwaarde)	3,15	1,68	1,15	1,01	6,99

Het resultaat van deze analyse is teleurstellend. Weliswaar be- draagt het aantal aspecten volgens verwachting 4 - het aantal onafhan- kelijke variabelen bedroeg immers 4- de causale achtergrond van de vergelijkingen 1, 2 en 3 daarentegen is grotendeels verdwenen. Het is moeilijk zelfs met onze voorkennis enkele aspecten causaal te duiden. Een relatief groot gedeelte van de variantie is in aspect 1 terecht gekomen. Elk aspect heeft betrekkelijk hoge correlaties met meer dan een x-variabele. Identificatie en daarmee de definitieve oplossing worden weer verkregen door toepassing van een zodanige draaiing, dat de referentie-assen met de x-vectoren samenvallen. Men krijgt een oplossing die in overeenstemming met de uitgangspunten is. Tabel 8 geeft het resultaat van deze draaiing.

Tabel 8. Correlatiecoëfficiënten a_{im} tussen aspect en variabelen, na draaiing; gegevens berekend uit de vergelijkingen 1, 2 en 3

Aspect Variabele	F_1	F_2	F_3	F_4	Som a_{im}^2
x_1	1,00	0	0	0	1,00
x_2	0	1,00	0	0	1,00
x_3	0	0	0	1,00	1,00
x_4	0	0	1,00	0	1,00
y_3	0,86	-0,30	0,32	-0,28	1,01
y_2	-0,88	0,36	0,23	0,19	0,99
y_1	0,29	-0,81	0,27	0,43	1,00

De bruikbaarheid van de uitgevoerde draaiing en de juistheid van het draaiingsprincipe komen uit dit resultaat duidelijk naar voren. De draaiing heeft een identificatie van de aspecten gegeven, terwijl de interpretatie is vereenvoudigd. Een landbouwkundige beschrijving van de resultaten valt buiten de opzet van dit artikel.

De verkregen resultaten hebben de betekenis van de volgende gedragslijn getoond. Ten behoeve van de causale interpretatie moet op de met de aspect-analyse verkregen gegevens steeds een draaiing worden toegepast. Het criterium bij deze draaiing moet bij de onafhankelijke variabelen liggen, waarvan het aantal gelijk is aan het aantal gevonden aspecten. Men tracht elke referentie-as zo goed mogelijk met een onafhankelijke variabele te laten samenvallen; hierbij moeten de correlaties tussen het aspect en de andere onafhankelijke variabelen zo klein mogelijk gehouden worden. Het effect van de hieruit resulterende draaiing wordt tenslotte voor alle variabelen berekend. Deze draaiing, waarvoor wij de naam caumax-draaiing voorstellen, beoogt de aspecten te identificeren en de mogelijkheid tot causale interpretatie maximaal te maken.

Discussie

De aspect-analyse biedt mogelijkheden, maar heeft ook een aantal beperkingen waarvan men zich steeds bewust moet blijven. De analyse heeft het nadeel, dat het model weinig gespecificeerd en niet-geïdentificeerd is. In dit opzicht is zij steeds de mindere van een meer gespecificeerd causaal model. De reeds gemaakte vergelijking tussen het model van figuur 1 en dat van figuur 4 geeft een duidelijk beeld van de verschillen en van wat men van een aspect-analyse mag verwachten. De analyse biedt aan de andere kant de onderzoeker de mogelijkheid niet-gespecificeerde samenhang tussen vele variabelen tegelijk te bestuderen, waardoor hij informatie over relevante variabelen en relaties krijgt. Hierdoor bestaat de mogelijkheid meer gespecificeerde modellen op te stellen. Het hangt van onderzoeker en van probleem af in welke mate de analyse bruikbare resultaten kan opleveren.

In dit verband is de beantwoording van de vraag belangrijk of de resultaten van een aspect-analyse, al of niet met draaiing, sterk door storende omstandigheden beïnvloed worden. Men kan hierbij denken aan de betekenis van bepalingfouten, van correlaties tussen de onafhankelijke variabelen enz. Andere moeilijkheden hebben betrekking op de vaststelling van het aantal aspecten en de feitelijke

uitvoering van de caumax-draaiing. Wij zullen in het volgende enkele van deze punten behandelen; wij verwijzen voor het probleem van de vaststelling van het aantal aspecten naar de betreffende literatuur (Harman, 1960).

