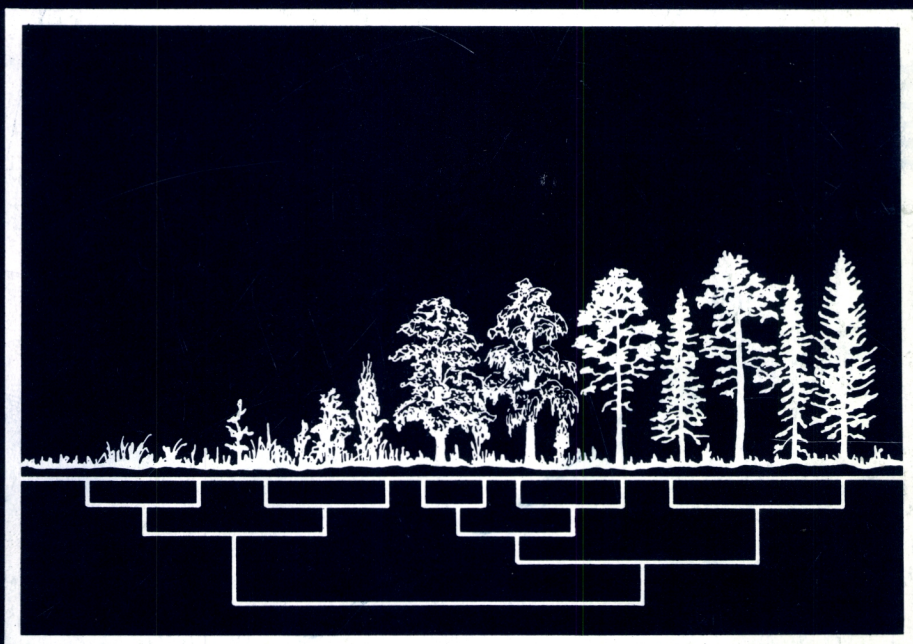




Metsänhoidon tutkimusosasto



Kari Mikkola  
Eeva-Liisa Jukola-Sulonen

YHTEISÖEKOLOGISTEN AINEISTOJEN  
KÄSITTELY JA ANALYSOINTI  
VAX - TIETOKONEELLA

HELSINKI 1984

# METSÄNTUTKIMUSLAITOS

## Jalostusosasto

*Kansikuva: Kangasmetsän sukessiogradientti avohakkuun jälkeen.  
Kuvan alaosassa on dendrogrammi, joka kuvaa kasvillisuuden ryhmittelyä TWINSPAN-ohjelmalla.*

*Piirros: Sara Niini*

Veikko Kumpulainen

Eeva-Liisa

Kari Mikkola & Eeva-Liisa Jukola-Sulonen

YHTEISÖKOLOGISTEN AINEISTOJEN KÄSITTELY JA  
ANALYSOINTI VAX-TIETOKONEELLA

SAATTEEKSI

Tässä oppaassa esitellään suppeassa muodossa Metsäntutkimuslaitoksen VAX -tietokoneissa käytettävissä olevat yhteisöekologiset monimuuttujamenetelmät. Tekstin pohjana on ollut Helsingin yliopiston kasvitieteen laitoksen monisteita -sarjassa ilmestyvän vastaavan laajemman oppaan luku 2 (Mikkola ym. 1984).

Menetelmien matemaattiseen perustaan ja tulosten tulkintaan ei syvennyttä; asiasta kiinnostuneet voivat etsiä tietoja alan laajasta englanninkielisestä kirjallisuudesta.

Ohjelmien alkuperä:

Cornell Ecology Programs (CEP) - Ecology and Systematics, Cornell University, Ithaca, New York. Alunperin laaja yhteisöekologisten ohjelmien kokonaisuus, josta tässä oppaassa esitellään METLassa hieman modifioituneet versiot CONDENSE, DATAEDIT, DECORANA ja TWINSpan.

CLUSTAN - Iäkäs versio Edinburghin Yliopistossa alunperin kehitellystä klusterointianalyysistä. Jouko Sarvala on lisännyt ohjelmaan eräitä etäisyysindeksejä ja vuorovaikutteisen version on muokannut K. Mikkola.

TRANS ja CONVERT on laadittu Helsingin Yliopiston Kasvitieteen laitoksella (Mikkola ym. 1984) ja ohjelma DIVER on Jaakko Heinosen käsialaa.

CEP - ohjelmista on olemassa laajat, englanninkieliset käsikirjat sekä matemaattisella että metsänhoidon osastolla.

Yhteisöekologiset monimuuttujamenetelmät soveltuvat periaatteessa hyvin monenlaisten havainnot-muuttujat -matriisien käsittelyyn. Metsäntutkimuksen piirissä kerätään vuosittain valtavia määriä monipuolista tietoa erilaisista ekosysteemeistä. Perinteisin menetelmin (korrelaatiot, regressiot, varianssianalyysit ym.) voi olla vaikeaa selvittää aineiston sisäisiä erittäin monimutkaisia syy-seuraussuhteita. Monimuuttujamenetelmät tavallaan pelkistävät aineiston sisältämän informaation monikymmenulotteisesta helpommin tajuttavaan 2-3-ulotteisuuteen.

