

Overdruk uit het Landbouwkundig Tijdschrift
74ste jaargang no. 17, oktober 1962

Methode van de pad-coëfficiënten in het landbouw-ecologische onderzoek

Methode der Pfad-Koeffizienten in ökologischen Untersuchungen
Zusammenfassung siehe Seite 735

TH. J. FERRARI met medewerking van G. J. J. TARIS,
Instituut voor Bodemvruchtbaarheid, Groningen

INLEIDING

In het artikel over de vergelijking tussen proeven met en zonder ingreep (Ferrari, 1960) hebben wij de methode met de zg. pad-coëfficiënten genoemd om bepaalde bezwaren te kunnen ondervangen, die in het bijzonder bij bewerking van gegevens van proeven zonder ingreep optreden. Een van de bezwaren hangt samen met de onvolkomenheid van het gebruikte regressiemodel. In het normale model wordt namelijk de veronderstelling gemaakt, dat de zg. onafhankelijke factoren elkaar niet beïnvloeden. Dit betekent, dat een verandering van een factor geen verandering van de andere onafhankelijke factoren tot gevolg heeft. Deze veronderstelling is in vele gevallen niet in overeenstemming met de werkelijkheid.

Een van de voorbeelden, waarbij het gebruik van het regressiemodel niet juist is, geeft het onderzoek van Sluijsmans (1962) over de samenhang tussen het MgO- en K₂O-gehalte van het gras enerzijds en bodemkundige en andere factoren anderzijds. De gegevens verkregen bij een proef zonder ingreep, werden onderworpen aan een regressie-analyse, waarbij het MgO- of het K₂O-gehalte gras als de afhankelijke variabelen, het MgO- en het K₂O-gehalte van de grond, het humusgehalte, het percentage kruiden en het gehalte aan ruw eiwit van het gras als de onafhankelijke of verklarende variabelen beschouwd werden. Schematisch wordt dit model in figuur 1 gegeven. Hierin wordt de veronderstelde causale samenhang met een pijl aangegeven, waarbij het effect aan de punt van de pijl is gelegen. Bij de berekening wordt de invloed van de correlaties tussen de onafhankelijke factoren geëlimineerd. In dit regressiemodel wordt dus de veronderstelling gelegd, dat een verandering van bijv. het K₂O-gehalte van de grond alleen een verandering van het MgO-gehalte van het gras, niet een verandering van het ruw-eiwitgehalte gras tot gevolg heeft. Deze veronderstelling is zoals bekend, vermoedelijk onjuist, zodat de invloeden van K₂O-gehalte grond en ruw-eiwitgehalte op het MgO-gehalte gras foutief berekend kunnen worden. Voor praktische doeleinden is echter kennis van de totale verandering van betekenis, dus niet alleen de directe invloed van het K₂O-gehalte van de grond op het MgO-gehalte van het gras, maar ook de indirecte via een verandering van het ruw-eiwitgehalte en het

Fig. 1 Diagram van een regressiemodel, waarin het MgO-gehalte gras als de afhankelijke variabele, de andere factoren als de onafhankelijke, verklarende variabelen opgenomen zijn

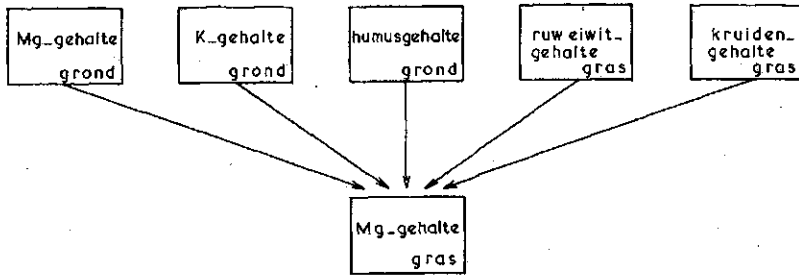


Fig. 1 Diagramm eines Regressionsmodelles in dem der MgO-Gehalt vom Gras als die abhängige Variable, die anderen Faktoren als die unabhängigen, erklärenden Variablen aufgenommen sind

percentage kruiden. De methode met de pad-coëfficiënten biedt o.a. een mogelijkheid wel met deze indirecte invloeden rekening te houden.

Het principe van de methode met de pad-coëfficiënten (path coefficient) is door Wright (1921) geïntroduceerd. Deze heeft in latere publikaties zijn opvattingen toegelicht en verder uitgewerkt. Toepassing heeft voornamelijk in de erfelijkheidsleer plaatsgevonden, hoewel Wright ook toepassingen op andere gebieden zoals de fysiologie bespreekt. Behalve deze en het voorbeeld dat Turner en Stevens (1959) behandelen, zijn ons geen toepassingen buiten het gebied van de erfelijkheidsleer bekend. De methode is overigens lange tijd bij wiskundigen en genetici niet erg populair geweest (Kempthorne, 1957). De laatste jaren wordt de methode in de erfelijkheidsleer echter steeds meer toegepast. De literatuur over de methode is over het algemeen niet gemakkelijk, ook al omdat de meningen op bepaalde punten nogal eens uiteenlopen. Naast de genoemde kunnen ook de artikelen van Li (1956) en Wright (1934, 1960) aanbevolen worden. Van betekenis zijn verder de publikaties van Wright en van Tukey in het boek 'Statistics and Mathematics in Biology' (Kempthorne, 1954).