Wij hebben gezien, dat de caumax-draaiing de onafhankelijke variabelen als criteria neemt; men neemt zoveel variabelen als er aspecten zijn. In de literatuur zijn een aantal draaiingsmethoden bekend die geen onderscheid tussen onafhankelijke en afhankelijke variabelen maken; de meest bekende is de draaiing tot de eenvoudige structuur. Een keuze over de als criteria te nemen variabelen behoeft men hierbij niet te maken; alle variabelen worden opgenomen en de draaiing kan zonder tussenkomst van de onderzoeker uitgevoerd worden. De caumax-draaiing vraagt daarentegen van de onderzoeker een stellingname over de te verwachten en de feitelijke inhoud van de aspecten, over de onafhankelijke variabelen die in de analyse opgenomen moeten worden en die als criteria van de draaiing moeten dienen. Een automatische oplossing wordt hier niet verkregen. Is de keuze van de variabelen, waarop gedraaid moet worden, eenmaal gedaan, dan is de draaiing zelf grafisch gemakkelijk uit te voeren.

Wij hebben enkele berekeningen over de invloed van fouten (standaardafwijkingen ongeveer 15 %) en van correlaties tussen de onafhankelijke variabelen uitgevoerd. Hieruit bleek, dat deze storingen een geringe invloed hadden en dat de aspect-analyse tamelijk consistente resultaten oplevert. Ter illustratie geven wij het resultaat van een aspect-analyse (na draaiing) van gegevens van het model uit figuur 1, waarin de onafhankelijke variabelen betrekkelijk sterke correlaties (0,20, -0,45, 0,64, -0,53, 0,20 en -0,27) hadden. Het resultaat wordt in tabel 9 gegeven.

Tabel 9. Correlatiecoëfficiënten a_{im} tussen aspect en variabelen, na draaiing; model met correlaties tussen de onafhankelijke variabelen

Aspect	F_1	F_2	F_3	F_4	Som a_{im}^2
Variabele					
x_1	0,89	-0,03	0,45	-0,13	1,01
x_2	0,11	0,95	0,23	-0,20	1,00
x_3	-0,27	-0,27	-0,03	0,95	1,01
x_4	0,22	-0,07	0,97	-0,02	1,00
y_3	0,74	-0,10	0,59	-0,28	1,00
y_2	-0,90	0,32	-0,12	0,28	1,00
y_1	0,09	-0,85	0,11	0,51	1,01

Een vergelijking tussen deze resultaten en die van tabel 8 toont aan, dat de aanwezigheid van betrekkelijk sterke correlaties tussen de onafhankelijke variabelen de inhoud van de aspecten niet wezenlijk veranderd heeft.

Samenvatting

Een aantal waarnemingen werden kunstmatig verkregen door de onafhankelijke variabelen in een model over de invloed van een aan aantal bodemfactoren op het MgO-gehalte van gras verschillende waarden te laten aannemen. Met behulp van dit model konden de waarden van de afhankelijke variabelen berekend worden. Vervolgens werden op deze gegevens aspect-analyses uitgevoerd, waarvan de resultaten vergeleken werden met de inhoud van het model.

Het blijkt, dat het aantal aspecten gelijk is aan het aantal onafhankelijke variabelen. Zonder toepassing van een draaiing kunnen de resultaten moeilijk geïnterpreteerd worden. Deze draaiing moet volgens het caumax-principe uitgevoerd worden: de mogelijkheid voor een causale interpretatie wordt zo groot mogelijk gemaakt door uitvoering van een draaiing, waarbij elke referentie-as zoveel mogelijk op een onafhankelijke variabele gedraaid wordt. Men bereikt hiermede een identificatie van het model van de aspect-analyse. De uitkomsten worden slechts in geringe mate door het optreden van waarnemingsfouten en door de aanwezigheid van correlaties tussen de onafhankelijke variabelen beïnvloed.