Yhteisöekologisten tutkimusten peruskohteena on useimmiten eliöiden läsnäolo/poissaolo tai runsaus tietyssä rajatussa otoksessa. Havaintoarvojen jakaumat ovat usein hyvin vinoja ja laji- ja biotooppikohtaisesti vaihtelevia. Useat tilastolliset menetelmät ovat vahvasti sidottuja oletuksiin jakaumien luonteesta ja riippuvuuksien lineaarisuudesta. Esim. korrelaatioanalyysin käyttö lajien ja ympäristömuuttujien suhteiden selvittämiseen ei ainakaan heterogeenisissä aineistoissa ole perusteltua. Tässä esitellyt menetelmät eivät ole varsinaisia tilastollisia metodeja - hypoteeseja ei testata vaan pareminkin luodaan. Moniulotteisen aineiston päävaihtelusuuntien ja ryhmittymisen tarkastelu selkeyttää tutkijan näkemystä aineistoonsa ja voi antaa perusteita myös aineiston osittamisiin ym. esitoimenpiteisiin ennen varsinaisten tilastollisten menetelmien käyttöä. Usein ryhmittely- ja ordinaatioanalyysien tulokset ovat jo sinänsä riittäviä lopputuloksia tutkimusongelman ratkaisun kannalta.

Parannusehdotukset oppaan sisältöön ovat tervetulleita. Ohjelmien mahdollisista puutteista tai virheistä pyydetään välittömästi ilmoittamaan korjaustoimenpiteitä varten.

Professori Erkki Lähde on lukenut käsikirjoituksen ja tehnyt eräitä korjausehdotuksia. Tutkija Carl-Gustaf Snellman on toiminut ATK-teknisenä asiantuntijana ja neuvonantajana sekä tarkastanut oppaan ohjeiden ja ohjelmien toimivuuden. Parhaat kiitoksemme tarkastajille sekä muille työn valmisteluissa mukana olleille henkilöille.

Helsingissä 28.12. 1984

Tekijät

## SISÄLLYSLUETTELO

1. TIEDOSTON MUODOSTAMINEN, RAKENNE JA MUOKKAUS ....	1-1
1.1 Perusmuotoinen havaintomatriisi .....	1-1
1.1.1 Format-lause .....	1-2
1.1.2 TRANS - muunnos- ja transponointiohjelma ....	1-5
1.1.2 CEP-tiedosto .....	1-6
1.2.1 CONDENSE - CEP-tiedoston muodostus .....	1-8
1.2.2 DATAEDIT - CEP-tiedoston muokkaus .....	1-9
1.2.2 CONVERT - CEP-tiedoston muunto perusmuotoon	1-14
2. ORDINAATIO- JA LUOKITTELUOHJELMAT .....	2-1
2.1 DECORANA .....	2-1
2.2 TWINSPAN -ryhmittely .....	2-3
2.3 CLUSTAN -ryhmittely .....	2-8
3. DIVERSITEETTI-INDEKSIT .....	3-1

## LÄHTEET

MUUTA AIHEESEEN LIITTYVÄÄ KIRJALLISUUTTA



## LUKU 1

### TIEDOSTON MUODOSTAMINEN, RAKENNE JA MUOKKAUS

#### 1.1 PERUSMUOTOINEN HAVAITOMATRIISI

Perusmuotoinen havaintomatriisi vastaa oikeastaan perinteistä kasvillisuustaulukkoa muodoltaan. Lajit ovat riveinä ja alat sarakkeina. Rakenne on selkeä ja tallennusvirheet kohtuullisella vaivalla paikannettavissa. Useat menetelmät (BMDP, CLUSTAN -ryhmittely ym.) edellyttävät syöttötiedostolta tällaista rakennetta. Haittana on matriisin "harvuus"; 0-peittävyudet vallitsevat. Tämä vaikeuttaa tallennustyötä jonkin verran ja myös tietokoneen muistitilan käyttö on tällöin epätaloudellista. Esim. 200 lajia ja 300 alaa sisältävä matriisi vie tietokoneen keskusmuistista tilaa 60000 muistialkion verran.

Kasvillisuusaineistojenkin keruussa tulisi pyrkiä välivaiheita säästävään suoraan tallennuslomakkeille tapahtuvaan kenttätietojen keräämiseen. Mitään yhtenäistä tiedostorakennestandardia ei toistaiseksi ole kuitenkaan saatu aikaan. Joissain tutkimuksissa on sovellettu vektorimuotoisia tietueita tai tiedostoja lajien numerokodeineen, mutta aineistojen käsitteilyyn ja muuntamiseen esim. BMDP -analyyseihin soveltuviksi tarvitaan räätälintyönä tehtyjä ohjelmia.

Ongelmana on lisäksi lajikoodiston ajan tasalla pito; ainakin sammu- ja jäkälälajien taksonomiassa on edelleen paljon selvittämistä ja muutoksia tapahtuu suhteellisen usein. Koodikirjan selaaminen on sekin lisätöyväaihe.

Nykyinen menetelmä erillisine kenttä- ja tallennuslomakkeineen kaivannee kuitenkin parannuksia. Uudet ideat olisivat kasvillisuustutkimustenkin kehittämiseksi tervetulleita.

Seuraavassa annetaan suositusluontoinen ohje perinteisen näytealat sarakkeina - lajit riveinä -havaintomatriisiin tallennusmuodosta.

Jos runsaudet tai frekvenssit tallennetaan prosentteina yhdellä desimaalilla, riittää havaintokentän leveydeksi kolme saraketta. Tällöin suurin mahdollinen runsausarvo on 99,9.

