

Canonische
Correspondentie
Analyse

J.T. van der Wal

Inhoudsopgave

1.	<u>Inleiding</u>	1
	1.1. <u>Canonische correspondentie analyse</u>	1
	1.2. <u>Vraagstelling</u>	1
2.	<u>Materiaal & Methode</u>	2
	2.1. <u>CANOCO</u>	2
	2.2. <u>Wijze van testen</u>	2
3.	<u>Canonische Correspondentie Analyse</u>	5
	3.1. <u>Uitwerking van het Algoritme.</u>	5
	3.2. <u>Beoordelen van de Resultaten.</u>	8
	3.3. <u>Test met kunstmatige dataset</u>	10
	3.4. <u>Test met Kwelderdataset met 'verzonnen' milieugegevens</u>	15
	3.5. <u>Gebruik van CANOCO in de praktijk.</u>	21
	3.5.1. <u>Gebruik CANOCO.con</u>	21
	3.5.2. <u>CANOLOT en Machine Readable Copy</u>	21
	3.5.3. <u>Actief versus Passief</u>	22
4.	<u>Conclusies.</u>	23
	<u>Literatuurlijst.</u>	25
	<u>Lijst van Figuren.</u>	27
	<u>Lijst van Tabellen.</u>	27
	<u>Index</u>	28

Canonische Correspondentie Analyse

1. Inleiding

1.1. Canonische correspondentie analyse

Canonische correspondentie analyse, is een methode voor de analyse van gegevens, die pas korte tijd bruikbaar is voor grote gegevenssets.

Een van de redenen hiervoor is dat deze rekenmethode veel rekenwerk vergt. Tot voor kort ontbrak het computers aan rekenkracht om deze analyse in aanvaardbare tijd uit te voeren.

Een andere reden is dat een geschikte rekenmethode (algoritme) nog niet was ontwikkeld voor het uitvoeren van deze analyse. Doordat zo'n algoritme ontwikkeld is, is het mogelijk geworden de analyse in een computerprogramma in te bouwen.

Voor gebruik binnen de Vegetatiekunde heeft de Canonische Correspondentie Analyse een aantal aantrekkelijke kanten. Het gaat om een 'direct gradient analysis' d.w.z. dat tijdens de ordinatie direct rekening gehouden met de vegetatie en met de milieuv variabelen. Kenmerk is dat het ordinatiediagram in principe beperkt is tot de mogelijkheden die de milieuv variabelen bieden. Dit houdt in dat de assen van het diagram te beschrijven zijn als, lineaire combinaties van de milieuv variabelen. Nieuw is ook dat bij deze analysemethode direct rekening kan worden gehouden met interacties tussen de factoren die het milieu bepalen.

De onderzoeker kan zich bij gebruik van Canonische Correspondentie Analyse veel tijd besparen doordat hij niet meer na de ordinatie de invloeden van de milieuv variabelen in het diagram hoeft te interpreteren, zoals bij 'indirect gradient analysis'. De milieuv variabelen kunnen rechtstreeks worden geplot in het ordinatiediagram, op basis van de uitgevoerde berekeningen.

Canonische Correspondentie Analyse is een verfijning van Correspondentie Analyse, een techniek die ook onder de naam Reciprocal Averaging bekend is.

Een voordeel dat Canonische Correspondentie Analyse samen met o.a. Reciprocal Averaging en Detrended Correspondance Analysis t.o.v. oudere analysemethodes zoals Principal Component Analysis heeft, is dat niet langer van een lineair (rechtlijnig) verband tussen de milieufactor en de reactie op verandering daarvan door de plant wordt uitgegaan. In plaats hiervan, wordt uitgegaan van een unimodale reactie van de plant op het veranderen van een milieufactor.

Een unimodale curve is de al van vele planten en vegetaties bekende 'optimum-curve'.

Voordat een nieuwe analysemethode toegepast kan worden zonder het gebruik van beproefde methoden moet deze methode getest worden.

1.2. Vraagstelling

De Vraagstelling die in dit verslag aan de orde komt luidt als volgt:

Zijn de resultaten van Canonische Correspondentie Analyse goed te interpreteren (handig/overzichtelijk). En welke stappen worden gemaakt tijdens de interpretatie.

Canonische Correspondentie Analyse

2. Materiaal & Methode

2.1. CANOCO

Canoco is een computerprogramma dat ontwikkeld is door het IWIS-TNO en is geschreven door C.J. Ter Braak. Het programma biedt naast CCA (Canonische Correspondentie Analyse) een veelvoud aan andere mogelijkheden om datasets te analyseren.

Canoco is speciaal geschreven voor gebruik binnen de Vegetatiekunde, maar is ook geschikt voor de verwerking van datasets die uit ander onderzoek voortkomen.

Het programma is geschreven in Fortran en onder andere beschikbaar voor VAX-computer's en P.C.'s. Op de R.U.Groningen is het programma beschikbaar op de VAX8600 van het RekenCentrum van deze universiteit.

Bij de ontwikkeling van Canoco is gebruik gemaakt van de broncode van het programma DECORANA, ontwikkeld door M.O. Hill & H.G. Gauch van de Cornell University (Hill M.O., Gauch H.G. jr. 1980).

CANOCO kan besteld worden bij het IWIS TNO, Postbus 100, 6700 AC Wageningen, Nederland. Het IWIS TNO is telefonisch bereikbaar op 08370-19100

2.2. Wijze van testen

De werking en methode van de canonische correspondentie analyse is getest met de analyse zoals in het programma CANOCO geïmplementeerd.

Hierbij is gebruik gemaakt van 2 verschillende datasets.

- 1) een kleine kunstmatige dataset ontleend aan Kramer, 1986. In Tabel 1 staat deze set gegevens. Tabel 1 bestaat uit twee delen.

In het eerste deel zijn de gegevens van de opnames weergegeven in een 'plaats'x'soort'-matrix. Het tweede deel is een 'plaats'x'milieuvariabele'-matrix.

- 2) een bestaande set opnamen is voorzien van kunstmatige milieugegevens. Deze kunstmatige gegevens zijn gebaseerd op de omstandigheden ter plaatse, maar zijn gefingeerd (geschat). Als Bijlage 1 is deze gegevensset bijgevoegd.

De resultaten van CCA zijn beoordeeld op kwaliteit en interpreteerbaarheid.

Bij de beoordeling is voor het maken van grafieken gebruik gemaakt van Canoplot een hulpprogramma bij Canoco. Canoplot leest de Machine Readable Copy-files die door Canoco geschreven kunnen worden. Op basis van de gegevens in deze MRC-files kan op een snelle wijze door het beantwoorden van enkele vragen een hele reeks grafieken geproduceerd worden. De grafieken zoals ze in het verslag zijn opgenomen zijn op een P.C. gemaakt met een grafisch pakket. Dit is gedaan omdat de grafieken van Canoplot, zogenaamde charactergraphics zijn, alleen geschikt voor beoordeling van een ordinatie tijdens het werk (Bijlage 3).

Canonische Correspondentie Analyse

Tabel 1: Kunstmatige Dataset (Kramer, 1986)
(deel a opnamegegevens, en deel b milieugegevens)

a) Kunstmatige Opnamegegevens

	Srt.1	Srt.2	Srt.3	Srt.4	Srt.5	Srt.6	Srt.7	Srt.8
Opn.1	40	10	4	4	4	20	0	20
Opn.2	20	50	20	4	4	20	10	4
Opn.3	4	10	40	4	4	20	0	20
Opn.4	50	4	0	10	20	10	0	20
Opn.5	30	20	10	10	20	10	20	1
Opn.6	10	20	30	10	20	10	4	10
Opn.7	0	4	50	10	20	10	0	20
Opn.8	40	10	4	20	4	4	0	20
Opn.9	20	50	20	20	4	4	0	20
Opn.10	4	10	40	20	4	4	0	20

b) Kunstmatige Milieugegevens

	Grad.1	Grad.2	Bijfactor
Opn.1	2	0	12
Opn.2	6	0	6
Opn.3	10	0	0
Opn.4	0	4	12
Opn.5	4	4	10
Opn.6	8	4	2
Opn.7	12	4	0
Opn.8	2	8	12
Opn.9	6	8	12
Opn.10	10	8	6

Opzet van de kunstmatige dataset.

De kunstmatige dataset is ontleend aan Kramer 1986. In dat verslag wordt de idee achter de opzet van de dataset uitgebreid behandeld. Deze opzet zal hier in het kort behandeld worden.

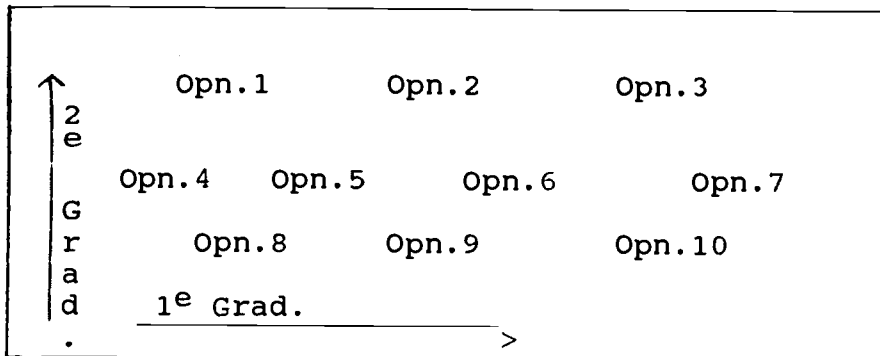
Als uitgangsmateriaal voor de dataset dient een denkbeeldig proefveldje. Twee onafhankelijke gradiënten zijn aanwezig in het proefveldje, deze gradiënten staan bovendien haaks op elkaar. De eerste gradiënt beïnvloedt het voorkomen van de soorten 1,2 en 3. Soort 1 komt het meest voor onderin de gradiënt, soort 3 juist bovenin, het soort 2 komt het meeste voor in het middengedeelte van de eerste gradiënt. De tweede gradiënt bepaaldt het voorkomen van de soorten 4,5 en 6. Het voorkomen van de soorten 4,5 en 6 op deze gradiënt is analoog met de soorten 1,2 en 3 en de eerste gradiënt. Deze beide gradiënten zijn globaal, ze beïnvloeden de vegetatie in het hele proefveld.

Lokale factoren kunnen ook invloed hebben op een vegetatie. De soorten 7 en 8 worden door een lokale factor bepaald. Soort 7 is afhankelijk van de lokale factor, "Bijfactor" genoemd, terwijl de bijfactor door soort 8 juist gemeden wordt. Zie, Figuur 1: Geografie van de Kunstmatige Dataset, en Kramer, 1986. Voor de analyse is gebruikt gemaakt de versie met hoge bedekkingswaarden. In Kramer, '86 staat ook een versie van dezelfde dataset met lage bedekkingswaarden, het verschil tussen deze versies beperkt zich tot de gebruikte schattingschaal voor de bedekkingen.

Canonische Correspondentie Analyse

Om de werking van het programma en de theorie van de Canonische Correspondentie Analyse beter te begrijpen is het gebruikte (algemene) algoritme bestudeerd (ter Braak & Prentice, 1986; ter Braak, 1987). Welke stappen er worden doorlopen tijdens de analyse wordt hierdoor duidelijk. Terwijl de werking van de analyse ook duidelijker wordt.

Figuur 1: Geografie van de Kunstmatige Dataset.