In de economie heeft men dikwijls met processen te maken, waarin variabelen zowel oorzaak als gevolg zijn. De econometrie heeft de methode van de *simultane vergelijkingen* toegepast om deze processen te analyseren. De methode met de pad-coëfficiënten is hieraan verwant, zodat ook een kennisneming van de hierop betrekking hebbende literatuur aan te bevelen is. Als algemene werken noemen wij de boeken van Tintner (1952) en Valavanis (1959) en de artikelen van Theil en Kloek (1959) en van Basmann (1960). Verder zijn de beschouwingen van Simon (1953, 1954) en Koopmans (1953) over identificatie zeer instructief.

De gedachtengang van de methode zal in het volgende kort besproken worden, waarbij de terminologie is afgestemd op ecologische problemen. De betekenis en de mogelijkheden zullen met behulp van gegevens uit een reeds op andere wijze uitgewerkt voorbeeld (Sluijsmans, 1962) worden toegelicht. De wiskundig-statistische aspecten blijven zoveel mogelijk buiten beschouwing, ook al omdat vele problemen hierbij nog op een oplossing wachten.

PAD-COEFFICIENTEN

Uitgangspunt is een model, dat uit een *gesloten causaal lineair systeem* met m primaire oorzaken x en n effecten y bestaat. Deze m en n variabelen worden verondersteld door een netwerk van causale paden met elkaar verbonden te zijn. Onder een gesloten lineair systeem wordt een netwerk verstaan, waarin een variabele óf een lineaire combinatie van een of meer andere variabelen van dit systeem óf een van de variabelen is die door geen van de variabelen uit dit systeem bepaald wordt. Deze laatste zijn de primaire oorzaken x . Het feit dat een variabele als een lineaire combinatie van een of meer variabelen beschouwd wordt, betekent dat deze als een lineaire functie van een of meer variabelen is uit te drukken. Bij de primaire oorzaken x is dit, althans in het gekozen model, niet het geval. De parameters, die in deze functies de sterkte van de invloed aangeven, worden pad-coëfficiënten genoemd.

Deze pad-coëfficiënten die uit de waarnemingen berekend moeten worden, hebben niet de pretentie causale relaties af te leiden. De onderzoeker zal zelf het causale schema moeten formuleren op basis van de reeds aanwezige kennis en/of op basis van de hypothese die hij wil toetsen. Men tracht door middel van de pad-coëfficiënten de betekenis van de veronderstelde primaire oorzaken en andere variabelen kwantitatief aan te geven of eventueel te verwerpen.

Het is gebruikelijk het causale model in een diagram weer te geven. De causale samenhang tussen twee variabelen wordt hierbij met een pijl aangegeven, waarbij het effect bij de punt van de pijl wordt geplaatst. Aan de hand van een dergelijk schema wordt het effect als een lineaire functie van een of meer variabelen x en y geschreven. De lineariteit is overigens een niet-noodzakelijke veronderstelling.

De berekening van de pad-coëfficiënten geschiedt met de veronderstelling dat de x -en foutloos zijn. Een andere veronderstelling is, dat de fouten van de y -en (meet- en reactiefouten) normaal verdeeld zijn, een gemiddelde gelijk aan nul hebben en niet met elkaar en met de x -en gecorreleerd zijn. De statistische bewerking kan in dit geval volgens de methode van de kleinste kwadraten plaatsvinden.

De uitwerking tot pad-coëfficiënten komt hierop neer, dat de totale invloed van een primaire oorzaak x op een effect y , uitgedrukt als een gewone of partiële regressie-coëfficiënt (*totale pad-regressie*) wordt weergegeven als een som van een aantal *samengestelde pad-regressies*. Een samengestelde pad-regressie geeft de werking van een oorzaak op een effect via één bepaalde reeks van variabelen; zij is gelijk aan het produkt van pad-coëfficiënten in deze reeks.

Theoretisch verkrijgt men de samengestelde pad-regressies door het effect y dat verklaard moet worden, uit te drukken als een functie van de variabelen waarvan het directe resultaat is (*structurele vergelijkingen*). Indien deze laatste variabelen zelf weer geen primaire oorzaken zijn, worden zij op hun beurt als functie van variabelen uitgedrukt. Dit proces van eliminatie gaat

door totdat het eerstgenoemde effect y als een functie van alleen maar primaire oorzaken x uitgedrukt is (*gereduceerde structuurvergelijkingen*). Men heeft dan tevens een partiële regressievergelijking verkregen, waarin de regressiecoëfficiënten als een som van de reeds eerder genoemde samengestelde pad-regressies uitgedrukt zijn. Door deze samengestelde pad-regressies voor alle effecten op te stellen of uit te rekenen en gelijk aan de overeenkomstige regressiecoëfficiënten te stellen verkrijgt men een stelsel van simultane, niet-lineaire vergelijkingen, waarin de regressiecoëfficiënten bekend zijn en waaruit de pad-coëfficiënten *soms* op te lossen zijn.