Jos runsaudet on arvioitu luokka-asteikolla (esim. 1-9 -asteikko), riittää leveydeksi yksi sarake.

Tietueen pituus on käytännön syistä syytä rajoittaa 80 sarakkeeseen. Jos tämä ei riitä, voidaan muodostaa jatkorivejä eli jatketaan vain tallennuslomakkeen täyttöä seuraaville riveille. Vaikka kaikilla lajeilla ei olisikaan havaintoja jatkorivien näytealoilla, on niillekin silti muodostettava tyhjät jatkorivit. Ohjelmat lukevat jokaista tietuetta mekaanisesti samalla tavalla annetun formaatin mukaisesti.

Tässä yhteydessä suositellaan käytettäväksi seuraavaa lajitietueen muotoa:

4 ensimmäistä saraketta: lajin juokseva numero  
 8 seuraavaa: lajinimi 2 x 4 merkin lyhenteenä  
 n x a seuraavaa: peittävyys, frekvenssi, biomassa  
 tms.

jossa n = näytealojen lukumäärä, a = havaintokentän leveys.

Lajinimi lyhennetään seuraavasti: 4 kirjainta suku- ja lajinimestä yhteen kirjoitettuna. Esim: VACCVITI, PLEUSCHR, RUMEACLA, RUMEACET. Nimi on tällöin yleensä täysin tunnistettava ja käy eräiden ohjelmien luettavaksi suoraan ilman muutoksia.

Koealojen tunnukset voidaan kirjoittaa tiedoston loppuun formaatilla (10A8) eli kymmenen kahdeksan merkin kenttää / rivi. Viimeinen rivi saa jäädä vajaaksi, mutta tunnusten järjestyksen ja lukumäärän on ehdottomasti oltava sama kuin edeltävässä numeerisessa aineistossa. Tunnuksen sisältö on vapaa, esim. LEHT122a tai 26 OMT.

Lajinimi- ja näytealatunnusten sijainnin ja formaatin näinkin tarkka määrittäminen on tarpeen, jos tiedosto muutetaan tiivistettyyn CEP-muotoon CONDENSE -ohjelmalla. Lajinimi voi tosin sijaita missä tahansa lajirivillä (ks. s. 1-7).

Taulukossa 1 on malli tällaisella rakenteella tallennetusta tiedostosta.

#### 1.1.1 FORMAT - LAUSE

FORMAT -lause eli formaatti on FORTRAN -ohjelmointikielessä luettavan ja tulostettavan tiedon muotoa määrittelevä lause. Koska FORTRAN tuntee useita eri tietomuotoja (mm. kokonaisluvut, reaailuvut, looginen ja merkkietieto), joudutaan kutakin tietomuotoa käsittelemään erikseen määrätyllä tavalla. Yleisimmin esiintyvä virhe syöttö- ja tulostustoiminnoissa onkin se, että tieto ja sen käsittelytapa eivät tyypeiltään vastaa toisiaan.



37 SKILPISJÄRVI KASVIDATA 19.3.1984 TC\*\* S

(14,9(I3,FS.1))

1	1	16.0	3	7.0	5	11.0	11	0.1	13	0.1	14	0.1	16	0.1	17	0.1	19	18.0
1	22	0.3	24	1.3	25	1.3	28	2.5	29	0.1	31	0.3	32	0.1	35	0.1	36	8.0
1	37	0.1																
2	1	1.0	2	0.1	3	4.8	4	0.9	5	0.5	6	1.8	7	0.3	8	0.1	10	0.4
2	11	0.6	12	0.8	13	0.1	14	0.5	15	0.1	16	0.3	17	0.1	18	0.1	19	1.3
2	20	34.0	21	0.3	23	1.0	24	2.5	25	0.1	26	0.5	28	0.1	29	0.1	31	0.1
2	35	0.5	36	2.4														
3	1	2.3	2	0.1	3	1.8	5	3.5	6	5.0	7	0.1	8	0.1	9	0.3	10	6.0
3	11	0.3	12	0.1	13	0.1	14	0.1	16	0.1	17	0.1	18	0.6	19	0.9	20	4.3
3	21	0.1	23	0.1	24	0.8	25	1.9	27	4.3	30	0.1	33	0.1	34	0.1	36	3.3
3	37	0.4																
4	1	4.5	2	0.1	3	13.0	5	0.4	6	9.3	11	0.1	19	72.0	22	3.8	23	2.5
4	24	1.8	27	0.1	28	0.1	30	0.4	31	0.5	32	0.4	33	0.1	34	0.1	35	0.1
4	36	2.4	37	0.1														
5	1	1.7	2	0.3	3	0.7	4	0.1	5	0.5	6	4.3	7	0.3	8	1.4	9	0.1
5	10	2.3	14	0.6	15	0.1	16	0.5	17	0.1	18	0.1	19	0.8	20	41.0	21	3.3
5	25	0.3	26	0.1	27	0.3	35	0.1	36	2.0	37	0.5						

0  
 VACMYRTVACVITTEMPEHERMVACULIGFESTOVINDESCFLEXCARAVAGIANTHODORBARTALPIGERASYLV  
 LINNBRELYCOANNOMELAPRATEDILAPPYROMINOSOLIVIRGRIEUEUROVIOLBIFLEUSCHRHYLOSPL  
 POLYCOMMPOLYJUNIDICRSCOPDICRUSBRACSALEBRACREFLPOHLNUTACETRISLACLADCORNLADCRAC  
 CLADRANGLADMITICLADPYXICLADPLEUPELTAPHTBARBCOLLORHTATTE  
 ICOEM 2SEMT 3LEHT 4SEMT 5LEHT

Taulukko 2. CEP-tiedosto.