Canonische Correspondentie Analyse

3. Canonische Correspondentie Analyse

3.1. Uitwerking van het Algoritme.

Het algoritme (rekenmethodiek) dat voor CCA wordt gebruikt is een afgeleide van het algemene algoritme zoals dat behandeld wordt in ter Braak & Prentice, 1986 en ter Braak, 1987.

Bij deze omschrijving wordt gebruikt gemaakt van matrixalgebra. En de oplossing van de formules is algebraïsch juist zoals deze genoteerd staan. Voor het oplossen van deze formules moeten echter niet deze formules worden opgelost maar moet door middel van een numerieke aanpak een voorloper van de formule worden opgelost. Dit omdat de formules zelf bewerkingen vereisen die niet toelaatbare bronnen van fouten (kunnen) zijn of die anderszins problemen opleveren.

Het gebruikte algoritme is een iteratief algoritme, hetgeen inhoud dat in opeenvolgende berekeningscycli de werkelijke waarde beter benaderd wordt. Indien de vordering gemaakt in twee opeenvolgende bewerkingssreeksen kleiner is dan een ingestelde waarde, wordt de benadering als optimaal beschouwd. Er wordt slechts een beperkt aantal iteratieslagen gemaakt. Wordt niet binnen dit aantal cycli de afslagwaarde gehaald, dan wordt de benadering afgebroken en gaat het programma verder. Deze minder wenselijke afbreking van de benadering wordt gemeld.

De gebruikte notaties zijn:

$Y = [y_{ik}]$ = de matrix met de 'plaats'x'soorten'-gegevens.

$Z_2 = [z_{i1}]$ = de matrix met de 'plaats'x'milieuvar.'-gegevens.

$Z_1 = [z_{ij}]$ = de matrix met de 'plaats'x'covar.'-gegevens.

(indien aanwezig. Covariabelen zijn variabelen waarvan de invloed bekend is of wordt verondersteld.)

$\hat{u} = [u_k]$ = soort scores

$\hat{x} = [x_i]$ = plaats score

$\hat{c} = [c_j]$ = canonische coëfficiënten > voor de s^e ordinatie-
voor de milieuvariabelen | as.

A = verzamelmatrix plaatsscores van de 'vorige' (s-1) ordinatie-assen.¹

i,n = aantal plaatsen

k,m = aantal soorten

j = aantal milieuvariabelen

l = aantal covariabelen

\ > q = j + l = totaal milieuvar's
/

¹ In de voor het algemene algoritme genoemde bronnen word een andere vorm van het algoritme behandeld. Het verschil is da de gegevens anders in de matrices staan. Rijen en Kolomme wisselen van plaats. Hierdoor zijn de formules verschillend maa is het rekenwerk (practisch) en het eindresultaat (volledig hetzelfde. De verschillen in de formules zitten in de plaatsinge van de accenten bij de matrices. Deze accenten houden ook ee verwisseling van rijen en kolommen in de matrix in, nl. d getransponeerde vorm van de matrix.

Canonische Correspondentie Analyse

Voordat de gegevens over worden gelaten aan het iteratieve algoritme worden enkele inleidende berekeningen uitgevoerd. Deze berekeningen zijn:

- 1) Berekenen van soortstotalen = Y_{+k} \

plaatstotalen = Y_{i+}	/	randtotalen van Y
groottotaal = Y_{++}		

 $r_k = Y_{+k}$ = reeks met soortstotalen
 $w_i = Y_{i+}$ = reeks met plaatstotalen
 $w_i^* = Y_{i+} / Y_{++}$ = reeks met relatieve plaatstotalen

- 2) Standaardiseren van de milieuvariabelen tot een gemiddelde 0 en variantie 1.

Hierbij wordt gebruik gemaakt van de relatieve plekttotalen. Voor Milieuvariabele j wordt gemiddelde \underline{z} en variantie v als volgt berekend:

$$\underline{z} = \sum_i \dot{E}_i w_i^* z_{ij}$$

$$v = \sum_i \dot{E}_i w_i^* (z_{ij}^2 - \underline{z}^2)$$

Vervolgens wordt gestandaardiseerd volgens:

$$z_{ij} = (z_{ij} - \underline{z}) / v$$

Dan komt het Iteratieve Algoritme, dat uit de volgende stappen bestaat.

- 1) Kies arbitraire, maar ongelijke scores voor de plaatsen.

$$\tilde{x} = [x_i] \text{ en stel } x_i^0 = x_i$$

- 2) Bereken nieuwe soort-scores uit de plaats-scores volgens:

$$u_k = \sum_i \dot{E}_i Y_{ik} x_i / r_k$$

- 3) Bereken nieuwe plaats-scores $\underline{x}^* = [x_i^*]$ uit de soort-scores (zie 2) volgens:

$$x_i^* = \sum_k \dot{E}_k Y_{ik} u_k / w_i$$

- 4) Indien $q \leq s_A$ (d.w.z. er zijn minder milieuvariabelen dan er reeds kolommen in A staan) dan $x_i = x_i^*$

Anders (dus wanneer $q > s_A$): bereken de multipele regressie van \tilde{x}^* op Z2 volgens:

$$\tilde{c} = (\tilde{Z}2'W\tilde{Z}2)^{-1} \cdot \tilde{Z}2'W\tilde{x}^*$$

In dit geval zijn er meer milieuvariabelen dan er kolommen in A zijn opgeslagen.

Numeriek komt dit neer op het volgende probleem. Een onbekende vector (c) berekenen uit een bekende vector (b) en dito matrix (A). $c = A^{-1} \cdot b$, omdat delen door een matrix niet mogelijk is en vermenigvuldigen met de inverse vaak problemen oplevert doordat de inverse slecht te bepalen is, moet het numerieke probleem $A \cdot c = b$ opgelost worden.

Stel vervolgens: $\tilde{x} = \tilde{Z}2\tilde{c}$

Canonische Correspondentie Analyse

5) Indien $s > 0$ (1 of meer kolommen opgeslagen in A, dus 1 of meer iteratieslagen afgesloten) maak $\tilde{x} = [x_i]$ ongecorrleerd met de vorige assen door de residuen te berekenen van de multipele regressie van \tilde{x} op A.

$$\tilde{x} = \tilde{x} - A'(AWA')^{-1} AW \tilde{x}$$

6) Standaardiseer $\tilde{x} = [x_i]$ tot gemiddelde 0 en variantie 1 volgens:

$$\begin{aligned} \bar{x} &= \hat{E}_i w^*_i x_i \\ s^2 &= \hat{E}_i w^*_i (x_i - \bar{x})^2 \\ x_i &= (x_i - \bar{x})/s \end{aligned}$$

7) Controleer op convergentie (Is de benadering genoeg verbeterd in de laatste bewerkingscyclus om een volgende cyclus te verantwoorden?)

Indien $\hat{E}_i w^*_i (x^{\circ}_i - x_i)^2 < 10^{-10}$; Ga naar stap 8;
anders stel $\tilde{x}^{\circ}_i = x_i$ en ga naar stap 1.

8) Stel de eigenwaarde λ gelijk aan s (van stap 6) en voeg $\tilde{x} = [x_i]$ als een nieuwe rij aan A toe.

9) Stel $s = s+1$ en ga naar stap 0,
indien nodig en er nog meer ordinatie-assen geëxtraheerd kunnen worden
anders STOP.

Wanneer in de Canonische Correspondentie Analyse wel gebruikt wordt gemaakt van wat in het programma covariabelen wordt genoemd, moeten aan het algoritme enkele stappen worden toevoegd.

Ten eerste: er komt een stap 3 bij, bij de Inleiden Berekeningen. In deze stap worden de restwaarden van de milieuvariabelen bepaald na multiple regressie van de milieuvariabelen op de covariabelen. (De invloed van de covariabelen die bekend wordt verondersteld wordt op deze wijze uit de gegevens verwijderd.) De gebruikte formules zijn:

$$\begin{aligned} \tilde{c}^*_j &:= (Z_1 W Z_1')^{-1} Z_1 W \tilde{z}_{2j} \\ \tilde{z}_{2j} &:= \tilde{z}_{2j} - Z_1 \tilde{c}^*_j \end{aligned}$$

In deze formules staat W voor diagonale matrix van de plaatstotalen w.

Ten tweede: er moet een stap tussengevoegd worden tussen stap 3 en 4 van het iteratieve algoritme. In deze stap worden plaatsscores ongecorrleerd gemaakt met de covariabelen door de restwaarden te bepalen van de multiple regressie van de plaatsscores op de covariabelen.

De gebruikte formule is:

$$\tilde{x}^* := \tilde{x}^* - Z_1' (Z_1 W Z_1')^{-1} Z_1 W \tilde{x}^*$$

3.2. Beoordelen van de Resultaten.

Bij het beoordelen van de resultaten van CCA is van verschillende methoden gebruik gemaakt.

Ten eerste:

De uitvoergegevens van CANOCO zijn beoordeeld. Deze gegevens bieden n.l. een aantal aanknopingspunten op grond waarvan de kwaliteit van de gehele analyse kan worden geschat.

Eigenwaarde van de ordinatie-assen: Hoe groter de eigenwaarde van een as, des te belangrijker is deze as binnen het geheel. Eigenwaardes liggen altijd tussen de 0 en de 1. Normaliter worden de assen in volgorde van afnemende eigenwaarde geëxtraheerd. Binnen vegetatiekundige toepassingen zouden waarden van 0.3 en groter veelvuldig voorkomen (Ter Braak 1987b). Voor 'Weighted Averaging'-methodes geldt dat de eigenwaardes van de assen niet, zoals bij lineaire methodes, eenvoudig als de 'fractie variantie verklaard' beschouwd kunnen worden. De eigenwaarde is een maat voor de spreiding van de soortendistributie langs de ordinatie-assen.

Om toch deze beoordeling te kunnen maken, wordt in de uitvoer het 'percentage variatie verklaard door de eerste s assen van de soorten-omgevingbiplot' gegeven.

VIF = Variance Inflation Factor: wanneer deze grootte een waarde aanneemt groter dan 20 (>20), is interpretatie van de betekenis van de bijbehorende milieuvariabele niet zinvol. Een waarde >20 geeft aan dat de variabele in kwestie binnen de analyse voor $\pm 100\%$ gecorreleerd is met de andere milieuv variabelen. En geen unieke bijdrage aan de analyse heeft. De regressiecoëfficiënt van zo'n variabele zal instabiel zijn. Variabelen met een hoge VIF kunnen zonder nadeel buiten de analyse gelaten worden.

Ten tweede:

Met behulp van CANOPLLOT zijn van het MRC-file (Machine Readable Copy's) van CANOCO, grafieken gemaakt van het resultaat. Deze grafieken zijn geïnterpreteerd en op kwaliteit beoordeeld.

Van de grafieken die CANOPLLOT maakt kan een zogeheten biplot worden gemaakt. In een biplot worden twee soorten variabelen in het zelfde assenstelsel geplott. De assen van het stelsel zijn eenheidsloos. Bij een biplot van resultaten van CCA is het handig de punten die bij de milieuv variabelen horen te beschouwen als de koppen van pijlen vanuit de oorsprong. De lengte van de pijl is (relatief t.o.v. de andere pijlen) een indicatie van het belang van de variabele in de analyse - Hoe langer, hoe belangrijker. De richting van de pijl geeft aan in welke richting in het ordinatiediagram de variabele het sterkst verandert, en dus het duidelijkst zijn invloed laat merken.