Literatuur

- Cattell, R. B. Factor analysis: an introduction to essentials. *Biometrics* 21 (1965) 190-215, 405-435.
- Ferrari, Th. J. Factor analysis in agricultural research. Pijl, H. and Venekamp, J.T.N. *Neth. J. Agric. Sci.* 5 (1957) 211-221.
- Ferrari, Th. J. Causal soil-plant relationships and path coefficients. *Plant and Soil* 19 (1963) 81-96.
- Ferrari, Th. J. Auswertung biologischer Kettenprozesse mit Hilfe von Pfadkoeffizienten. *Biometrische Zeitschrift* 6 (1964) 89-102.
- Ferrari, Th. J. Models and their testing: considerations on the methodology of agricultural research. *Neth. J. Agric. Sci.* 13 (1965) 366-377.
- Harman, H. H. Modern factor analysis. Chicago (1963).
- Liberg, A. H. J. and Mol, J. Multiple factor analysis as a method of agricultural research. *Neth. J. Agric. Sci.* 9 (1961) 269-280.
- Mol, J. Modèles d'exploitations agricoles. C. E. E. Collection études. Serie Agriculture Nr. 13. Bruxelles (1964).
- Mol, J. De economische onderzoeker en zijn ruimte. *Economie* 30 (1966) 361-376.
- Seal, H. Multivariate statistical analysis for biologists. London (1964).

Groningen, juni 1966

- Figuur 1. Causaal model van de invloed van 4 onafhankelijke variabelen (primaire oorzaken) op 3 afhankelijke variabelen (effecten)
- Figuur 2. Vectorenconfiguratie bij 1 onafhankelijke variabele
- Figuur 3. Vectorenconfiguratie bij 2 onafhankelijke variabelen
- Figuur 4. Aspectanalyse-model met 3 afhankelijke en 2 onafhankelijke variabelen
- Figuur 5. Ligging van referentie-assen en vectorenconfiguratie, door analyse gegeven.