1VACMYRT 160 10 23 45 17

2VACCVITI	01	01	01	03	
3EMPEHERM	70	48	18	130	07
4VACCVILIG	09	09	01		
5FESTOVIN	110	05	35	04	05
6DESCFLEX	18	50	93	43	
7CAREVAGI	03	01	03		
8ANTHODOR	01	01	14		
9BARTALPI	03	01			
10GERASYLV	04	60	03	23	
11LINNBORE	01	06	03	01	
12LYCOANNO	08	01			
13MELAPRAT	01	01	01		
14PEDILAPP	01	05	01	06	
15PYROMINO	01	01			
16SOLIVIRG	01	03	01	05	
17TRIEURO	01	01	01	01	
18VIOLBIFL	01	06	01		
19PLEUSCHR	180	13	09	720	08
20HYLOSPL	340	43	410		
21POLYCOMM	03	01	33		
22POLYJUNI	03		38		
23DICRSCOP	10	01	25		
24DICRFUSC	13	25	08	18	
25BRACSALE	13	01	19	03	
26BRACREFL	05			01	
27POHLNUTA	43	01	03		
28CETRISLA	25	01	01		
29CLADCORN	01	01			
30CLADCRAC			01	04	
31CLADRANG	03	01	05		
32CLADMITI	01		04		
33CLADPYXI			01	01	
34CLADPLEU			01	01	
35PELTAPHT	01	05	01	01	
36BARBCOLL	80	24	33	24	20
37ORTHTATTE	01	04	01	05	
ICOEM	2SEMT	3LEHT	4SEMT	5LEHT	

Taulukko 1. Perusmuotoinen tiedosto.

Kasviekologisten ohjelmien käyttäjän on tunnettava havaintoaineistonsa täsmällinen formaatti.

CEP -tiedostossa formaatti on sisäisenä mukana tiedoston toisena tietueena eikä sitä tarvitse antaa ulkopuolelta.

Formaatin yleinen muoto on seuraava:

$$(q_1, t_1, z_1, \dots, t_{n-1}, z_{n-1}, t_n, q_2)$$

jossa:  $q_1$  ja  $q_2$  voivat olla joko tyhjiä tai yksi tai useampi / - merkki,  $t_1$  on kentän kuvaus tai ryhmä niitä,  $z_i$  on kentän erotin, joko pilkku tai yksi tai useampi / - merkki.

Formaatti annetaan aina suluissa.

Samassa formaatissa voidaan määritellä useita tietueita (kuten lajin kaikki tietueet). Tietueiden kuvaukset erotetaan kauttaviivalla.

Aritmeettinen kenttä voidaan yleisimmin käytetyissä lukuformaateissa määritellä seuraavasti:

rIw tai rFw.d

Joissa r = toistotekijä (kokonaislukuvakio), joka ilmoittaa kuinka monta kertaa peräkkäin kenttä toistuu tietueessa.

w = kentän leveys merkkeinä, kokonaislukuvakio.

d = desimaaliosan leveys, kokonaislukuvakio.

On olemassa useita muitakin kenttämäärittelytyyppejä, mutta niihin ei tässä puututa. Kokonaislukujen käsittelyyn käytetään ainoastaan I-koodia.

ESIMERKKI:

Tietue on seuraavan muotoinen:

.500,250,300

Tällöin lukuformaatti olisi: (3I4). Kentän alkuun jätetyt tyhjät tulkitaan nolliksi.

Kentät voivat olla eri kokoisia, esim. (I2,I4,I6) lukee ensin kahden merkin levyisen kentän, sitten neljän ja lopuksi kuuden.

Reaalilukujen (käytännössä desimaalilukujen) käsittelyyn voidaan käyttää F-koodia. Useat ohjelmat edellyttävät lukuformaatilta tätä muotoa.

## ESIMERKKI:

12345678

Kun ylläoleva tietue luetaan formaatilla (2F4.2) saadaan arvot 12.34 ja 56.78. Jos formaatti olisi (2F4.0) saataisiin vastaavasti 1234.0 ja 5678.0.

X-koodilla voidaan hypätä haluttujen sarakkeiden yli. Jos esim. tietueessa lajin:o ja -nimi ovat 11:11ä ensimmäisellä sarakkeella ja seuraavana 20 näytealan peittävydet kolmen sarakkeen levyisinä kenttinä, olisi lukuformaatti seuraavanlainen: (11X,20I3) tai (11X,20F3.1).

Formaatissa on usein tarpeen, että sama kenttämäärittelyryhmä esiintyy useamman kerran peräkkäin. Ryhmää ei tarvitse kirjoittaa useampaan kertaan, vaan se voidaan ympäröidä suluilla ja sen eteen sijoittaa toistotekijä, esim: (I4,8(3X,I3,F5.2)).

Jatkorivit erotetaan / -merkillä. Sitä käytettäessä ei tarvita pilkkua erottamaan kenttämäärittelyjä toisistaan.

ESIMERKKI: Lajia kohti on kaksi riviä:

126CALAEP	IG	5	10	1	200	15	1	1	5	2
2		10		1	250		360	1	1	6

Tällöin tiedoston lukuformaatti on: (11X,14F3.1/17F3.1). Kauttaviivan voi tarvittaessa sisällyttää myös suluilla erotettavaan toistettavaan osaan.