Een voorbeeld ontleend aan Turner en Stevens (1959), kan dit verduidelijken. Wij stellen hiertoe een eenvoudig causaal model op, dat in figuur 2 is weergegeven. Dit model geeft aan, dat y_3 door de primaire oorzaken x_1 en x_2 bepaald wordt. Het effect y_4 staat direct alleen onder invloed van x_2 en y_3 .

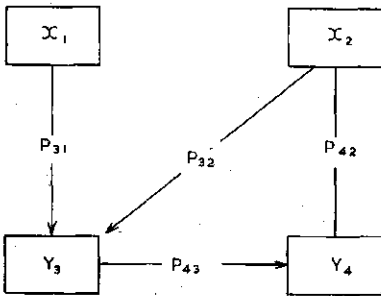


Fig. 2 Diagram van een gesloten causaal systeem met x_1 en x_2 als primaire oorzaken en met een indirecte inwerking van x_2 op y_4 via y_3

Fig. 2 Diagramm eines geschlossenen, kausalen Systems mit x_1 und x_2 als primäre Ursachen und mit einer indirekten Einwirkung von x_2 auf y_4 via y_3

De primaire oorzaak x_1 oefent zijn werking op y_4 alleen indirect via y_3 uit. De notatie van de pad-coëfficiënten p spreekt verder voor zichzelf. Onze ervaring is, dat het gemakkelijker is een doorlopende nummering van de variabelen zonder onderscheiding tussen primaire oorzaken en effecten te gebruiken.

De structurele vergelijkingen die volgens bovenstaande richtlijnen van dit model opgesteld kunnen worden, zijn als volgt:

$$(1) \quad y_3 = a_3 + p_{31} x_1 + p_{32} x_2$$

$$(2) \quad y_4 = a_4 + p_{43} y_3 + p_{42} x_2$$

Hiervan is vergelijking (1) reeds een van de gereduceerde structuurvergelijkingen. De tweede verkrijgt men door (1) en (2) te substitueren:

$$(3) \quad y_4 = a_4 + a_3 p_{43} + (p_{31} p_{43}) x_1 + (p_{42} + p_{32} p_{43}) x_2$$

Deze uitdrukkingen (1) en (3) voor y_3 en y_4 zijn nu ook als partiële regressievergelijkingen te beschouwen, waarin de regressiecoëfficiënten $b_{31.2}$, $b_{32.1}$, $b_{41.2}$ en $b_{42.1}$ uit te rekenen zijn. Door deze regressiecoëfficiënten gelijk aan de overeenkomstige termen uit vergelijkingen (1) en (3) te stellen verkrijgt men een aantal vergelijkingen, waaruit de vier onbekende pad-coëfficiënten voor dit geval op te lossen zouden zijn, nl.:

$$(4) \quad b_{31.2} = p_{31} \text{ en } b_{32.1} = p_{32}$$

$$(5) \quad b_{41.2} = p_{31}p_{43} \text{ en } b_{42.1} = p_{42} + p_{32}p_{43}$$

Men vindt dus direct p_{31} en p_{32} , waarna p_{42} en p_{43} uit (5) op te lossen zijn. Dergelijke stelsels van vergelijkingen zijn niet altijd oplosbaar. In het besproken voorbeeld zijn de pad-coëfficiënten juist geïdentificeerd, omdat zij eenduidig uit de regressiecoëfficiënten berekend kunnen worden. Hiervoor is het noodzakelijk (maar niet altijd voldoende zoals bijv. in gevallen met terugkoppelingssystemen), dat het totale aantal pad-coëfficiënten gelijk is aan het totale aantal regressiecoëfficiënten. Dikwijls treedt *overidentificatie* of *onderidentificatie* op. Het extreme geval van onderidentificatie is het model van de factoranalyse (Ferrari, Pijl en Venekamp, 1957), waarbij zeer weinig regressiecoëfficiënten bekend zijn. Voor een bespreking van de problemen bij over- en onderidentificatie verwijzen wij naar de genoemde econometrische literatuur en naar de artikelen van Simon (1953, 1954).

TOEPASSING OP EEN ECOLOGISCH VRAAGSTUK

Wij laten de mogelijkheden van de methode met pad-coëfficiënten zien aan de hand van resultaten verkregen bij toepassing op dezelfde gegevens, die Sluijsmans (1962) voor de kwantificering van de invloeden op het MgO-gehalte van gras heeft gebruikt.