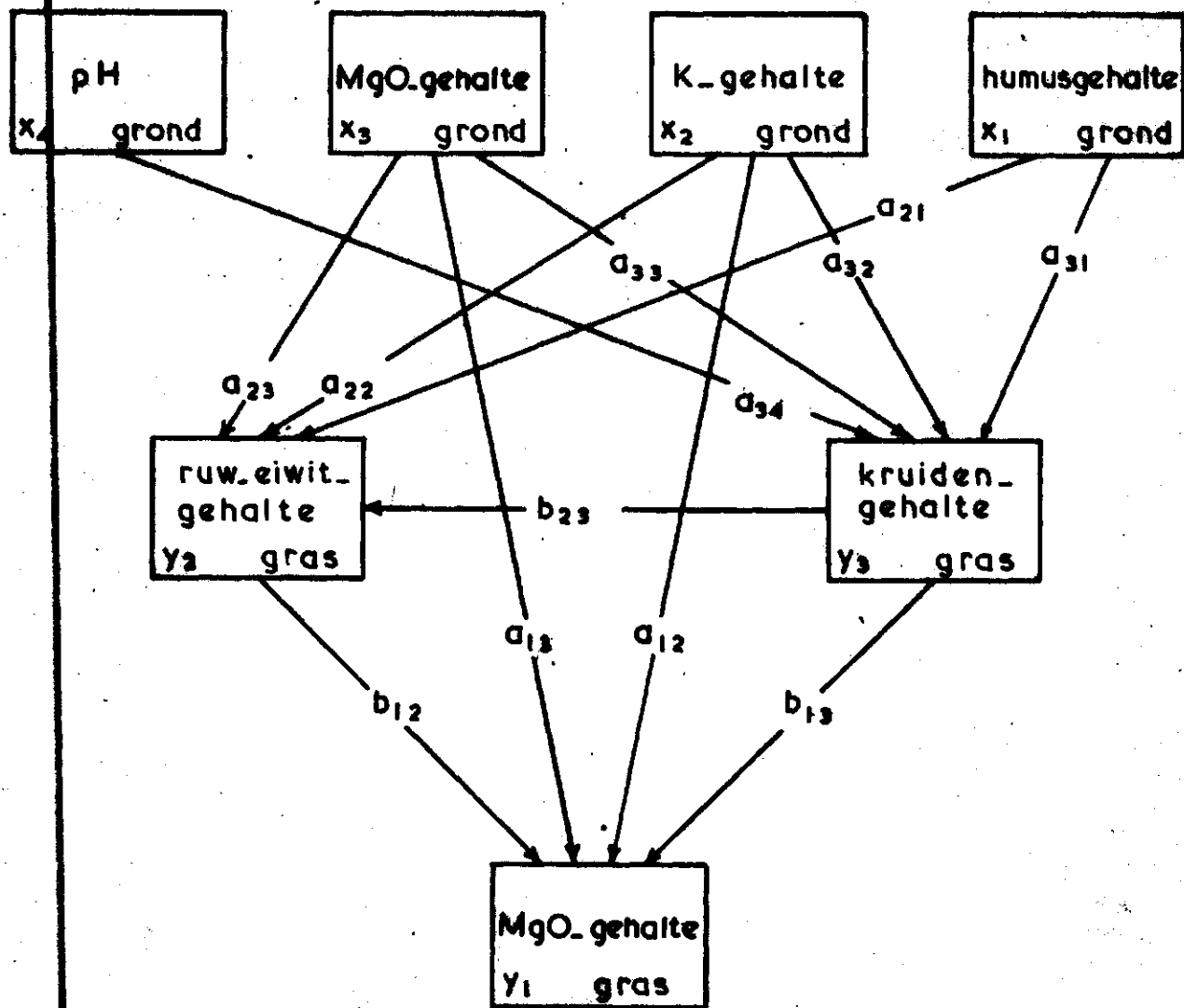


fig. 1

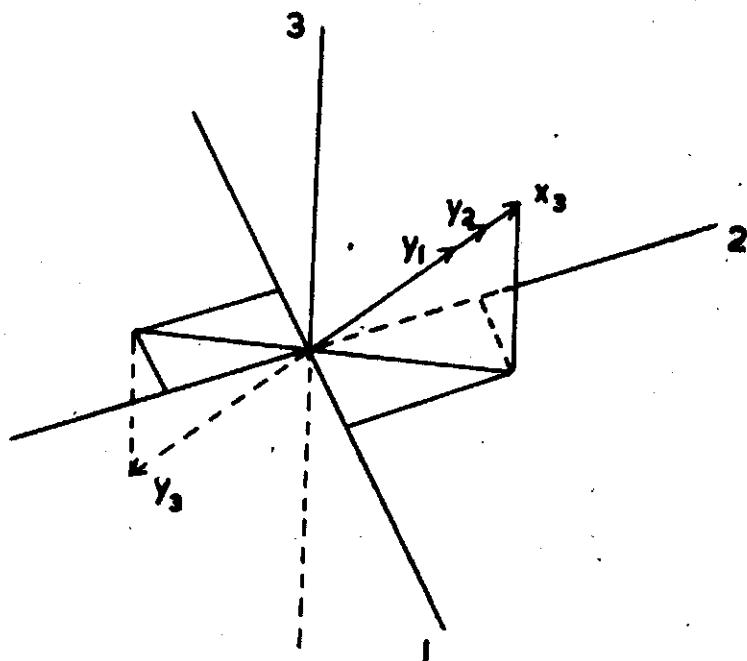


fig. 2

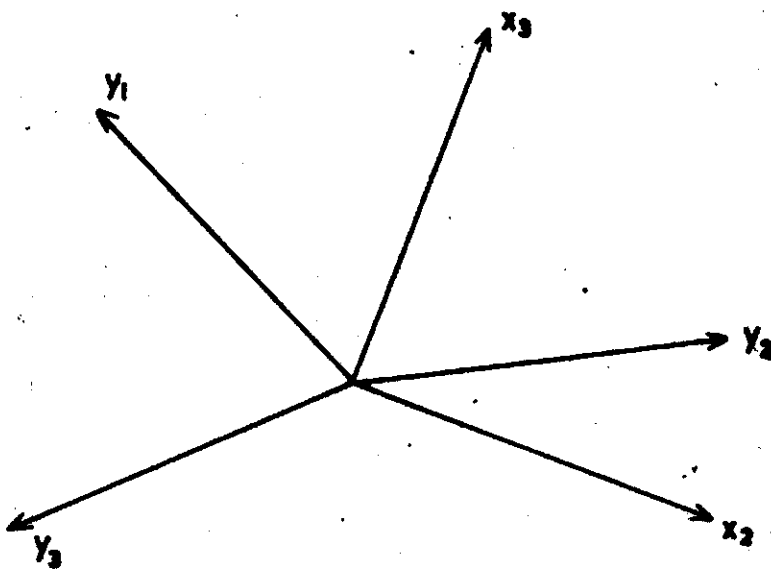