Merkitietoa käsitellään A -koodilla. Lajinimiä lukeva formaatti voi olla esim. (10A8).

Lisätietoja FORTRAN -ohjelmointikielen ominaisuuksista voi etsiä mm. Toivosen ja Yrjölän (1978) FORTRAN-77 -ohjelmointioppaasta.

## 1.1.2 TRANS - MUUNNOS- JA TRANSPONOINTIOHJELMA

Havaintomatriisin transponoinnilla tarkoitetaan sen "kääntämistä" 90 astetta. Uudessa matriisissa sarakkeet ja rivit ovat vaihtaneet paikkaa. Tavallisesti perusmuotoisissa kasvillisuusaineistoissa lajit ovat riveinä ja alat sarakkeina, transponoidussa matriisissa tilanne on siis päinvastainen.

Kasvillisuusaineistoissa lajit ovat havaintoja ja näytealat muuttujia. Usein tilastomatematiikassa käsitetään rivit havainnoiksi (cases, observations) ja sarakkeet muuttujiksi (variables, attributes). Kaavalausekkeissa edellisten indeksinä on i ja jälkimmäisten j. Tämä terminologia on hyvä muistaa sekaannusten välttämiseksi. CEP -tiedostot ovat tavallaan transponoituja; näytealat ovat riveinä (i).

Käytännössä matriisin transponointia voi käyttää CLUSTAN-ohjelman yhteydessä (klusterointi lajeille), sekä suoritettaessa BMDP -faktorianalyysi lajien väliseen korrelaatiomatriisiin perustuen.

TRANS -ohjelmalla voidaan suorittaa matriisille transpoosi. Sillä on mahdollista myös tehdä muunnos  $\ln(x+1)$  sekä standardointi sarakkeittain. Logaritmuunnos vaimentaa dominanttien vaikutusta ja tekee lajit "tasa-arvoisiksi". Standardointi suoritetaan kaavalla:

$$X_{mi} = \frac{100}{\sum X_i} * X_i$$

jossa  $X_i$  = lajin i alkuperäinen runsausarvo

$X_{mi}$  = "-" muunnettu "-"

Jokainen sarake (näyteala) skaalataan siis siten, että runsauksien summaksi tulee sata. Tällöin näytealojen runsauksien absoluuttisten arvojen vaikutus vaimenee, vain suhteelliset erot jäävät voimaan.

TRANS käynnistetään seuraavasti:

\$ RUN CEP:TRANS

Aluksi ohjelma kysyy syöttö- ja tulostustiedostojen nimiä. Kysely jatkuu:

ANNA RIVIMUUTTUJIEN LUKUMÄÄRÄ

Jos käsiteltävä tiedosto on perusmuotoinen havaintomatriisi (ks. s. 1-1), annetaan lajien lukumäärä.

## ANNA SARAKEMUUTTUJEN LUKUMÄÄRÄ

Sarakemuuttujien lukumäärä vastaa perusmuotoisessa matriisissa näytealojen lukumäärää.

### HALUATKO MUUNNOKSEN?

LN(X+1) (=1)

STANDARDOINTI ALOILLE (=2)

EI MUUNNOSTA (=3)

### HALUATKO TRANSPOOSIN?

KYLLÄ = 1, EI = 2

Näihin kysymyksiin vastataan halutun toimenpiteen mukaisesti. Transpoosi jätetään tietysti suorittamatta jos aineistolle halutaan tehdä pelkkä muunnos.

## ANNA KIRJOITUSFORMAATTI

Annetaan sopiva tulostusformaatti. Jos suoritetaan standardointi, on tulostuville uusille suhteellisille runsausarvoille varattava riittävän tilava kenttä (esim. F5.1).

Alkuperäisen tiedoston tyhjät havaintokentät kirjoittuvat tulostiedostoon nolliksi.

Ohjelmaa voi käyttää myös minkä tahansa taulukkomuotoisen tiedoston formaatin muuttamiseen; muunnos ja transpoosi jätetään vain pois ja määritellään kirjoitusformaatti haluttuun muotoon. Tällä tavoin voi myös taulukoida havaintotiedostonsa selkeinä kenttinä.

Huom! F-muotoisen tulostusformaatin kentät on syytä määrittellä aina yhtä saraketta lukuformaattia leveämmiksi, koska tulostuva desimaalipiste vie tilaa yhden sarakkeen verran. Liian ahdas tulostusformaatti saa aikaan vain \* -merkkien kirjoittumisen. Tämä koskee myös muita ohjelmia, joissa tulostusformaatti annetaan ulkopuolelta.

## 1.2 CEP - TIEDOSTO

CEP on lyhennys nimestä "Cornell Ecology Programs". Tämä Cornellin yliopistossa USA:ssa kehitetty yhteisöekologisten aineistojen analysointiin tarkoitettu ohjelmistopakkaus käyttää uudempien ohjelmien syöttöaineistona tiivistettyä havaintomatriisia; kultakin näytealalta luetaan tietueeseen vain nollaa suuremmat havainnot. Käytännössä tällainen tietue on seuraavanmuotoinen:

Alan n:o, lajin:o, runsausarvo, lajin:o, runsausarvo jne.

Pienissä matriiseissa tilansäästö on kyseenalaista, mutta suu-

rissa etu on selvä. Tällainen ns. vektorimuotoinen tiedosto nopeuttaa myös laskentaa huomattavasti.