Hiertoe is allereerst een gesloten causaal systeem opgesteld, dat in figuur 3 is weergegeven. Wij hebben hierin onze kijk op dit vraagstuk verwerkt. Wij brengen verder in herinnering, dat de methode met de pad-coëfficiënten naar voren gebracht werd om het bezwaar te ondervangen om het percentage kruiden en het ruw-eiwitgehalte als onafhankelijke variabelen in het regressiemodel op te nemen. Het model van figuur 3 heeft dit bezwaar niet. Het is

Fig. 3 Diagram met directe en indirecte inwerking van de 4 primaire oorzaken op het MgO-gehalte gras

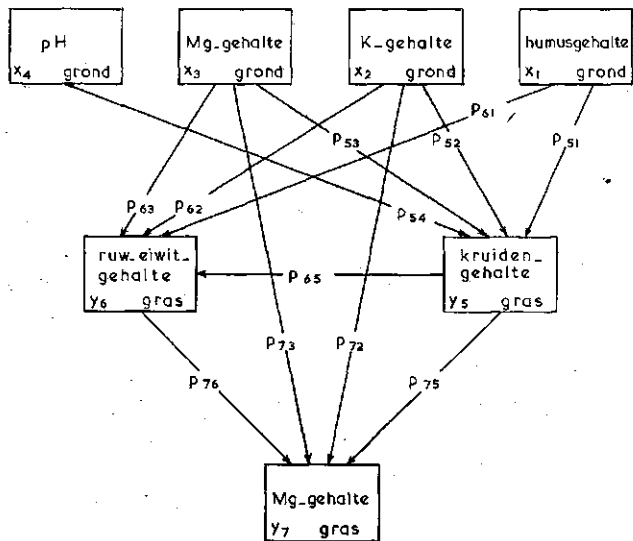


Fig. 3 Diagramm mit direkter und indirekter Einwirkung der 4 primären Ursachen auf den MgO-Gehalt vom Gras

zo opgebouwd, dat beide variabelen zowel oorzaak als effect zijn. Als primaire oorzaken zijn in het model het K_2O - en MgO -gehalte van de grond, het humusgehalte en de pH opgenomen. Twee van deze factoren oefenen direct hun invloed op het MgO -gehalte van het gras uit, nl. het MgO - en het K_2O -gehalte van de grond via de pad-coëfficiënten p_{73} en p_{72} . Het humusgehalte en de pH beïnvloeden het MgO -gehalte van het gras alleen indirect, nl. via het percentage kruiden en/of het ruw-eiwitgehalte. Het ontbreken van een directe invloed van pH en humus is een gevolg van de overweging dat wij ons niet kunnen voorstellen op welke wijze dit zou plaatsvinden. Dezelfde overwegingen gelden ook voor het ontbreken van een causale verbinding van pH met ruw-eiwitgehalte. Een verbinding van humusgehalte met het percentage kruiden en met het ruw-eiwitgehalte is op een mogelijke invloed via resp. vochtvoorziening en stikstoflevering gebaseerd. Het ruw-eiwitgehalte is verder afhankelijk van de botanische samenstelling van het gras. Het is verder duidelijk, dat de inhoud van dergelijke causale modellen nauw met kennis, intuïtie en visie van de onderzoeker samenhangt. Overigens is de redelijkheid van het besproken model voor het betoog niet doorslaggevend. De invloed van het MgO -gehalte van de grond op het MgO -gehalte van het gras gaat via verschillende paden. Allereerst is er de reeds genoemde directe beïnvloeding, aangeduid met pad-coëfficiënt p_{73} . De indirecte invloed ontstaat, omdat een verandering van het MgO -gehalte van de grond direct een verandering in het percentage kruiden en in het ruw-eiwitgehalte via resp. p_{53} en p_{63} teweegbrengt, die zelf ook weer het MgO -gehalte van het gras beïnvloeden. Bovendien beïnvloedt het percentage kruiden ook het ruw-eiwitgehalte. Het MgO -gehalte van het gras hangt dus op een directe en op drie indirecte wijzen af van het MgO -gehalte van de grond. In pad-coëfficiënten kan men deze invloed als volgt uitdrukken:

$$(6) \quad p_{73} + p_{53}p_{75} + p_{63}p_{76} + p_{53}p_{65}p_{78}$$

Overeenkomstige formuleringen zijn ook voor de andere factoren te geven. Deze worden verkregen óf uit de af te leiden gereduceerde structuurvergelijkingen óf direct uit het model door toepassing van een (hier niet) gegeven regel. Het model is oplosbaar, o.a. omdat het aantal pad-coëfficiënten gelijk aan het aantal regressiecoëfficiënten is.

Wij laten nu enkele uitkomsten van de bewerking volgen, maar geven eerst in tabel 1 een korte beschrijving van de gebruikte gegevens. Deze zijn nuttig om de betekenis van de pad-coëfficiënten in het geheel van de ecologische relaties goed te kunnen begrijpen. Evenals de regressiecoëfficiënt geeft de pad-coëfficiënt de verandering van het effect aan onder invloed van een eenheid van verandering bij de oorzakelijke variabele.