Wanneer de soorten in een species/environment-biplot ook als pijlen worden weergegeven (beschouwd), is de hoek tussen de pijl van een soort en de pijl van een milieuv variabele indicierend voor de correlatie tussen de milieuv variabele en de abundantie van de soort. D.w.z. dat bij een kleine hoeken (b.v. $0-45^\circ$) en bij grote hoeken (b.v. $135-180^\circ$) er een duidelijke positieve resp. negatieve correlatie tussen soort en milieuv variabele wordt vermoed. Bij hoeken tussen de pijlen in de buurt van de 90° is het aannemelijk

Canonische Correspondentie Analyse

dat de soort ongecorreleerd is met de betreffende milieuvariabele en dus onafhankelijk van de milieuvariabele.
Het kan wenselijk zijn de milieuvariabelen voor het maken van een species-environment biplot te herschalen.
In het CANOCO manual wordt daarvoor de volgende formule gesuggered:

$(\bar{Y} * \text{biplotwaarde milieuvariabele}) + \text{gemiddelde soort-as}$.

\bar{Y} is in deze formule een gekozen waarde om de biplotwaarde van de milieuvariabele op een prettige wijze in de biplot te kunnen weergeven. Het gemiddelde van de soort-as is nul met uitzondering van analyses waarbij voor 'non-linear rescaling' van de assen is gekozen.

In de uitvoer van CANOCO worden scores voor de soorten (en de opnames) gegeven. Dat zijn niet de werkelijke scores. De berekende scores zijn m.b.v. een multiplier binnen het interval [999, 999] gebracht. Door de scores uit de uitvoertabel te delen door de multiplier is de werkelijke score terug te vinden. De multiplier staat bovenaan in de tabellen. Soorten (Opnames) die in de tabel WEIGHT 0 hebben zijn passieve soorten. Passieve opnames staan in de tabel van 'sample scores' achteraan, na de actieve opnames. In het algemeen geldt dat soorten die dicht bij de rand van het ordinatiediagram staan, een laag gewicht hebben. Deze soorten zijn van weinig invloed op de analyse. En kunnen zonder een sterke verslechtering van de betrouwbaarheid van de analyse eventueel weggelaten worden.

CANOCO schrijft bij de Canonische Correspondentie Analyse twee verschillende MRC-files. Het ene file bevat o.a. species scores en sample scores, regression/canonical coefficients, het andere (MRC of the Environmental Scores) o.a. biplot scores of environmental variables en sample scores, which are linear combinations of environmental variables.

Opgemerkt moet worden dat in beide files 'sample scores'-opnamescores- worden geschreven. Deze opnamescores verschillen duidelijk van elkaar.

In het eerste geval gaat het om opnamescores die zijn afgeleid van de soortssamenstelling van de opnames.

In het tweede geval zijn de opnamescores berekend volgens het gefitte multiple regressie-model (waarbij de soortscores als invoer voor het model worden gebruikt).

De 'sample scores, which are linear combinations of environmental variables' zijn in het algemeen beter gespreid over het diagram. De 'sample scores' hebben de neiging om samen te klonteren in het midden van het diagram.

De uitvoer van CANOCO bevat in sommige gevallen ook waarden van de t-toetsgrootte, van de Student-t-toets. In het CANOCO manual (blz. 49) staat een voorschrift voor het testen van hypothesen aangaande regressiecoëfficiënten m.b.v. deze t-waarden. Canonische coëfficiënten kunnen niet op deze manier op hun significantie getest worden. De t-waarden kunnen wel gebruikt worden als vuistregel. Wanneer de absolute waarde van 't' van een canonische

Canonische Correspondentie Analyse

coëfficiënt kleiner is dan 2.1, is de bijdrage van de variabele aan de "fit" van het model klein.

3.3. Test met kunstmatige dataset

Deze dataset komt uit Kramer 1986, en is onaangepast voor deze test toegepast. Tijdens de analyse is van de gegevens zoals die in tabel 1 staan gebruikt gemaakt.

Van de milieugegevens (Gradiënt1, Gradiënt2 en Bijfactor) is de Bijfactor door CANOCO uitgeschakeld. Deze variabele heeft zo weinig variatie, dat hij voor het programma onbruikbaar is voor de analyse.

In Figuur 2 is het verband te zien tussen de soorten en de beide gradiënten die in de dataset voorkomen. De gradiënten zijn in de biplot weergegeven als pijlen. Deze pijlen geven met hun lengte aan hoe belangrijk de milieuv variabelen zijn binnen de analyse, relatief ten opzichte van elkaar. De richting van de pijlen komt overeen met de richting waarin de bijbehorende milieufactor het sterkst zijn invloed doet blijken. De pijlen kunnen vanzelfsprekend ook in negatieve richting verlengd worden, ook in die richting maakt de betrokken milieufactor zich duidelijk kenbaar. De plaatsing van de soorten ten opzichte van elkaar en ten opzichte van de gradiënten, in Figuur 2, komt goed overeen met de verwachting. Baserend op hoe de dataset is samengesteld wordt de aanwezige structuur door CANOCO goed gereconstrueerd. Soorten 1 t/m 3 langs gradiënt 1, soorten 4 t/m 6 langs gradiënt 2. Soort 7 en Soort 8 laten zich minder goed langs deze beide gradiënten plaatsen. Dit klopt met de opzet van de dataset. Soort 7 wordt niet door de gradiënten beïnvloed, maar door de bijfactor. Soort 8 wordt beïnvloed door de gradiënten, maar kan bij de waarden die ze binnen het 'proefveldje' aannemen overal groeien. Echter op plaatsen waar de Bijfactor sterk is zal Soort 8 minder voorkomen dan elders.

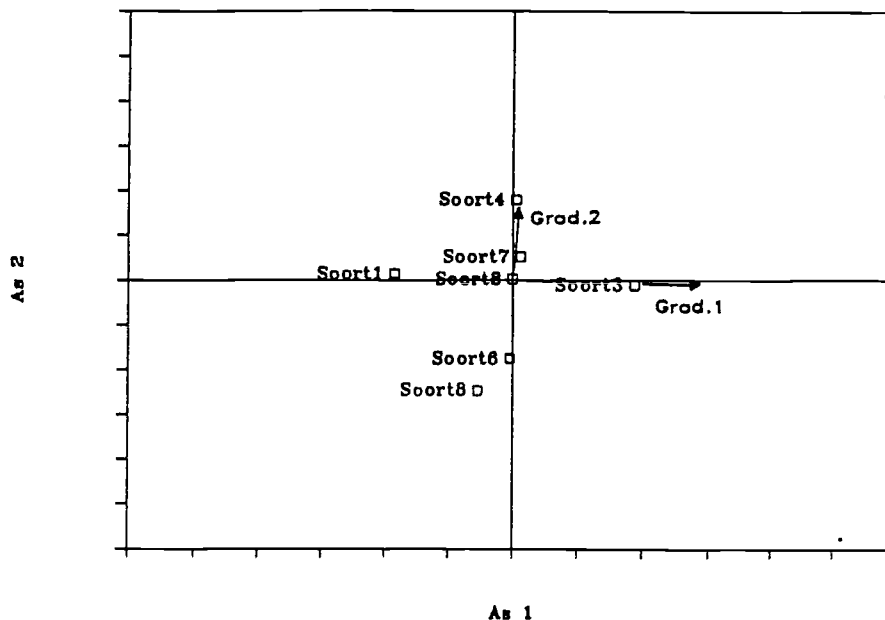
De Bijfactor is door het programma niet in de analyse gebruikt wegens te weinig variatie. Dit neemt niet weg dat de bijfactor toch in de analyse herkend kan worden. In Figuur 3, een biplot van soorten en omgeving van de eerste en de derde ordinatie-as blijken de soorten 7 & 8 de uiterste plaatsen op de derde as in te nemen. De derde ordinatie-as, die door het wegvallen van de Bijfactor een vrije ordinatie-as is geworden, is toch in grote lijnen een as die door de Bijfactor verklaard kan worden.

Dat de bijfactor niet duidelijk in de analyse naar voren komt is ook een gevolg van de gebruikte analysemethodiek. Ordinaties zijn door hun karakter met name gericht op het naar voren halen van factoren/gradiënten die van invloed zijn op 'alle' opnames. Lokaal werkende factoren zoals de bijfactor kunnen het beste worden gevonden met clusteringstechnieken.

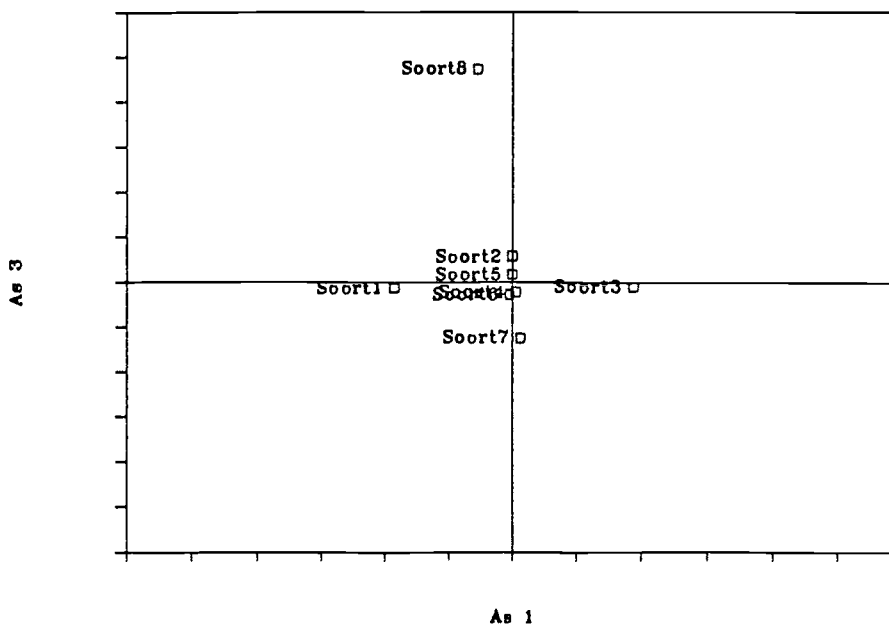
Bij een CCA-analyse worden de ordinatieassen beperkt tot de mogelijkheden geboden binnen het kader van de milieuv variabelen. Indien er minder milieuv variabelen zijn dan assen -minder dan 4- dan zijn hooguit de eerste zoveel ordinatieassen beperkt tot combinaties van de milieuv variabelen. CANOCO extraheerd tijdens de analyses ten hoogste 4 ordinatieassen. Wanneer meer assen gewenst zijn, kunnen deze worden gevonden door de berekende assen als covariabelen te gebruiken bij verdere berekeningen (Ter Braak 1987b).