Koko tiedoston muoto on seuraavanlainen (taulukko 2):

1. Otsikkorivi:

60	300TSIKKO	TC	** S
----	-----------	----	------

- Sar. 1-5: Lajien lukumäärä
- Sar. 6-10: Alojen lukumäärä
- Sar. 11-70: Vapaamuotoinen otsikko, max. 60 merkkiä
- Sar. 71: T
- Sar. 72: C
- Sar. 77: \* (= Lajinimet ovat mukana)
- Sar. 78: \* (= Näytealojen tunnuksset ovat mukana)
- Sar. 80: S

2. Formaattirivi.

(I5,8(I3,F5.2))

8

- Sar. 1-60: Formaatti. Ensin viiden merkin levyinen kenttä näytealan numerolle, sitten 8 kolmen merkin levyisen kokonaisluvun ja viiden merkin levyisen reaaliluvun muodostamaa paria. I3 -kenttään tulee lajinumero ja F5.2 -kenttään runsausarvo. Tässä käytetty esimerkkiformaatti ei ole ainoa mahdollisuus. Kenttien leveydet ja toistomäärät ovat vapaasti määriteltävissä.
- Sar. 61-70: Lajinumero - runsaus -parien maksimimäärä tietuetta kohden.

3. Varsinainen tiedosto näytealanumero-lajinumero-peittävyys-tietoineen. Jos lajinumero-peittävyys -pareja on enemmän kuin 80 merkin riville mahtuu, muodostetaan jatkorivejä. Näytealanumero on kirjoitettava näidenkin rivien alkuun.

4. Nollatietue. Rivin alkuun kirjoitetaan nolla; tämä ilmoittaa ohjelmalle numeerisen tiedon loppumisen ja nimitiedon alkamisen.

5. Lajinimet formaatilla (10A8) = kymmenen kahdeksan merkin kenttää / tietue. Esim. VACCVITIPLEUSCHRCALLLVULGDICRSCOP jne. Viimeinen rivi saa jäädä vajaaksi.

6. Alojen tunnuksset samalla formaatilla.

Esim. 0126 OMT0127 VT0128 MT jne.

Huom! Näytealanumeroina on käytettävä juoksevaa numerointia 1-n nousevassa järjestyksessä. Lajinumeroiden on myös oltava nousevassa järjestyksessä kullakin näytealalla.

### 1.2.1 CONDENSE - CEP-TIEDOSTON MUODOSTUS

CONDENSE muodostaa perusmuotoisesta matriisista CEP-tiedoston. Alkuperäinen versio on kehitetty Cornellin Yliopistossa USA:ssa (CEP-ohjelmat). Laitoksen käyttöön on alkuperäisestä reikäkorttikäyttöön soveltuvasta versiosta muokattu vuorovai-  
kutteisesti toimiva ohjelma.

Ohjelman esittämät kysymykset ovat selkeitä ja itsensä selit-  
täviä, joten niiden tarkempi läpikäyminen lienee tarpeetonta. Aineiston lajien ja alojen lukumäärä sekä formaatti on kui-  
tenkin selvitettävä ennen ajoa.

Jos lajinimi on mukana perusmuotoisen matriisin lajitietu-  
eessa, voidaan se lukea mukaan tulostuvaan CEP-tiedostoon. Lajinimen on oltava 2 x 4 merkin lyhenteenä (esim. VACCMYRT).

Jos näytealojen vastaavat tunnuksot ovat mukana lukutiedoston  
lopussa saadaan nekin mukaan (ks. s. 1-2).

Vaikka paikkojen tai lajien nimiä ei lähtötiedostossa olisi-  
kaan, onnistuu muunnos tästä huolimatta. Kun ohjelman esittä-  
miin kysymyksiin nimien läsnäolosta ja sijainnista vastaa nol-  
lalla, kirjoittuu muodostuvan CEP-tiedoston loppuun keino-  
koiset nimet "SPC 1 ... SPC n" ja "SMP 1 ... SMP n". Ilman  
minkäänlaisia nimitietueita CEP-tiedosto on käyttökelpoton,  
eikä kelpaa millekään ohjelmalle syöttötiedostoksi.

Harvinaisimmat lajit on myös mahdollistaa pudottaa pois muo-  
dostettavasta tiedostosta antamalla luku N, jota harvemmin  
esiintyvät lajit jäävät pois.

Muodostuvan CEP-tiedoston formaatti on ohjelmansisäisenä ole-  
tusarvona asetettu kiinteäksi. Ohjelmoinnin alkeita ymmärtävä  
käyttäjä voi helposti muuttaa asetusta FORTRAN -kielistä  
koodia muuttamalla. Oletusformaatti on: (I5,8(I3,F6.2)).

CONDENSE käynnistyy komennolla:

```
§ RUN CEP:CONDENSE
```

Tämän jälkeen ilmestyvät päätteelle kysymykset syöttö- ja tu-  
lostustiedostoista, lukuformaateista ym. Lajien ja näytea-  
lojen lukumäärät on selvitettävä tarkasti ennen ajoa.

Muodostettu CEP-tiedosto on sellaisenaan valmis käsiteltäväksi  
ohjelmilla DATAEDIT, TWINSPAN ja DECORANA.

CONDENSEN kapasiteetti on 1000 alaa ja 500 lajia.