In tabel 2 worden de berekende pad-coëfficiënten van het model gegeven. Ter vergelijking geven wij ook in tabel 3 de regressiecoëfficiënten, berekend volgens het model in figuur 1.

Een ideale vergelijking van de gegevens uit tabel 2 met die uit tabel 3 kan niet plaatsvinden, omdat de pH als factor alleen in het model met de

PAD-COEFFICIENTEN IN LANDBOUW-ECOLOGISCH ONDERZOEK

Tabel 1 Laagste, hoogste en gemiddelde waarde van de in het model opgenomen variabelen

variabele	laagste - hoogste waarde	gemiddelde waarde
humusgehalte grond	3,4 — 12,6	7,8
K ₂ O-gehalte grond	7 — 30	16,5
MgO-gehalte grond	50 — 227	125
pH grond	4,9 — 5,9	5,5
percentage kruiden	1,1 — 37,5	13,5
ruw-eiwitgehalte gras	9,5 — 24,6	14,1
MgO-gehalte gras	0,209 — 0,383	0,293
K ₂ O-gehalte gras	1,95 — 4,05	2,82

Tabelle 1 Niedrigste, höchste und mittlere Wert der im Modell aufgenommenen Variablen

Tabel 2 Berekende waarden van de 12 pad-coëfficiënten uit het model van figuur 3

effect \ oorzaak	humus (x ₁)	K ₂ O-gehalte grond (x ₂)	MgO-gehalte grond (x ₃)	pH (x ₄)	percentage kruiden (y ₅)	ruw-eiwitgehalte (y ₆)
percentage kruiden (y ₅)	1,67	-0,23	-0,031	5,26		
ruw-eiwitgehalte (y ₆)	-0,74	0,11	0,011		0,20	
MgO-gehalte gras (y ₇)		-0,0038	0,0004		0,0041	0,0083

Tabelle 2 Berechnete Werte der 12 Pfad-Koeffizienten aus dem Modell von Figur 3

Tabel 3 Regressiecoëfficiënten berekend volgens het model in figuur 1

effect \ oorzaak	humus	K ₂ O-gehalte grond	MgO-gehalte grond	percentage kruiden	ruw-eiwitgehalte
MgO-gehalte gras	0,0001	-0,0038	0,0004	0,0029	0,0059
K ₂ O-gehalte gras	-0,078	0,040	0,003	-0,002	0,092

Tabelle 3 Regressionskoeffizienten berechnet nach dem Modell in Figur 1

pad-coëfficiënten en niet in het regressiemodel is opgenomen. Desondanks is een vergelijking instructief.

Het blijkt dat de regressiecoëfficiënten en de pad-coëfficiënten p₇₂ en p₇₃ als maat voor de rechtstreekse invloeden van het K₂O- en het MgO-gehalte van de grond op het MgO-gehalte van het gras van gelijke grootte zijn. Het is verder opvallend, dat de (hier niet opgegeven) totale pad-coëfficiënten aan deze waarden gelijk zijn. De rechtstreekse invloeden van het percentage kruiden en het ruw-eiwitgehalte verschillen daarentegen sterk van de overeenkomstige regressiecoëfficiënten; de verhouding is ongeveer 1½. In dit opzicht is een advisering waarbij met het bestand en het ruw-eiwitgehalte van het gras op basis van een regressiemodel rekening gehouden wordt, dus inderdaad onjuist.

Tabel 2 geeft verder nog verschillende interessante gegevens. Wij noemen de negatieve invloed van het K₂O- en MgO-gehalte van de grond op het percentage kruiden, de positieve invloed op het ruw-eiwitgehalte, de uit te rekenen positieve totale invloed (rechtstreekse plus niet-rechtstreekse via het bestand) op het ruw-eiwitgehalte, de negatieve invloed van het K₂O-gehalte

van de grond op het MgO-gehalte van het gras, de onverwachte negatieve invloed van het humusgehalte op het ruw-eiwitgehalte enz. Deze laatste negatieve invloed van het humusgehalte wijst erop, dat de werking van de humus zeker niet (alleen) in een levering van stikstof gezocht moet worden. Het ligt buiten de opzet van deze bijdrage hierop verder in te gaan. Het bovenstaande moge voldoende zijn om de mogelijkheden van de methode met de pad-coëfficiënten te demonstreren.

DISCUSSIE

Ons uitgangspunt was de onjuiste formulering van het regressiemodel indien variabelen die zelf afhankelijk van andere opgenomen variabelen zijn, alleen als onafhankelijke factoren beschouwd worden. Wij hebben laten zien welke mogelijkheden de methode met de pad-coëfficiënten hierbij biedt. De methode heeft echter nog enkele andere aspecten, die van belang zijn en die nog kort behandeld zullen worden. Men moet zich echter realiseren, dat de ervaringen bij toepassing op ecologische vraagstukken nog gering zijn en dat nog vele vragen van wiskundige aard, o.a. de statistische toetsing van de pad-coëfficiënten, om een oplossing vragen (Basmann, 1960).