Canonische Correspondentie Analyse

Figuur 2: Soorten-Omgeving Biplot Kunstmatige Dataset van Ordinateassen 1 en 2. Soorten weergegeven als vierkantjes, met tekst er naast. Milieufactoren als pijlen.1



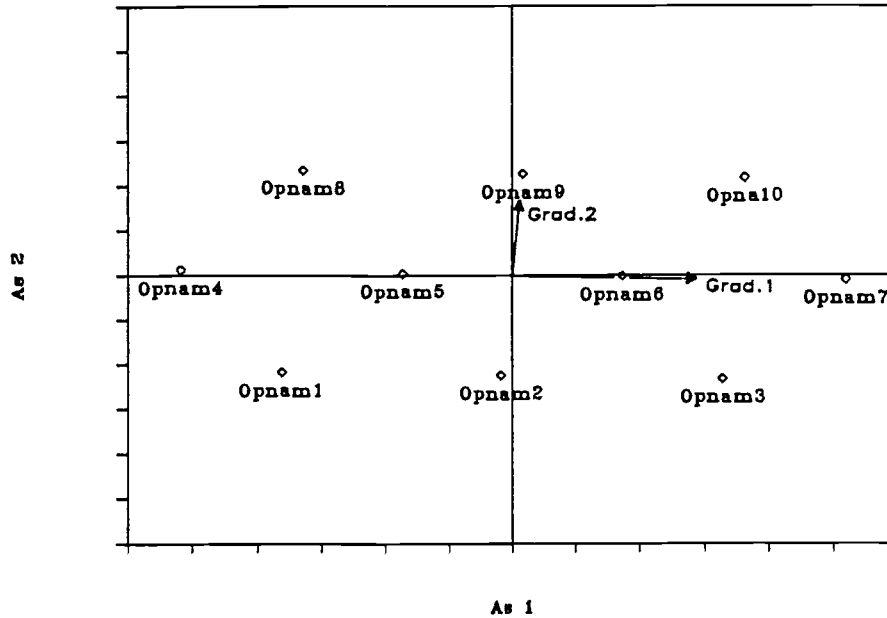
Figuur 3: Soorten-Omgeving Biplot Kunstmatige Dataset van Ordinateassen 1 en 3. Soorten weergegeven als vierkantjes, met tekst er naast. Milieufactoren als pijlen.



Canonische Correspondentie Analyse

Figuur 4: Opnames-Omgeving Biplot Kunstmatige Dataset van Ordinatieassen 1 en 2.

Opnames weergegeven als ruitjes, met tekst er naast. Milieufactoren als pijlen.



De op basis van het gefitte model berekende scores voor de opnames, zoals geplot in Figuur 4, vormen in combinatie met de biplot scores voor de milieuvariabelen een 'kopie' van de afbeelding van de plaats van de opnames op het 'proefveldje' en richting van de gradiënten in Kramer 1986 en Figuur 1. Met dien verstande dat Figuur 4 in de X-as gespiegeld is t.o.v. Figuur 1.

Canonische Correspondentie Analyse

Tabel 2: Resultaatoverzicht Kunstmatige Dataset.

- a) Correlatie matrix
- b) Gemiddelden, Standaarddeviaties en VIF's
- c) Percentages verklaard per as
- d) Eigenwaardes van de assen

a) Gewogen Correlatie Matrix (Gewicht = Opname Totaal)

SPECAX1	1.00								
SPECAX2	0.06	1.00							
SPECAX3	-0.10	-0.66	1.00						
SPECAX4	0.00	0.04	0.00	1.00					
ENVIAX1	0.99	0.00	0.00	0.00	1.00				
ENVIAX2	0.00	0.75	0.00	0.00	0.00	1.00			
ENVIAX3	0.00	0.00	0.00	0.00	0.00	0.00	1.00		
ENVIAX4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	
Grad.1	0.99	-0.03	0.00	0.00	0.99	-0.04	0.00	0.00	1.00
Grad.2	0.04	0.75	0.00	0.00	0.04	0.99	0.00	0.00	1.00
	SPEAX1	SPEAX2	SPEAX3	SPEAX4	ENVAX1	ENVAX2	ENVAX3	ENVAX4	
Grad.1	1.00								
Grad.2	0.00	1.00							
	Grad.1	Grad.2							

b) Variabele	Gewogen Gemiddelde	Standaard Deviatie	Variatie Inflatie Factor
SPEC AX1	0.0000	0.3201	
SPEC AX2	0.0000	0.2324	
SPEC AX3	0.0000	0.3366	
SPEC AX4	0.0000	0.1763	
ENVI AX1	0.0000	0.3181	
ENVI AX2	0.0000	0.1744	
ENVI AX3	0.0000	0.0000	
ENVI AX4	0.0000	0.0000	
Grad.1	5.9976	3.6707	1.0000
Grad.2	3.9895	3.0812	1.0000

c) Percentage variatie verklaard door de eerste S assen van de Soorten-Omgeving Biplot

S	PERC
1	75.7
2	100.0

d) Eigenwaardes van de Ordinatie-assen

Axis 1	EIG=0.092
Axis 2	EIG=0.030
Axis 3	EIG=0.102
Axis 4	EIG=0.030

De structuur van de kunstmatige dataset, die uiteraard volledig bekend is, is door CANOCO goed gereconstrueerd op basis van de opname- en milieugegevens.

De gegevens uit Tabel 2 zijn ontleend aan de uitvoer van CANOCO. De correlaties tussen de Assen en de milieufactoren Tabel 2 a) zijn erg hoog.

Canonische Correspondentie Analyse

De eigenwaardes van de gevonden assen Tabel 2 d) zijn allemaal aan de lage kant, dit doet echter geen afbreuk aan de correctheid waarmee de structuur van de gegevens is gereconstrueerd. Bovendien kunnen eigenwaardes bij CCA niet als 'fractie variatie verklaard' door de bijbehorende as worden beschouwd. Het opmerkenwaard is dat de derde ordinatieas de hoogste eigenwaarde heeft. De oorzaak is dat deze as niet meer beperkt is tot lineaire combinaties van de milieuv variabelen. Deze beperking geldt bij twee milieuv variabelen slechts voor de eerste twee ordinatieassen.

De VIF's van de milieufactoren Tabel 2 b) zijn allen kleiner dan twintig zodat interpretatie van deze factoren binnen de ordinatie zinvol is.

Samen zijn de eerste twee assen voldoende om alle in de dataset aanwezige variatie te kunnen verklaren (Tabel 2 c)).

Van getalsmatige zijde bekeken, is de ordinatie redelijk geslaagd. Redelijk omdat de eigenwaardes van de gevonden assen betrekkelijk laag zijn.

Canonische Correspondentie Analyse

3.4. Test met Kwelderdataset met 'verzonnen' milieugegevens

Deze set gegevens is afkomstig van onderzoek op de Oude Beweide Kwelder (O.B.K.) op Schiermonnikoog (Figuur 5). Het ging hierbij om onderzoek van de vakgroep Gedragsbiologie van de R.U. Groningen, waarbij het gedrag van de Rotgans in het voorjaar bestudeerd werd. De vegetatie is binnen dit onderzoek van groot belang, omdat deze belangrijke aanwijzingen kan geven voor het gedrag van de rotganzen. Het onderzoek aan de rotgans (*Branta bernicla*) door de R.U. Groningen op Schiermonnikoog loopt al vele jaren. De opnames die hier zijn gebruikt zijn in het voorjaar van 1986 gemaakt door R. de Winter en J.T. van der Wal.

Gebaseerd op kennis van terrein is deze set opnames voorzien van fictieve omgevingsgegevens. Deze omgevingsgegevens zijn niet gemeten maar geschat. Daarom zijn de gegevens niet in overeenstemming met de werkelijkheid. Ze geven echter een redelijk beeld van wat er uit zulke metingen zou kunnen komen.

De opnames zijn in een aantal groepen op te delen.

De eerste tweedeling is die tussen opnames afkomstig van de Lage Kwelder ('Puccinellia') en die afkomstig van de Hoge Kwelder ('Festuca'). Deze tweedeling is door gebruik van een covariabele aan het begin van de analyse buiten spel gezet.

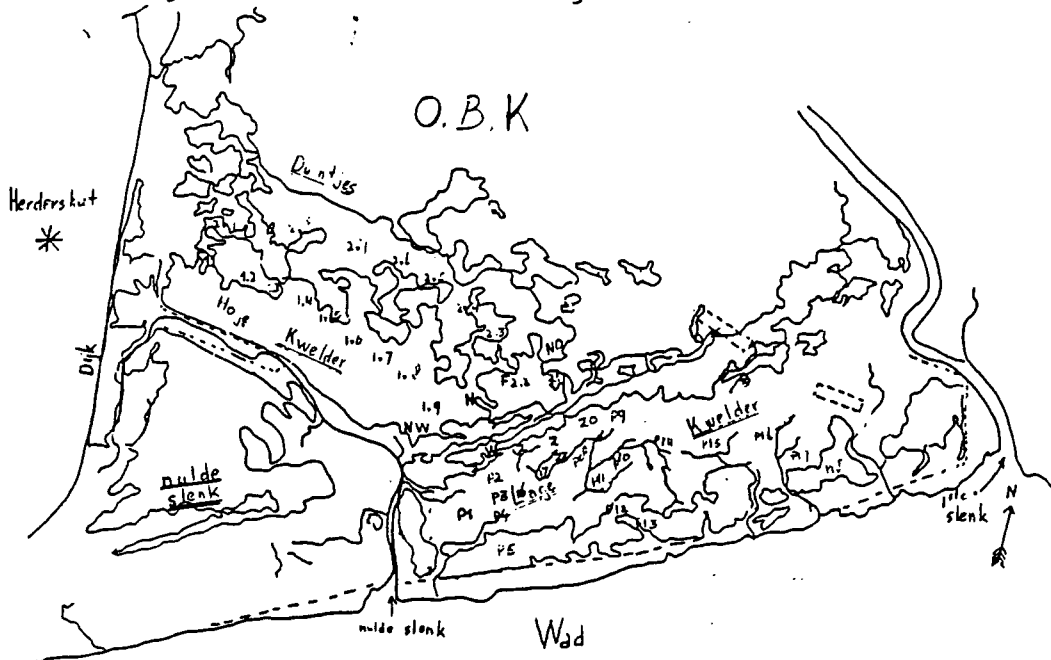
Op de Lage Kwelder zijn opnames gemaakt op een visgraatvormige raai, die van West naar Oost liep, deze opnames zijn gekenmerkt door Pnn (nn is een getal tussen 1 - 18).

Op de Hoge Kwelder zijn de opnames vnl. geplaatst in twee raaien die ongeveer in Noord-Zuid-richting liepen. De eerste (F1n) ten westen van het midden van het Hoge deel, de tweede (F2n) meer naar het oosten onderlangs de Duintjes.

Daarnaast zijn ook opnames gemaakt in een aantal andere vakken op de kwelder, gelegen nabij het westelijk deel van de raai. Deze vakken heten Z, ZZ, ZW, W en ZO. Op de Hoge Kwelder liggen er een aantal grotere vakken die resp. NW, N en NO heten. Deze vakken waren in een sterpatroon rondom een waarnemingstoren die op de kwelder stond gerangschikt en ontleen hun naam aan de richting waarin deze vakken waargenomen werden vanuit die toren. B.V. NO in Noord-Oostelijke richting, ZZ is Zuid-Zuid, een vak ten zuiden van het vak Zuid.