fig. 3

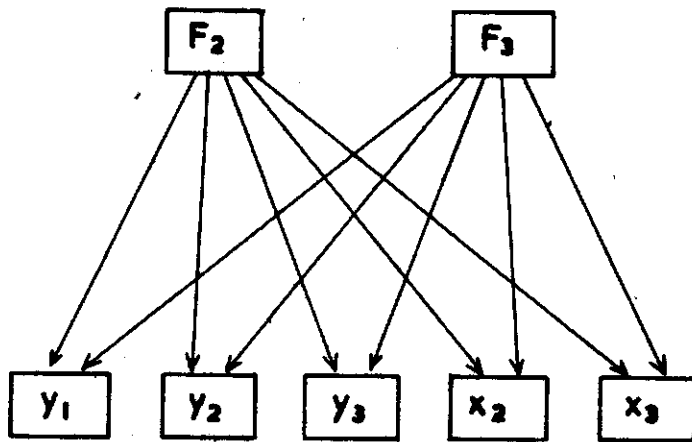


fig. 4

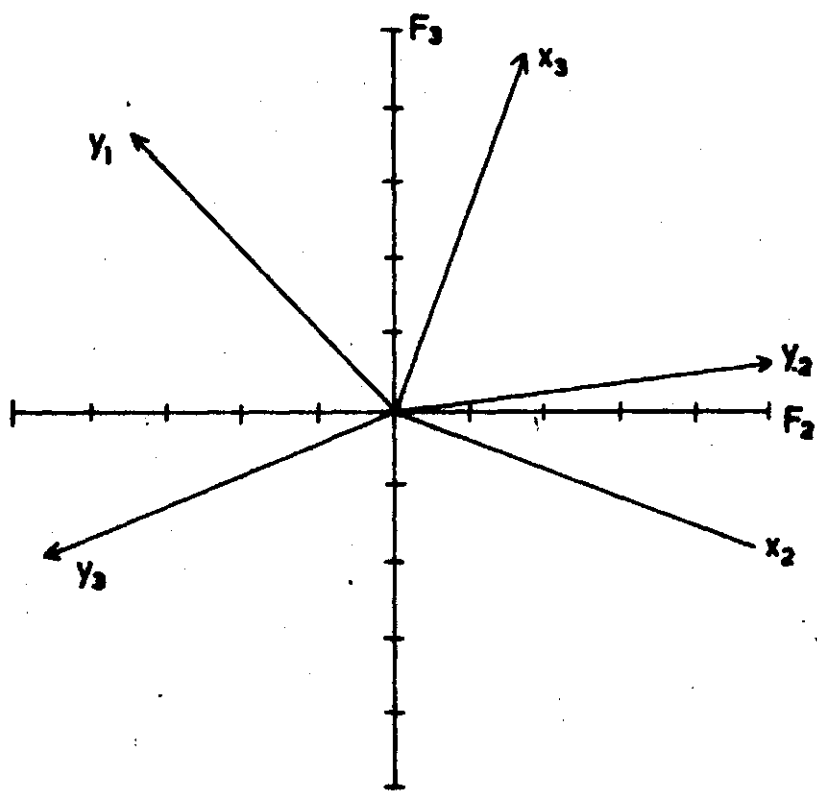


fig. 5