Wij hebben de methode toegepast op gegevens, die van een proef zonder ingreep afkomstig waren. De methode is echter ook bruikbaar bij de opstelling en uitwerking van proeven met ingreep. Het kan voordelen hebben deze methode te gebruiken. In een proef met ingreep wordt nl. verondersteld, dat de niet in beschouwing genomen factoren niet veranderen (Ferrari, 1960). De noodzaak een uitgebreider causaal model te moeten opstellen, de afweging van de noodzakelijkheid om bepaalde variabelen wel of niet op te nemen enz. maken de onderzoeker erop attent welke metingen hij moet verrichten om een zo juist mogelijk beeld van de werkelijkheid te krijgen. Het model maakt de onderzoeker er verder opmerkzaam op, dat het resultaat van een ingreep op verschillende wijzen kan plaatsvinden. Een resultaat kan verkregen worden zonder dat er een direct causaal verband tussen twee variabelen bestaat en omgekeerd. Hieruit blijken dus ook de beperking van een proef met ingreep en de grote betekenis van de invloed van de steekproef op het resultaat.

De methode met de pad-coëfficiënten dwingt de onderzoeker, die met gegevens van een proef zonder ingreep werkt, zich te realiseren op welke wijze bepaalde verschijnselen in werkelijkheid tot stand komen, en een hypothese op te stellen. Kennis, intuïtie, combinerend vermogen enz. van de onderzoeker worden hierdoor op de proef gesteld, hetgeen alleen een verrijking kan zijn.

De methode moet ongetwijfeld van grote betekenis voor het meer synthetisch ingestelde ecologische onderzoek zijn, hoewel de mogelijkheden uiteraard hierbij niet onbeperkt zijn. Het besprokene gaf hiervan reeds een voorbeeld. Het is echter mogelijk hierin nog verder te gaan.

Wij lieten in het model van figuur 3 het K₂O-gehalte van het gras buiten beschouwing; er werd wel een directe causale beïnvloeding door het K₂O-

Fig. 4 Diagram met directe en indirecte inwerking van de 4 primaire oorzaken op het MgO-gehalte gras, waarbij de werking van het K₂O-gehalte grond indirect via het K₂O-gehalte gras plaatsvindt

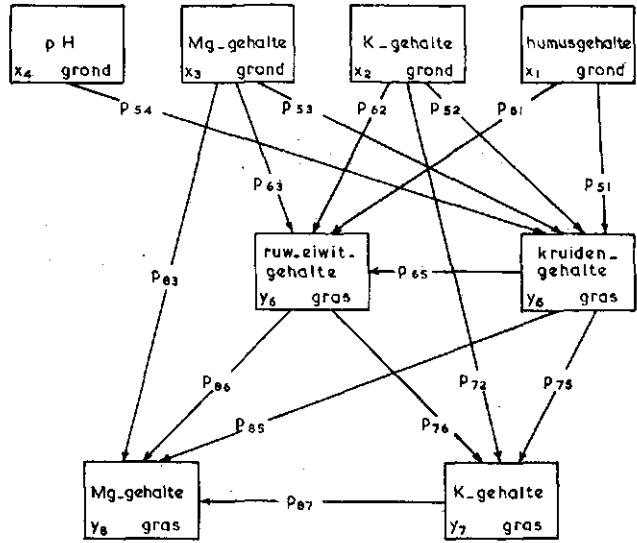


Fig. 4 Diagramm mit direkter und indirekter Einwirkung der 4 primären Ursachen auf den MgO-Gehalt vom Gras, wobei die Wirkung des K₂O-Gehaltes vom Boden indirekt über den K₂O-Gehalt vom Gras stattfindet

gehalte van de grond via de pad-coëfficiënt p₇₂ verondersteld. Een uitspraak over de aard van deze beïnvloeding werd niet gedaan. In het model van figuur 4 is echter de variabele K₂O-gehalte van het gras opgenomen, waarbij causale verbindingen van het K₂O-gehalte van het gras met het K₂O-gehalte van de grond, het percentage kruiden en het ruw-eiwitgehalte en verder van het MgO-gehalte met het K₂O-gehalte van het gras aangenomen zijn. Een en ander betekent dat de werking van het K₂O-gehalte grond op het MgO-gehalte van het gras van plantenfysiologische en niet van bodemchemische aard verondersteld is. Het model is overigens een voorbeeld van overidentificatie, omdat de pad-coëfficiënten p₇₂, p₇₅ en p₇₆ niet eenduidig te bepalen zijn. De eenduidige oplossing is verkregen door een beperkende relatie, nl. dat het percentage kruiden het K₂O-gehalte van het gras niet beïnvloedt, te veronderstellen. Dit betekent, dat de coëfficiënt genomen is, die het minst van nul afwijkt. Het resultaat wordt in tabel 4 gegeven.