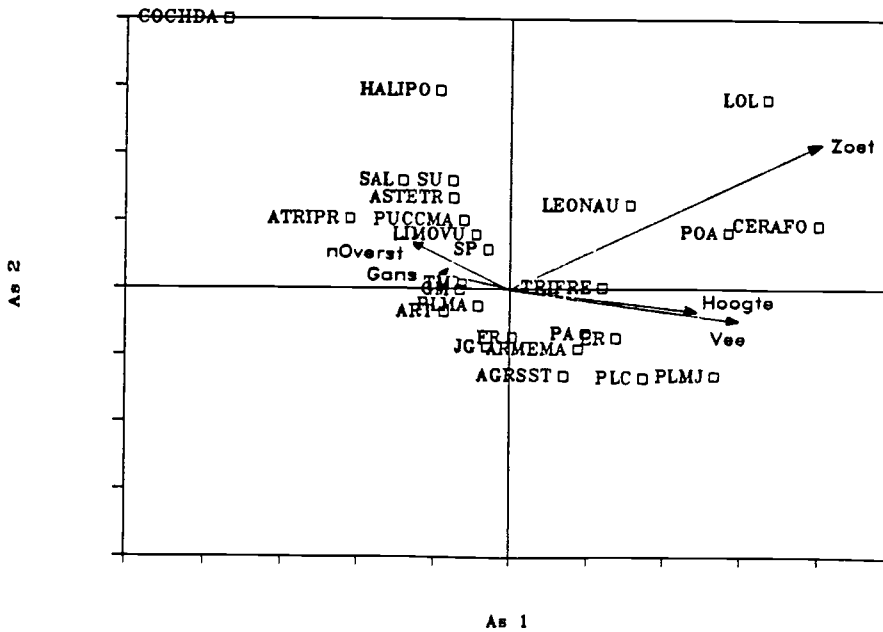
Canonische Correspondentie Analyse

Figuur 5: Kaartje van Schiermonnikoog.



Figuur 6: Soorten-Omgeving Biplot Kwelder-Dataset van Ordinatiessassen 1 en 2.

Soorten weergegeven als vierkantjes, met Soortscodes (6 tekens) ernaast, m.u.v. Tm=Trigma, Gm=Glama, Pa=Potean, Er=Elymre, Jg=Juncge, Fr=Festru, Sal=Salspp, Su=Suaema, Art=Artema, Plc=Planco, Plma=Planma, Plmj=Planmj, Sp=Spegma, Poa=Poaspp, Lol=Lolper. Milieufactoren als pijlen.



Canonische Correspondentie Analyse

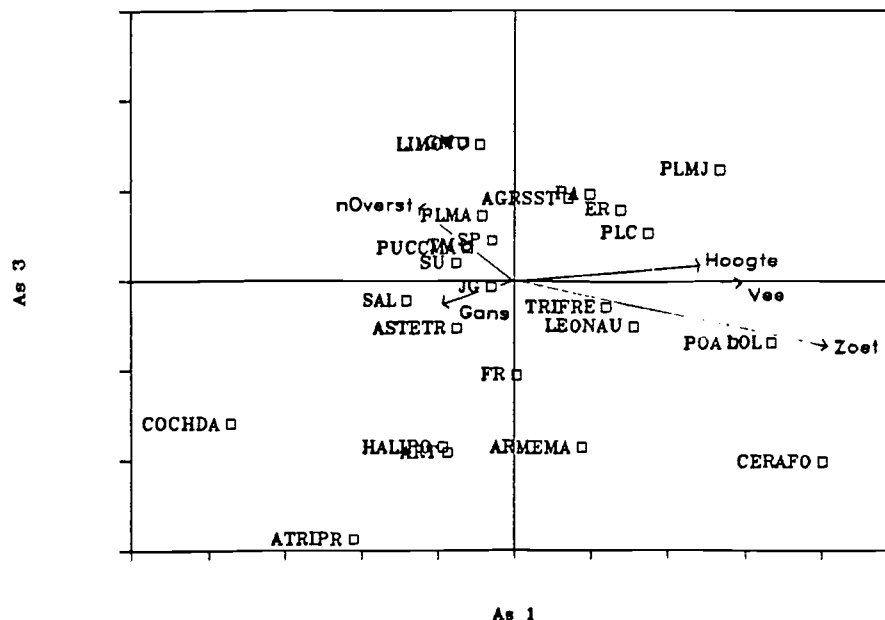
De milieufactoren in Figuur 6 hebben hun belangrijkste component langs de 1^e ordinatieas. Dit is logisch, tussen deze factoren bestaat een duidelijk verband. Op de kwelder bestaat een duidelijk verband tussen de hoogte in het terrein (Hoogte), het aantal overstromingen dat in de loop van het seizoen optreedt (nOverst) en de 'zoetheid' van de standplaats (Zoet).

Ook de richting van de biotische milieugegevens is goed verklaarbaar. Ganzen (rotganzen, branta bernicla) grazen tijdens het voorjaar voornamelijk op het lage deel van de kwelder, dat deel waar de kwelder het meest frequent onder water staat. Het vee, dat de boeren vanaf het late voorjaar tot in het najaar op de kwelder laten grazen, graast meer op de hogere delen van de kwelder. De reden is dat de hogere delen zoeter zijn. En vee moet zoet water, binnendijks, drinken om voor te veel met het voedsel opgenomen zout te compenseren. Het is dus voor het vee gunstig om voornamelijk de zoetere delen van de kwelder te begrazen en op die manier de zoutopname te beperken. Voor de ganzen bestaat deze noodzaak niet, omdat deze dieren een klier hebben waarmee ze overtollig zout kunnen uitscheiden.

Het is goed dat deze verbanden ook door CANOCO tijdens CCA worden opgemerkt.

Figuur 7: Soorten-Omgeving Biplot Kwelder-Dataset van Ordinatiesassen 1 en 3.

Soorten weergegeven als vierkantjes, met Soortscodes (6 tekens) ernaast, m.u.v. Tm=Trigma, Gm=Glama, Pa=Potean, Er=Elymre, Jg=Juncge, Fr=Festru, Sal=Salspp, Su=Suaema, Art=Artema, Plc=Planco, Plma=Planma, Plmj=Planmj, Sp=Spegma, Poa=Poaspp, Lol=Lolper. Milieufactoren als pijlen.



Canonische Correspondentie Analyse

Ook ten opzicht van de derde ordinatieas hebben de milieufactoren hun hoofdcomponent in de zelfde richting als de eerste ordinatie, zoals te zien in Figuur 7.

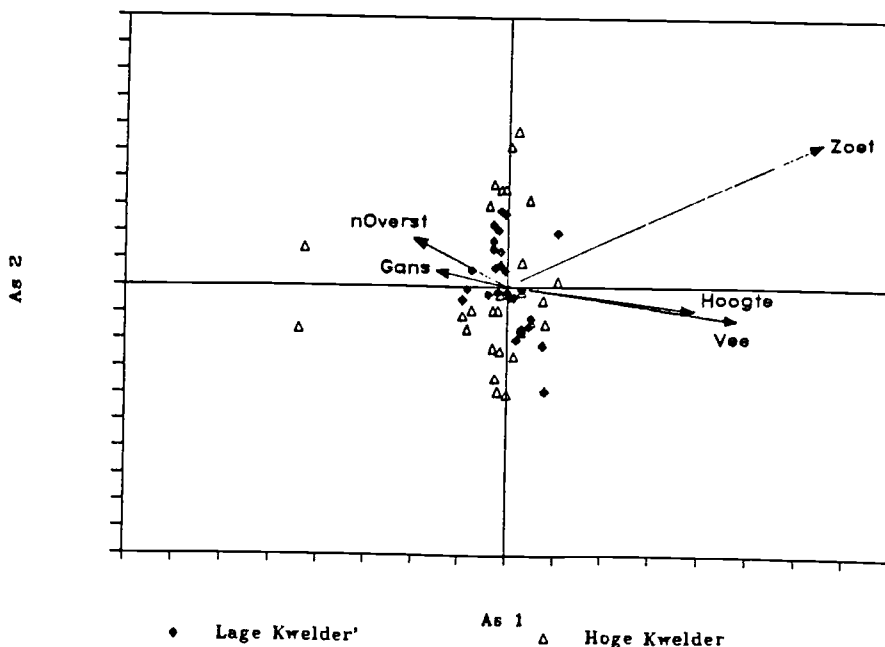
De soorten die in de opnames voorkomen zijn in Figuur 6 en 7 staan meerendeels op een goede d.w.z. verklaarbare plaats. Soorten die te maken hebben met zoete invloeden op de kwelder, *Lolium perenne*, *Poa*-soorten, *Cerastium fontanum* en *Leontodon autumnalis*, staan bij elkaar in de buurt en bij de pijl voor Zoet.

Andere soorten met een bekende voorkeur voor zoute standplaatsen, staan bij elkaar in de buurt rondom de pijl voor Overstromingen gegroepeerd. *Puccinellia maritima*, *Limonium vulgare*, *Triglochin maritima*, *Salicornia*-soorten, *Aster tripolium* etc. zijn soorten die in deze hoek hun plaats vinden.

Ten slotte is er nog een groep betrekkelijk indifferente soorten die met name in Figuur 6 centraal in het ordinatiediagram aan te treffen zijn. Voorbeelden uit deze groep zijn soorten als *Festuca rubra*, *Juncus gerardii*, *Agrostis stolonifera*, *Armeria maritima* en *Plantago*-soorten.

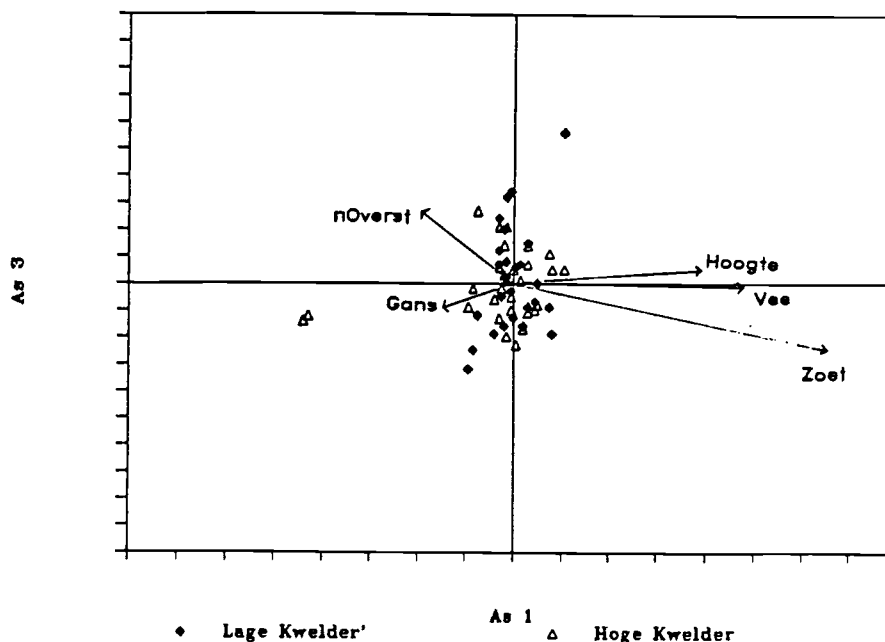
De volgende twee Figuren geven een beeld van het verband tussen de opnames en de milieufactoren. De opnames zijn in deze Figuren opgesplitst in twee groepen. De ene groep is van opnames afkomstig van het lage deel van de OBK. De andere is de groep van opnames afkomstig van het hoge(re) deel van de kwelder.

Figuur 8: Opnames-Omgeving Biplot Kwelder-Dataset van Ordinatiesassen 1 en 2.
Milieufactoren als pijlen.



Canonische Correspondentie Analyse

Figuur 9: Opnames-Omgeving Biplot Kwelder-Dataset van Ordinateassen 1 en 3. Milieufactoren als pijlen.



De lage kwelder opnames zijn in Figuur 8 meer centraal te vinden dan de hoge kwelder opnames. Een probleem bij het plotten van opnames in een biplot is, dat de opnames sterk de neiging hebben in het centrum van het diagram samen te klonteren.