### 1.2.2 DATAEDIT - CEP-TIEDOSTON MUOKKAUS

Kerätty numeerinen aineisto voi joskus olla sellaisenaan sopimaton monimuuttujamenetelmien syöttöaineistoksi. Tällöin aineiston tietty muuntelu voi tulla kysymykseen.

On tärkeää muistaa, että kaikkien monimuuttujamenetelmien tarkoitus on tehdä kerätty numeerinen tieto ymmärrettäväksi, tiivistää informaatiota, kun taas editointien tulisi korostaa aineiston eri piirteitä. Järkevästi tehty muunnos voi edesauttaa tulosten tulkintaa enemmän kuin koko analyysimenetelmän vaihtaminen.

On tietysti mahdollista, että koko aineistossa ei kertakaikkiaan ole mitään mielekästä tulkittavaa. Tällöin on olemassa vaara, että aineistoa sopivasti ja monipuolisesti manipuloidulla saadaan tällöinkin lopulta "oikeat", tutkijan omaa näkemystä mukailevat tulokset. Objektiivisuuteen pyrkimisen pitäisi kuitenkin olla koko numeeristen metodien käytön taustasy, joten liialliseen syöttöaineiston muunteluun ei missään tapauksessa pidä ryhtyä.

DATAEDIT on CEP-ohjelmiin kuuluva monipuolinen editointiohjelma. Paitsi erilaisia muunnoksia DATAEDITillä voi yhdistellä haluttuja näytealoja "keskiarvonäytealoiksi", pudottaa pois lajeja tai näytealoja, viipaloita aineisto osiin annettujen parametrien mukaan, tuottaa erilaisia yhteenvetoja ja taulukoita ym. Ohjelmalla voi myös muodostaa joko TWINSPAN-analyysin tuottamasta luokittelusta haluttujen jakotasojen ryhmistä muodostettuja tai subjektiivisen luokittelun mukaisia keskiarvonäytealoja.

Ohjelma ei ole vuorovaikutteinen. Suorituskäskyt kirjoitetaan käsiteltävän tiedoston alkuun ja loppuun. Vain CEP-tiedostot kelpaavat syöttöaineistoksi.

Tässä yhteydessä ei ole mahdollista antaa täydellistä käyttöohjetta. Joitain käyttökelpoisia käskyjä esitellään lyhyesti. Mm. METLAN matemaattisella osastolla on kopio alkuperäisestä englanninkielisestä käsikirjasta.

Aluksi käsiteltävän tiedoston alkuun kirjoitetaan seuraava käsky:

```
READ MATRIX
```

Nyt tiedosto on DATAEDITin luettavissa. Varsinaiset toimintakäskyt kirjoitetaan tiedoston loppuun, heti koealatunnuksia seuraaville riveille. Vain isot kirjaimet ovat käytössä.

DATAEDIT käynnistyy lähettämällä seuraava käsky:

```
$ RUN CEP:DATAEDIT
```

Tämän jälkeen ohjelma kysyy syöttötiedoston nimeä. Tulostus-



tiedostoja on kaksi:

- (1) EDIT.LIS : - Ohjelman sisäiset virheilmoitukset
  - Tiedot eri vaiheiden suorituksesta, poistetut lajit ja näytealat, yhdistettyjen näytealojen rakenne ym.
  - Aineiston tilastotiedot, pyydetyt taulukot ym.
- (2) EDIT.OUT : - Uusi CEP-tiedosto.

Seuraavassa malleja DATAEDITin käskyistä:

Esimerkki 1.

- Pudotetaan yhden ja kahden esiintymän lajit pois.  
Kirjoitetaan uusi tiedosto.

OUTPUT FLAG MATRIX PUNCH KEEPFLAG (OCCUR 3)

Esimerkki 2.

- Poistetaan aloja ja lajeja.  
Kirjoitetaan uusi tiedosto.  
Jatkorivi aloitetaan & -merkillä.

EXCLUDE SPECIES 3,4,5 SAMPLES 13,19-30,  
&39,42,45

OUTPUT FLAG MATRIX PUNCH

Esimerkki 3.

- Kirjoitetaan uusi tiedosto aloista 16-25.

EXCLUDE SAMPLES ALL  
INCLUDE SAMPLES 16-25  
OUTPUT FLAG MATRIX PUNCH

Esimerkki 4.

- Suoritetaan runsausarvojen muunnoksia.

TRANSFORMATION XXXX  
OUTPUT FLAG MATRIX PUNCH

- XXXX voi olla
  - a) OCTAVE
    - Oktaaviskaala = (lähes) logaritminen 1-9 -asteikko.
  - b) BRAUN-BLANQUET
    - Keskieurooppalainen 1-5 -asteikko.
  - c) MULTIPLY X
    - X on reaalityyppi, jolla kaikki matriisin alkiot kerrotaan.
  - d) PRESENCE/ABSENCE
    - Vain lajin läsnäolo mukaan.  
( 0 tai 1 arvoiksi)

- e) SQUAREROOT  
- Neliöjuurimuunnos.
- f) LOGARITHM  
-  $\log(X+1)$
- g) STANDARDIZE  
- lajien arvot skaalataan 0 - 100  
-vaihteluvälille kaavalla:

$$X_{mj} = \frac{100}{\max(X_j)} * X_j$$

- h) RELATIVIZE  
- runsauksien summaksi 100  
kullakin näytealalla:

$$X_{mj} = \frac{100}{\sum X_j} * X_j$$

joissa:  $X_{mj}$  = muunnettu runsaus  
lajille j,  
 $X_j$  = alkuperäinen runsaus  
lajille j,  
 $\max(X_j)$  = lajin j korkein  
runsausarvo.