Tabel 4 Berekende waarden van de pad-coëfficiënten met betrekking tot het K₂O-gehalte van het gras uit het model van figuur 4

effect \ oorzaak	K ₂ O-gehalte grond (x ₂)	percentage kruiden (y ₅)	ruw-eiwit-gehalte (y ₆)	K ₂ O-gehalte gras (y ₇)
K ₂ O-gehalte gras (y ₇)	0,029	-0,007	0,184	
MgO-gehalte gras (y ₈)				-0,1307

Tabelle 4 Berechnete Werte der Pfad-Koeffizienten in Bezug auf den K₂O-Gehalt vom Gras aus dem Modell von Figur 4

De antagonistische werking komt in de negatieve waarde van de pad-coëfficiënt p₈₇ tot uiting.

Het laatste model is niet ideaal, omdat ook een invloed van het MgO-

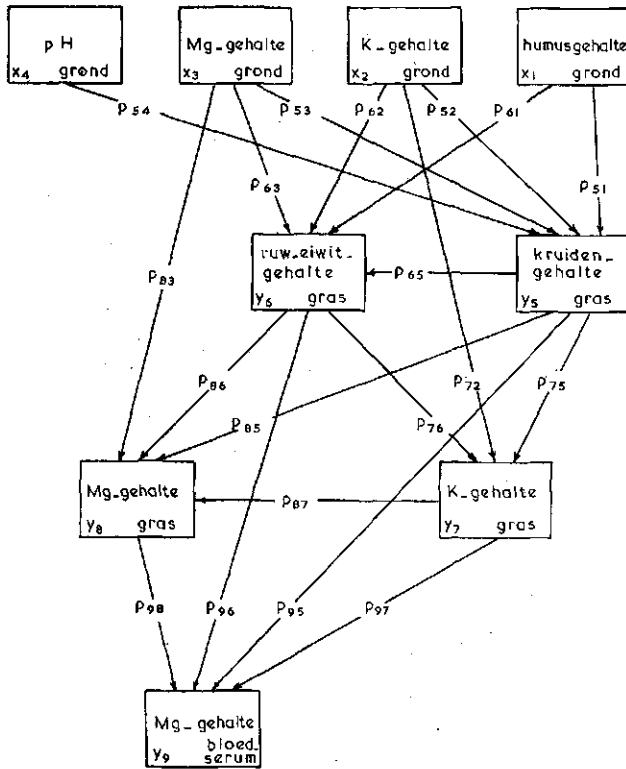


Fig. 5 Diagram met indirecte werking van de 4 primaire oorzaken op het MgO-gehalte van het bloedserum

Fig. 5 Diagramm mit indirekter Wirkung der 4 primären Ursachen auf den MgO-Gehalt des Blutserums

gehalte op het K₂O-gehalte van het gras verondersteld mag worden. Dit betekent, dat ook een pijl van y_8 naar y_7 getrokken moet worden. Men heeft dan een systeem van *terugkoppeling*, dat echter in dit geval niet oplosbaar is (onderidentificatie). Een sterk vereenvoudigd model gaf als oplossing een positieve invloed van het MgO en een negatieve van het K₂O in het gras. Dergelijke terugkoppelingssystemen met tegengestelde coëfficiënten laten een *tegenkoppeling* zien, waarin men in dit geval een verklaring voor de constante som van de kationen kan zien.

De methode met de pad-coëfficiënten stimuleert om nog verder te gaan met de synthese. Volgens Kemp (1960) is het MgO-gehalte van het bloedserum van grote betekenis voor het optreden van hypomagnesaemie. Aan de andere kant moet het MgO-gehalte van het gras belangrijk voor het MgO-gehalte van het bloedserum zijn. Verder is het bekend, dat het ruw-eiwitgehalte en het K₂O-gehalte van het gras ongunstig op het Mg-gehalte van het bloed werken. Een en ander is in het model van figuur 5 weergegeven. Een bestudering laat zien op welke wijze het Mg-gehalte van het bloedserum beïnvloed kan worden. Een kwantificering van de pad-coëfficiënten kan aangeven op welke wijze dit het meest efficiënt gebeuren kan. Dit model is inderdaad oplosbaar. Het is echter te betreuren, dat de hiervoor benodigde gegevens over het Mg-gehalte van het bloedserum in dit onderzoek ontbreken.

ZUSAMMENFASSUNG

Die in einem Experiment ohne Eingriff erzielten Ergebnisse werden oft mit einer Regressionsgleichung ausgewertet. Im Modell dieser Gleichung wird eine Variable durch die sonstigen sog. unabhängigen Variablen „erklärt“. Die Regressionskoeffizienten geben dann jedesmal die Zunahme des Effektes an wenn eine Ursache um 1 wächst, unter Annahme dass die sonstigen erklärenden Variablen durch die Aenderung dieser Ursache selbst nicht geändert werden (partielle oder Teilregression). In vielen Fällen entspricht diese Annahme aber nicht der Wirklichkeit.