In Ter Braaks 1987b staan ook opmerkingen in deze richting. In Ter Braak '87b is het plotten van opnames (in dezelfde Figuur als soorten en milieufactoren) achterwege gelaten om de Figuur overzichtelijk te houden. Om dit probleem te omzeilen is in Figuren 8 & 9 gebruik gemaakt van de opnamescores op basis van het door het programma gefitte model. Deze waarden hebben minder neiging tot samenklonteren dan de opnamescores puur op basis van de soortssamenstelling.

Als geheel bieden Figuren 8 & 9 geen mogelijkheid te concluderen dat de opnames van de verschillende delen van de kwelder duidelijk van elkaar verschillen. De soorten hebben een dusdanige ecologische amplitude dat binnen deze ordinatie de opnames niet eenduidig in twee groepen uiteenvallen. Meer dan een tendens in Figuur 8 voor de hoge kwelder opnames om richting 'Hoogte', 'Zoet' naast de lage kwelder opnames te liggen kan niet worden opgemerkt.

Canonische Correspondentie Analyse

Tabel 3: Resultaatoverzicht Kwelder Dataset.

- a) Correlatie matrix
- b) Gemiddelden, Standaarddeviaties en VIF's
- c) Percentages verklaard per as
- d) Eigenwaardes van de assen

a) Gewogen Correlatie Matrix (Gewicht = Opname Totaal)

SPECAX1	1.00							
SPECAX2	0.01	1.00						
SPECAX3	-0.02	-0.04	1.00					
SPECAX4	-0.08	0.01	0.06	1.00				
ENVIAX1	0.84	0.00	0.00	0.00	1.00			
ENVIAX2	0.00	0.75	0.00	0.00	0.00	1.00		
ENVIAX3	0.00	0.00	0.59	0.00	0.00	0.00	1.00	
ENVIAX4	0.00	0.00	0.00	0.47	0.00	0.00	0.00	1.00
Hoogte	0.66	-0.15	0.11	0.24	0.79	-0.20	0.18	0.51
nOverst	-0.24	0.31	0.50	-0.08	-0.29	0.41	0.84	-0.18
Zoet	0.73	0.31	-0.15	-0.02	0.87	0.42	-0.26	-0.05
Gans	-0.22	0.12	-0.21	-0.08	-0.27	0.16	-0.35	-0.18
Vee	0.69	-0.32	-0.02	-0.17	0.83	-0.43	-0.03	-0.36
	SPEAX1	SPEAX2	SPEAX3	SPEAX4	ENVAX1	ENVAX2	ENVAX3	ENVAX4

Hoogte	1.0000							
nOverst	-0.2347	1.0000						
Zoet	0.5396	-0.2890	1.0000					
Gans	-0.2119	-0.0613	-0.0286	1.0000				
Vee	0.5553	-0.3759	0.5641	-0.2127	1.0000			
	Hoogte	nOverst	Zoet	Gans	Vee			

b) Var.	Gew.			Gew.		VIF
	Gem.	St. Dev.		Var.	St. Dev.	
SPECAX1	0.00	0.43		Hoogte	17.57	5.00
SPECAX2	0.00	0.29		nOverst	3.84	1.93
SPECAX3	0.00	0.27		Zoet	0.34	2.18
SPECAX4	0.00	0.23		Gans	5.93	2.63
ENVIAX1	0.00	0.36		Vee	2.57	4.07
ENVIAX2	0.00	0.22				
ENVIAX3	0.00	0.16				
ENVIAX4	0.00	0.11				

c) Percentage variatie verklaard door de eerste S assen van de Soorten-Omgeving Biplot

S	PERC
1	55.9
2	77.9
3	90.4
4	96.1

d) Eigenwaardes van de Ordinatie-assen

Axis 1	EIG=0.114
Axis 2	EIG=0.045
Axis 3	EIG=0.025
Axis 4	EIG=0.012

Canonische Correspondentie Analyse

De getallen in Tabel 3: Resultaatoverzicht Kwelder Dataset, vormen een beeld, dat vergelijkbaar is met Tabel 2. De belangrijkste overeenkomsten zijn :

- 1) lage eigenwaardes voor de geëxtraheerde assen
- 2) hoge percentages variatie verklaard door de eerste assen
- 3) VIF (Variatie Inflatie Factor) kleiner dan 20
- 4) goede correlaties tussen assen en variabelen.

Ook het geheel van deze ordinatie vormt al met al toch het beeld van een redelijk geslaagde analyse.

3.5. Gebruik van CANOCO in de praktijk.

Door gebruik van CANOCO zijn enkele punten naar voren gekomen, die binnen het kader van dit verslag enige nadere aandacht verdienen.

3.5.1. Gebruik CANOCO.con

Tijdens het werken met CANOCO wordt door het programma in de files CANOCO.con en CANOC2.con bijgehouden wat de gebruiker antwoordt op de vragen. Deze '.con'-files zijn normale tekstfiles die met een editor bewerkt kunnen worden en die van begeleidend commentaar worden voorzien door het programma.

Wanneer een bepaalde combinatie van antwoorden (d.w.z. een bepaalde analyse) herhaald moet worden, of wanneer zo'n analyse vaker zal moeten uitgevoerd -op dezelfde wijze door verschillende personen- kan dat vergemakkelijkt worden door de 'con'-files na het uitvoeren van een "sjabloon"-analyse te hernoemen. Zo'n hernoemd '.con'-file krijgt dan, vanzelfsprekend, een naam die aangeeft wat voor analyse erin is opgeslagen. Een kopie van dit file kan m.b.v. een editor aangepast worden (namen van de files e.d.), waarna automatisch de juiste analyse wordt uitgevoerd. Bij het editen van een '.con'-file kan het aanwezige commentaar als gids dienen bij de aanbrengen wijzigingen. Hierbij zijn toevallige tikfouten en andere vergissingen bij het beantwoorden van de vragen en het kiezen van opties uitgesloten.

3.5.2. CANOPLLOT en Machine Readable Copy

CANOCO vraagt tijdens de loop van het programma of de gebruiker al of niet het schrijven van een Machine Readable Copy (MRC) wenst. Wanneer na afloop van de analyse met de hulp van CANOPLLOT grafieken gemaakt, worden moet een Machine Readable Copy beschikbaar zijn, die afkomstig is van de juiste analyse.

Wanneer een MRC geschreven zal worden vraagt CANOCO naar een naam voor deze (dit) file(s). Afhankelijk van het type analyse worden 1 of 2 MRC-files geschreven. Bij DCA is er slechts een zo'n file, bij CCA zijn er twee '.MRC'-files.

In elk geval is er een file voor de 'Species Scores' en de 'Sample Scores'. Indien nog wordt een file met gegevens over de 'Environmental Variables' geschreven. Als extensie voor deze files, waar geen default voor bestaat, is '.mrc' aan te bevelen.

Canonische Correspondentie Analyse

Tabel 4: Overzicht inhoud CANOCO Machine Readable Copy-files.

1 ^e file:	species scores
	sample scores
	regr./canonical coefficients
	t-values regression coefficients
	interset correlation env. vars' =><= axes
2 ^e file:	biplot scores env. var's
	centroids of env. var's
	sample scores which are linear combinations of environmental variables

3.5.3. Actief versus Passief

CANOCO biedt de mogelijkheid om soorten en opnames passief te maken door ze een minimaal 'gewicht' toe te kennen. In sommige gevallen kan dit nuttig zijn. Zo kan op deze manier voorkomen dat bepaalde soorten of opnames de analyse bemoeilijken zonder dat het nodig is om deze soorten of opnames volledig buiten de analyse te laten.

Het is mogelijk dat opnames, om een onduidelijke reden en ongevraagd, door het programma passief worden gemaakt. Het gevolg hiervan is dat de opnames geen invloed uitoefenen op het resultaat, deze opnames worden na afloop van de analyse in de ordinatie geplaatst op basis hun samenstelling. De passieve opnames worden in de tabel van de resultaten na de actieve opnames geplaatst.

In extreme gevallen van passief maken van opnames door CANOCO blijven nul opnames over. In deze gevallen stopt het programma met een foutmelding. Deze foutmelding geeft aan dat een fout is opgetreden bij het aanpassen van de grootte van een 'adjustable array'.

De oorzaak van dit verschijnsel is te vinden in de CANOCO-Handleiding in paragraaf 4.1 blz 39. Wanneer het programma verschillen opmerkt in de naamgeving/nummering van een opname tussen het file met de opnamegegevens en het file met de milieugegevens wordt de opname passief gemaakt. Zorgvuldigheid bij het maken van de invoerfiles is daarom geboden. Een spatie meer of minder tussen de namen van de opnames kan veel ellende veroorzaken.

Canonische Correspondentie Analyse

4. Conclusies.

Voor dit verslag is met twee verschillende datasets Canonische Correspondentie Analyse uitgevoerd. Voor het uitvoeren van deze analyse is gebruik gemaakt van CANOCO, een computerprogramma voor het verwerken van vegetatiekundige datasets.

In beide gevallen leverde analyse ordinatiediagrammen (biplots) op die goed interpreteerbaar zijn. Bij deze datasets zijn de getalsmatige toetsingsmogelijkheden voor de kwaliteit in orde, met uitzondering, wellicht, van de eigenwaardes van de ordinatienassen. Deze eigenwaardes zijn betrekkelijk laag (<0.30). In Ter Braak 1987b wordt aangegeven dat eigenwaardes, bij oecologische toepassingen, vaak groter dan 0.30 zijn. Daar tegenover staat dat eigenwaardes van de assen bij 'Weighted Averaging'-methodes, als CCA, niet simpelweg als de 'fractie variatie verklaard' beschouwd worden. Op grond daarvan is het waarschijnlijk niet nodig zwaar te tillen aan de betrekkelijk lage eigenwaardes. En is het beter meer aandacht te schenken aan de 'percentages variatie verklaard door de eerste "s" assen' van de ordinatie.

Geconcludeerd kan worden dat de CCA-optie van CANOCO een handige verwerkingsmethode is bij de verwerking van vegetatiekundige datasets, wanneer zowel opnames als omgevingsgegevens beschikbaar zijn. In één bewerking worden soorten en opnames geordineerd en is bekend in welke mate en richting factoren vanuit de omgeving de vegetatie beïnvloeden. Nuttig is de uitvoer van Student-t-toetswaarden door het programma, waardoor het mogelijk is de coëfficiënten op significantie te testen. Zolang het om regressiecoëfficiënten gaat, canonische coëfficiënten kunnen niet getest worden, maar in het CANOCO-manual wordt een vuistregel hiervoor gegeven. Zodat de betrouwbaarheid van canonische coëfficiënten beoordeeld kan worden.

Het door Ter Braak (1986) geïntroduceerde type biplot, waarin de omgevingsfactoren als pijlen worden weergegeven, is na wat, gewenning een prettige en overzichtelijke wijze om de resultaten van de ordinatie weer te geven. Het is een wijze van weergeven, die het mogelijk maakt veel informatie op een duidelijke wijze in een diagram te stoppen.

Opgemerkt moet worden dat beide datasets fictief of deels fictief zijn. De eerste dataset (Kramer, 1986) is volledig fictief, de interpretatie van de resultaten van de ordinatie van deze gegevens is niet moeilijk. Dit omdat van deze gegevens alles bekend is, en bovendien is deze set gegevens moedwillig eenvoudig gehouden. Van de tweede -Kwelder- set is de reeks milieugegevens fictief, met uitzondering van de covariabele LageKwelder, de andere factoren zijn geschat, naar aanleiding van kennis van het gebied. Bij de schatting van de milieugegevens is de plaats van de opname op de kwelder, geografische ligging, gebruikt. De schatting van de factoren is dus niet op de soortensamenstelling van de opnames gebaseerd. Door het schatten deze gegevens betrekkelijk eenvoudig. Waardoor ook de interpretatie van de gegevens makkelijk is.