Die von *Wright* entwickelte Methode mit den Pfad-Koeffizienten gibt die Möglichkeit diese Schwierigkeiten bisweilen zu beseitigen. Hierzu muss der Forscher ein geschlossen kausales, lineares System mit m primären Ursachen x und n Effekten y aufsetzen. Unter einem geschlossen kausalen System wird ein Netzwerk verstanden in dem jede Variable entweder eine lineare Kombination einer oder mehrerer Variablen dieses Systems oder eine der Variablen ist, welche durch keine der Variablen des Systems bestimmt ist. Die letzten Variablen sind darin die primären Ursachen x . Die Grösse der Einflüsse wird durch die Pfad-Koeffizienten gegeben. Ein beispiel der Auswertung der Pfad Koeffizienten des einfachen Systems aus Figur 2 wird durch die Gleichungen 1—5 gegeben.

Die Möglichkeiten der Methode mit den Pfad-Koeffizienten werden vorgeführt an Hande einer Untersuchung nach den Einflüssen von Boden- und anderen Faktoren auf den MgO- und K₂O-Gehalt des Weidegrases in Bezug auf die Wichtigkeit dieser Zusammensetzung für das Auftreten von Hypomagnesaemie. Das Regressionsmodell dieser Untersuchung wird in Figur 1 gegeben. Figur 3 gibt ein mehr reelles Modell, worin die Variablen Prozentsatz an Kräutern und Rohweissgehalt des Grases Ursache sowohl wie Effekt sind. Die Pfad-Koeffizienten dieses Modelles sind lösbar und werden in Tabelle 2 gegeben. Zum Vergleich werden in Tabelle 3 die Regressionskoeffizienten des Modelles aus Figur 1 gegeben. Im Modell von Figur 4 wird der Einfluss von Kali im Boden mit Hilfe des Kaligehaltes vom Gras physiologisch gedeutet.

Die Methode mit den Pfad-Koeffizienten hat viele Vorzüge vor dem Regressionsmodell. Die Methode gibt weiter grosse Möglichkeiten für eine synthetische Auswertung der Ergebnisse. Im Modell von Figur 5 wird eine Synthese zwischen Bodenfaktoren, Zusammensetzung des Weidegrases und Mg-Gehalt des Blutes gegeben. Die Pfad-Koeffizienten dieses Modelles sind lösbar.

LITERATUUR

- BASMANN, R. L. (1960): An expository nota on estimation of simultaneous structural equations. *Biometrics* 16, 464—480.
- FERRARI, TH. J., H. PIJL en J. T. N. VENKAMP (1957): Factor analysis in agricultural research. *Neth. J. Agric. Sci.* 5, 211—221.
- FERRARI, TH. J. (1960): Vergelijking tussen proeven met en zonder ingreep. *Landbouwk. Tijdschr.* 72 792—801.
- KEMP, A. (1960): Hypomagnesaemia in milking cows: The response of serum magnesium to alterations in herbage composition resulting from potash and nitrogen dressings on pasture. *Neth. J. Agric. Sci.* 8, 281—304.
- KEMPTHORNE, O. (1957): An introduction to genetic statistics. New York.
- KOOPMANS, T. C. (1953): Identification problems in economic model construction. Chapter II, Studies in Econometric Method, Cowles Commission Monograph no. 14, ed. W. C. Hood and T. C. Koopmans.
- LI, C. C. (1956): The concept of path coefficients and its impact on population genetics. *Biometrics* 12, 190—210.

- SIMON, H. A. (1953): Causal ordering and identifiability. Chapter III, Studies in Econometric Method, Cowles Commission Monograph no. 14, ed. W. C. Hood and T. C. Koopmans, 1953.
- (1954): Spurious correlation: a causal interpretation. *Am. Stat. Ass.* 49, 467—479.
- SLUIJSMANS, C. M. J. (1962): Magnesium- en kaliumgehalten van gras in afhankelijkheid van bodem- en andere factoren. *Tijdschr. Diergeneesk.* 87, 457—556.
- THEIL, H. en T. KLOEK (1959): The statistics of systems of simultaneous economic relationships. *Statistica Neerlandica* 13, 65—89.
- TINTNER, G. (1952): *Econometrics*. New York.
- TUKEY, J. W. (1954): Causation, regression and path analysis. Chapter 3, *Statistics and Mathematics in Biology*, ed. Kempthorne.
- TURNER, M. E. en C. D. STEVENS (1959): The regression analysis of causal paths. *Biometrics* 15, 236—258.
- VALAVANIS, S. (1959): *Econometrics. An introduction to maximum likelihood methods*. New York.
- WRIGHT, S. (1921): Correlation and causation. *J. Agri. Res.* 20, 557—585.
- (1934): The method of path coefficients. *Ann. Math. Stat.* 5, 161—215.
- (1954): The interpretation of multivariate systems. Chapter 2, *Statistics and Mathematics in Biology*, ed. Kempthorne.
- (1960): Path coefficients and path regressions: alternative or complementary concepts? *Biometrics* 16, 189—202.
- (1960): The treatment of reciprocal interaction, with or without lag, in path analysis. *Biometrics* 16, 423—445.

Groningen, mei 1962