De test van Canonische Correspondentie Analyse zoals die in dit verslag besproken wordt, is niet volledig. Bij het testen is een

Canonische Correspondentie Analyse

aspect als 'ruis' of 'noise' in de gegevens niet aan de orde gekomen. Hoe goed CCA bestand is tegen ruis, valt dus nog niet te zeggen. Door de test te herhalen, met een volledig fictieve dataset kan de 'ruis'-bestendigheid van CCA getest worden. De dataset moet dan kunstmatig van 'ruis' worden voorzien. Dezelfde analyse uitvoeren op dataset met verschillende hoeveelheden 'ruis', kan de 'ruis'-vraag beantwoorden.

Het kan blijken, dat ruis van grote invloed is op de uitkomsten van CCA. Een mogelijkheid tot verbetering van de resultaten bestaat. Door opties die CANOCO biedt binnen Canonische Correspondentie Analyse anders te kiezen, kan het eindresultaat van de analyse verbeteren. Of de resultaten van de analyse door anders kiezen van opties verbeterd konden worden, is binnen dit onderzoek niet aan de orde gekomen. Een verbetering van de analyse door andere opties te kiezen, is mogelijk een zaak die afhangt van de gegevens die verwerkt worden. Wat optimaal is voor de meeste gegevens, hoeft niet noodzakelijkerwijs optimaal te zijn voor andere gegevens.

Mogelijkheden die er binnen CCA zijn om de analyse te beïnvloeden zijn o.a. data-transformatie, 'weights' voor opnames/soorten en 'downweighting of rare species'.

Bij dit onderzoek zijn geen 'weights' toegekend aan bepaalde soorten of opnamen. De analyses zijn beide uitgevoerd met wortel-transformatie en zonder 'downweighting of rare species'. Voor wortel-transformatie is gekozen, in verband met de gebruikte schattingsschaal voor de bedekkingen [0-100]. Soorten met een erg hoge bedekking kunnen dan minder makkelijk de ordinatie misvormen. Downweighting is niet gebruikt, omdat in beide datasets geen sprake was van zeldzame soorten, die in verhoudingsgewijs weinig opnames voorkwamen. Andere invloeden van de soortenrijkdom op de analyse is geen rekening meegehouden.

Al met al is Canonische Correspondentie Analyse, zoals geïmplementeerd in CANOCO, een krachtige analysemethode. Waarvan de resultaten op een overzichtelijke wijze in een ordinatiediagram kunnen worden weergegeven.

Canonische Correspondentie Analyse

Literatuurlijst.

Braak C.J.F. Ter (1986)
Canonical Correspondance Analysis; A new eigenvector technique for
multivariate direct gradient analysis
Ecology 67(5), 1167-1179

Braak C.J.F. Ter (1987a)
Canoco - a FORTRAN program for canonical community ordination by
[partial][detrended] correspondance analysis, principal compo-
nents analysis and redundancy analysis (version 2.1).
IWIS-TNO, Wageningen

Braak C.J.F. Ter (1985)
Correspondance Analysis of Incidence and Abundance Data: Proper-
ties in Terms of a Unimodal Response Model.
IWIS-TNO, Wageningen, report

Braak C.J.F. Ter (1987b)
The analysis of vegetation-environment relationships by canonical
correspondence analysis
Vegetatio 69 (in press)

Braak C.J.F. ter, Looman, C.W.N. (1987)
Weighted averaging, logistic regression and the Gaussian response
model.
Vegetatio (in press)

Braak C.J.F., Colin Prentice, I. (1986)
A theory of gradient analysis
ITI-TNO Wageningen

Forsythe G.E. et al.
Computer methods for mathematial computations
Prentice-Hall, Inc., Englewood Cliffs, N.J. 07632

Gauch H.G. jr. (1982)
Multivariate analysis in community ecology
Cambridge University Press

Heiser Willem J. (1986)
Joint Ordination of Species and Sites: The Unfolding Technique.
New Dev.s in Num. Ecol. ed. P. Legendre & L. Legendre
Springer Verlag, Berlin

Hill M.O. (1973)
Reciprocal Averaging: an eigenvector method of ordination.
J. of Ecology, 63, 237-250

Hill M.O., Gauch H.G. jr. (1980)
Detrended Correspondance Analysis: an improved ordination
technique.
Vegetatio 41, 47-58

Canonische Correspondentie Analyse

Kershaw K.A., Looney J.H.H. (1985)
Quantative and Dynamic Plant Ecology (3.ed)
Edward Arnolds Lim., London

Kramer, Koen (1986)
Evaluatie van een aantal hierarchisch-polythetisch-divisie
clustermethoden met betrekking tot de verwerking van vegetatie??
doctoraal verslag, R.U.G.

Legendre L. & Legendre P. (1983)
Numerical Ecology.
Elsevier Scientific Publ. Comp., Amsterdam

Lindeman R.H., Merenda P.F., Gold, R.Z. (1980)
Introduction to multivariate analysis
Scott, Foresman & Company

Orloci L. (1975)
Multivariate analysis in vegetation research.
Junk, Den Haag

Wal, J.T. van der (1987)
VG-Pack Manual
handleiding, R.U. Groningen

Wal, J.T. van der (1987)
VG-Pack Programmers Manual
handleiding, R.U. Groningen

Whittaker R. (ed.), Tuxen R. (ed.) (1973)
Handbook of Vegetation Science (V) Ordination and Classification
of Communities
Junk, Den Haag

Wiegand, Gerhard (1986)
Grenzen und Moglichkeiten der Datenanalyse in der PflanzenOkolo-
gie.
Tuexenia 6, 365-377

Canonische Correspondentie Analyse

Lijst van Figuren.

Figuur 1: Geografie van de Kunstmatige Dataset.	4
Figuur 2: Soorten-Omgeving Biplot Kunstmatige Dataset van Ordina- tieassen 1 en 2.	11
Figuur 3: Soorten-Omgeving Biplot Kunstmatige Dataset van Ordina- tieassen 1 en 3.	11
Figuur 4: Opnames-Omgeving Biplot Kunstmatige Dataset van Ordina- tieassen 1 en 2.	12
Figuur 5: Kaartje van Schiermonnikoog.	16
Figuur 6: Soorten-Omgeving Biplot Kwelder-Dataset van Ordina- tieassen 1 en 2.	16
Figuur 7: Soorten-Omgeving Biplot Kwelder-Dataset van Ordina- tieassen 1 en 3.	17
Figuur 8: Opnames-Omgeving Biplot Kwelder-Dataset van Ordina- tieassen 1 en 2.	18
Figuur 9: Opnames-Omgeving Biplot Kwelder-Dataset van Ordina- tieassen 1 en 3.	19

Lijst van Tabellen.

Tabel 1: Kunstmatige Dataset (Kramer, 1986)	3
Tabel 2: Resultaatoverzicht Kunstmatige Dataset. a) Correlatie matrix b) Gemiddelden, Standaarddeviaties en VIF's c) Percentages verklaard per as d) Eigenwaardes van de assen	13
Tabel 3: Resultaatoverzicht Kwelder Dataset. a) Correlatie matrix b) Gemiddelden, Standaarddeviaties en VIF's c) Percentages verklaard per as d) Eigenwaardes van de assen	20
Tabel 4: Overzicht inhoud CANOCO Machine Readable Copy-files. .	22

Canonische Correspondentie Analyse

Index

.con (21)
.mrc (21)
Biplot (8)
CANOCO
 M.R.C. (21)
 Machine Readable Copy (21)
Eigenwaarde (8)
Geen actieve 'samples' (22)
Machine Readable Copy (8)
MRC (8)
Multiplier (9)
Opzet van de kunstmatige dataset. (3)
Student-t-toets (9)
Variance Inflation Factor (8)
VIF (8)
WEIGHT (9)

Canonische Correspondentie Analyse

Bijlage 1: Test Dataset 2.

Op de volgende bladzijden staan de gegevens van test data set 2, weergegevens als in een Full Format gegevensbestand. Full Format is een wijze van bestandsopbouw nauw verwant aan Cornell Condensed, en het is een opbouw die door veel statistische pakketten wordt gebruikt (Ter Braak 1987a). Voor meer informatie over de opbouw van gegevensbestanden zie de CANOCO-handleiding (Ter Braak 1987a) en de handleiding voor VG-Pack (van der Wal 1987).

Canonische Correspondentie Analyse

Canoco test dataset 2: Kwelderopnames Rotgansonderzoek voorjaar 1986

(I2,X,18(F2.0,X)/3X,8(F2.0,X))

18

1	1	1	30	40	10	1	1	0	1	5	10	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0									
2	1	1	50	50	5	1	5	0	1	5	20	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0									
3	1	0	50	5	5	0	1	20	1	1	1	1	0	0	0	0	0
3	0	0	0	0	0	0	0	0									
4	1	0	40	10	10	0	1	10	1	1	5	1	0	0	0	0	0
4	0	0	0	0	0	0	0	0									
5	0	0	40	30	10	1	1	30	1	1	5	5	0	0	0	0	0
5	0	0	0	0	0	1	0	0									
6	1	1	40	0	10	1	5	10	1	5	10	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0									
7	1	1	70	5	10	1	1	5	1	5	10	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0									
8	5	5	30	20	1	1	1	5	5	1	1	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0									
9	1	1	50	10	1	1	1	5	1	10	20	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0									
10	5	1	40	5	1	1	1	30	5	1	5	1	0	0	0	0	0
10	0	0	0	0	0	0	0	0									
11	1	1	20	30	5	1	1	60	1	1	1	1	0	0	0	0	0
11	0	0	0	0	0	0	0	0									
12	1	1	30	10	10	1	1	40	1	1	5	1	0	0	0	0	0
12	0	0	0	0	0	0	0	0									
13	1	1	40	50	5	1	1	40	1	1	1	1	0	0	0	0	0
13	0	0	0	0	0	0	0	0									
14	1	1	30	10	10	1	1	30	1	5	10	1	0	0	0	0	0
14	0	0	0	0	0	0	0	0									
15	1	1	40	1	5	1	1	40	1	10	10	1	0	0	0	0	0
15	0	0	0	0	0	0	0	0									
16	1	1	5	1	1	1	1	90	0	1	1	5	5	0	0	0	0
16	0	0	0	0	0	0	0	0									
17	1	0	40	10	5	1	5	40	1	1	1	5	5	0	0	0	0
17	0	0	0	0	0	0	0	0									
18	1	1	10	5	5	1	60	5	1	0	5	0	20	0	0	0	0
18	0	0	0	0	0	0	1	0									
19	1	1	30	10	5	1	20	30	1	1	5	0	10	0	0	0	0
19	0	0	0	0	0	0	0	0									
20	0	1	5	10	5	1	1	80	1	5	5	10	10	0	0	0	0
20	0	0	0	0	0	1	0	0									
21	1	1	5	5	1	1	40	50	1	1	1	0	5	0	0	0	0
21	0	0	0	0	0	0	0	0									
22	1	1	20	20	1	1	20	40	1	1	1	1	5	0	0	0	0
22	0	0	0	0	0	0	0	0									
23	1	1	40	1	1	1	10	40	1	1	1	5	10	0	0	0	0
23	0	0	0	0	0	0	0	0									
24	1	1	30	10	5	1	5	40	1	5	10	10	5	0	0	0	0
24	0	0	0	0	0	0	0	0									
25	1	1	30	5	20	1	1	40	1	10	20	1	1	0	0	0	0
25	0	0	0	0	0	0	0	0									
26	1	1	40	20	10	1	20	10	1	5	10	1	1	0	0	0	0
26	0	0	0	0	0	0	1	0									

Canonische Correspondentie Analyse

27	1	1	30	10	10	1	10	40	1	1	1	1	5	0	0	0	0	0
27	0	0	0	0	0	0	1	0										
28	1	1	1	1	1	1	60	30	1	1	1	0	40	1	0	0	0	0
28	0	0	0	0	0	1	5	0										
29	1	1	5	1	5	1	20	1	1	5	1	0	30	1	0	0	1	0
29	0	0	0	0	0	1	10	1										
30	10	1	20	5	1	1	0	0	1	0	0	0	20	0	0	0	0	0
30	0	0	0	0	0	0	0	1										
31	1	0	20	1	1	0	1	50	1	1	1	0	10	5	1	1	0	0
31	0	0	1	0	0	0	0	0										
32	1	0	5	1	1	0	1	50	1	1	10	0	5	10	10	1	0	0
32	0	0	0	0	0	0	1	0										
33	1	0	0	0	5	0	1	20	1	5	1	0	30	20	0	10	10	0
33	0	0	0	0	0	0	1	0										
34	0	1	0	0	0	1	30	10	1	5	1	0	30	5	5	20	5	5
34	1	0	1	0	0	0	5	1										
35	0	0	0	0	1	0	1	20	1	1	5	0	50	10	5	5	5	1
35	0	0	0	0	0	0	0	0										
36	0	1	1	0	0	0	1	40	1	1	1	0	20	5	5	5	0	10
36	5	0	0	1	0	0	0	0										
37	0	0	0	0	1	0	1	30	0	1	1	0	30	10	1	1	30	40
37	5	0	1	1	0	0	0	0										
38	0	0	0	1	0	0	1	30	1	1	5	1	30	10	1	5	20	0
38	0	0	1	0	0	0	0	0										
39	0	1	0	0	0	0	1	20	5	5	5	0	30	30	1	5	10	0
39	0	0	1	0	0	0	0	0										
40	0	0	0	1	0	0	5	30	1	1	1	1	30	10	5	10	0	10
40	0	0	1	0	0	0	0	0										
41	0	0	0	1	1	0	1	10	1	1	1	0	60	1	1	0	10	0
41	0	0	1	0	0	0	0	0										
42	0	0	0	0	0	0	1	70	0	1	1	0	5	20	1	10	5	1
42	1	0	0	0	0	0	0	0										
43	0	0	0	0	1	0	1	20	1	1	10	0	40	30	1	5	0	0
43	0	0	0	0	0	0	0	0										
44	0	0	0	0	0	0	1	50	5	0	5	0	10	20	1	20	5	1
44	0	0	1	0	0	0	0	0										
45	0	0	0	0	0	0	1	5	0	1	0	0	50	30	1	1	5	10
45	1	1	5	0	1	0	1	5										
46	0	0	0	0	0	0	5	0	1	1	0	5	50	10	5	1	20	10
46	1	1	1	1	0	0	1	0										
47	0	1	0	0	0	1	1	5	1	0	0	0	20	10	5	10	10	50
47	0	1	0	0	0	0	0	0										
48	0	0	0	1	1	0	10	20	1	1	0	0	50	10	1	0	10	0
48	0	0	1	0	0	0	0	0										
49	0	0	0	0	0	0	5	0	1	5	0	1	40	20	5	0	10	10
49	5	1	1	0	0	0	5	0										
50	0	0	0	0	0	0	1	5	0	0	0	1	60	10	5	1	10	10
50	1	1	5	0	1	0	0	0										
51	0	0	0	0	0	0	1	30	0	1	0	0	10	10	1	50	0	5
51	5	0	1	0	0	0	0	1										
52	0	0	0	0	0	0	0	0	1	5	0	1	40	5	5	1	5	0
52	0	1	1	5	1	0	0	0										
53	0	0	0	0	0	0	1	30	0	1	0	1	40	10	1	0	10	0
53	10	0	1	0	0	0	0	0										
54	0	0	0	0	0	0	0	5	1	0	0	1	20	5	5	0	20	50
54	1	0	1	0	0	0	0	0										

Canonische Correspondentie Analyse

55 0 0 0 0 0 0 0 1 0 0 0 20 10 5 5 5 50
 55 10 1 1 0 1 0 0 0
 56 0 0 0 0 0 1 10 0 1 0 1 60 10 1 0 10 0
 56 0 0 1 0 0 0 0 0
 0000

SALSPP SUAEMA PUCOMA LIMOVU TRIGMA ASTETR ARTEMA JUNCGE
 SPEGMA PLANMA GLAXMA ARMEMA FESTRU AGRSST TRIFRE POTEAN
 ELYMRE LOLPER POASPP CERAFO PLANCO LEONAU PLANMJ COCHDA
 ATRIPR HALIPO

P2	P1	ZO	Z	P8	P5	A	P17	P3	P15	P11	P7	B
ZW	W	P10	P14	P4	P13	C	ZZ	P18	P16	P9	D	P6
P12	E	F	G	F16	F18	F11	H	F12	F25	F13	NW	F24
N F15	J	F14	F22	K	F21	F28	F17	F23	L	NO	F26	
F19	F20	F27	F110									

Canonische Correspondentie Analyse

Canoco test dataset 2: Kwelder artif. milieugeg. Rotgansond
 voorjaar 1986
 (I2,x,6(f3.0))

6							
1	1	2	8	0	10	1	
2	1	2	11	0	10	1	
3	1	3	7	0	15	2	
4	1	3	5	0	5	1	
5	1	3	10	0	12	1	
6	1	1	7	0	8	1	
7	1	2	7	0	7	1	
8	1	0	13	0	2	2	
9	1	3	11	0	9	1	
10		1	1	5	0	14	1
11		1	3	8	0	12	2
12		1	2	9	0	9	1
13		1	4	6	0	8	1
14		1	1	5	0	6	1
15		1	4	3	0	6	1
16		1	3	7	0	14	2
17		1	4	4	0	10	2
18		1	8	3	0	6	2
19		1	4	3	0	9	1
20		1	3	4	0	7	2
21		1	3	3	0	2	1
22		1	2	2	0	15	1
23		1	1	3	0	14	1
24		1	6	4	0	12	2
25		1	4	5	0	8	2
26		1	3	2	0	11	3
27		1	5	3	0	12	2
28		0	9	4	0	9	1
29		0	7	3	0	8	1
30		0	8	3	0	7	0
31		0	12	3	0	4	3
32		0	25	3	0	2	3
33		0	45	3	0	1	3
34		0	15	3	0	0	3
35		0	13	4	0	1	3
36		0	16	3	1	2	3
37		0	35	5	0	1	3
38		0	37	3	0	3	3
39		0	29	4	0	4	3
40		0	47	3	0	3	3
41		0	26	1	0	2	4
42		0	34	1	0	2	4
43		0	19	2	0	1	4
44		0	43	2	1	1	4
45		0	48	2	1	1	4
46		0	49	1	2	2	4
47		0	46	1	2	4	4
48		0	35	1	0	1	4
49		0	26	1	1	3	5
50		0	34	1	1	2	5
51		0	49	0	2	5	5
52		0	28	1	1	4	5

Canonische Correspondentie Analyse

53	0	36	2	1	3	5
54	0	41	1	3	2	5
55	0	43	0	2	2	5
56	0	27	0	1	3	5

0000

LageKw	Hoogte	nOverst	Zoet	Gans	Vee							
P2	P1	ZO	Z	P8	P5	A	P17	P3	P15	P11	P7	B
ZW	W	P10	P14	P4	P13	C	ZZ	P18	P16	P9	D	P6
P12	E	F	G	F16	F18	F11	H	F12	F25	F13	NW	F24
N F15	J	F14	F22	K	F21	F28	F17	F23	L	NO	F26	
F19	F20	F27	F110									

Canonische Correspondentie Analyse

Bijlage 2: Numerieke Programmatuur.

CANOCO lost met behulp van numerieke rekenmethoden een wiskundig probleem op. Dit probleem laat zich niet op een andere wijze oplossen, omdat bij een algebraïsche aanpak op allerlei plaatsen min of meer ernstige onnauwkeurigheden kunnen optreden. Indien een wiskundig probleem met numerieke rekenmethoden wordt opgelost, is het eigenlijk niet correct van een oplossing te bespreken. In feite wordt een goede, betrouwbare benadering van de werkelijke oplossing berekend.

Andere programmatuur die numerieke problemen kan verwerken is b.v. de NAG-reeks van routines. NAG bestaat uit een reeks (gecompileerde) Fortran-routines, die aan programma's kunnen worden gekoppeld. Op de CYBER en de VAX kunnen dit ook Pascal-programma's of b.v. Cobol-programma's zijn. NAG is ook beschikbaar voor de P.C. Aan de R.U.G. is NAG beschikbaar op de CYBER en de VAX8600.

Op de CYBER is een andere minder uitgebreide serie routines voor numerieke bewerkingen beschikbaar, als sourcecode. Deze serie heet GEF. Van GEF is een pascal-source beschikbaar, deze pascal-source is een machinevertaling van de oorspronkelijke fortran-source. Deze pascal-source, is geschreven in ISO-Pascal, en zal makkelijk in een willekeurig programma kunnen worden opgenomen.

Om de pascalsource op de CYBER beschikbaar te krijgen moet de volgende stuuropdracht worden ingegeven:

```
GET,PAS/UN=GEF
```

De GEF-bibliotheek is gebaseerd op fortran-routines die behandeld worden in:

Computer methods for mathematical computations

George E. Forsythe, Michael A. Malcolm, Cleve B. Moler.

Prentice-Hall, Inc., Englewood Cliffs, N.J. 07632

De naam is afgeleid van de initialen van de eerste auteur. Door medewerkers van het R.C. zijn in de loop der jaren routines aan GEF toegevoegd van andere oorsprong dan het hiervoor genoemde boek.

Canonische Correspondentie Analyse

Bijlage 3: Productie van Grafieken.

De gegevens van de resultaten van CCA worden door CANOCO ordelijk gerangschikt in een aantal file bewaard. Van deze files kan met het programma CANOPLOT volautomatisch een reeks grafieken gemaakt worden. Deze grafieken zijn dan zogenaamde character graphics. Zeer geschikt om snel een ordinatie te kunnen beoordelen, maar niet erg geschikt voor in verslagen e.d.

De uitvoerfiles zijn m.b.v. de benodigde apparatuur en programma-tuur daarvoor (modem, Kermit) uit de VAX-computer gehaald, naar een P.C. toe.

De files zijn vervolgens binnengehaald in een P.C.-programma, dat getalsmatige grafisch kan weergeven. Meestal gaat het dan om zogenaamde Spreadsheets. Voorbeelden van spreadsheets zijn Lotus 1-2-3, SuperCalc, Visicalc, Javelin, Multiplan etc.

De grafieken zoals ze in het verslag zijn opgenomen zijn gemaakt m.b.v. Lotus 1-2-3.

Deze werkwijze, met tussenkomst van een P.C., levert grafieken van een beduidend hogere kwaliteit dan de character graphics van CANOPLOT.