#### Esimerkki 5.

- Aineiston taulukointi. Tulokset levylle tiedostoon  
"EDIT.LIS".

OUTPUT FLAG MATRIX SORT (SPECIES BY XX, SAMPLES BY XX) PRINT  
OUTPUT FLAG MATRIX PROFILE

XX voi olla

- a) RA  
- Reciprocal averaging -ordinaation  
mukainen järjestys.
- b) RATABLE  
- Samoin, myös ordinaatiopisteet tuloste-  
taan.
- c) ABUNDANCE  
- Runsauksien mukainen järjestys.
- d) PRESENCE  
- Läsnäolon mukainen järjestys.

- Käskyparista ylempi tuottaa halutulla tavalla järjes-  
tetyt lajit-paikat -kasvillisuustaulukon. Jos tähän lisä-  
tään (SPECIES MAXIMUM) saadaan taulukko, jossa lajit on  
skaalattu + - 9. Muutoin koko matriisi skaalataan samalle  
asteikolle.

- "PROFILE" -käsky tuottaa aineiston lajien ja alojen tau-  
lukoinnin järjestettynä keskimääräisen runsauden sekä  
frekvenssien mukaan.

OUTPUT FLAG MATRIX PUNCH -käsky tuottaa aina uuden CEP-tie-

doston ("EDIT.OUT"), joka on sellaisenaan valmis käsiteltäväksi. Jos edeltävä käsittelykäsky aiheuttaa muutoksia kokonaislajimäärässä, ohjelma ottaa automaattisesti muutokset huomioon. Samanaikaisesti tulostuvassa EDIT.LIS -tiedostossa on luettelo annetuista käskyistä ja mahdolliset virheilmoitukset. Frekvenssien ja runsauksien mukaan järjestetyt laji- ja näytealaluettelot sekä kasvillisuustaulukot ym. kirjoittuvat myös tähän tiedostoon.

Jos TWINSPAN -analyysillä tuotetusta ryhmittelystä halutaan muodostaa ryhmäkohtaisia yhdistelmänäytealoja tai pilkkoa alkuperäinen aineisto halutun jakotason ryhmittelyn mukaan osiksi, on TWINSPANia ajettaessa vastattava kysymykseen "TYPE 1 IF MACHINE-READABLE..." ykkösellä (ks. s. 2-5). Tällöin levyille muodostetaan tiedosto FOR007.DAT johon analyysin tulos kirjoittuu erityiseen vektorimuotoon. Tälle tiedostolle suoritetaan seuraavat toimenpiteet:

- (1) Poistetaan alusta seuraavat 4 riviä:

```

--- Otsikko ---
    Tyhjä rivi
SAMPLE CLASSIFICATION
THERE ARE n ITEMS

```

- (2) Poistetaan näytealavektoria seuraava lajivektori otsikkoriveineen.

Lopputuloksena on pelkkä näytealaryhmittelyvektori.

Näytealavektori liitetään käsiteltävän CEP-tiedoston loppuun seuraavasti:

```

READ MATRIX
....
.... <CEP-tiedosto>
....
READ VECTOR SAMPLES HIERARCHY
    nnTWINSPAN STANDARD ANALYSIS <nn = näytealojen lkm. (15)>
(16X,15)
....
....
.... <luokitteluvektori tiedostosta for007.dat>
....
....
OUTPUT SUBSETS HIERARCHY LEVEL 2 PRINT <1>
OUTPUT COMPOSITE HIERARCHY LEVEL 3 PUNCH <2>

```

Yhdistäminen voidaan suorittaa joko COPY -komennolla tai editorin avulla.

Käskyllä <1> saadaan EDIT.LIS -tiedostoon jakotason 2 (4 ryhmää) kasvillisuustaulukot.

Käsky <2> tuottaa uuden CEP-tiedoston (EDIT.OUT) jossa on kahdeksan kolmannen jakotason ryhmistä muodostettua keskiarvonäytealaa.

Keskiarvonäytealojen muodostaminen on tarpeen erittäin suuria aineistoja käsiteltäessä. Varsinkin massiivisille aineistoille ajetut ordinaatioanalyysit voivat olla hankalaselkoisia ilman alkuperäisen tiedoston tiivistämistä keskiarvonäytealoiksi.

Keskiarvonäytealoja voidaan muodostaa myös halutuista näytealoista CONSTRUCT MATRIX -proseduurilla.

Esimerkki:

Muodostetaan näytealoista 1-3, 4-6 ja 7-9 keskiarvonäytealat. Uudet runsausarvot lasketaan aritmeettisinä keskiarvoina.

```
CONSTRUCT MATRIX NAMES TYPE(AVERAGES)
COMPOSITE
NAME UUSI0001
INCLUDE SAMPLES 1-3
ENDCOMPOSITE
COMPOSITE
NAME UUSI0002
INCLUDE SAMPLES 4-6
ENDCOMPOSITE
COMPOSITE
NAME UUSI0003
INCLUDE SAMPLES 7-9
ENDCOMPOSITE
ENDCONSTRUCT
OUTPUT CONSTRUCTED MATRIX PRINT,PUNCH TYPE(AVERAGES)
&TITLE (UUSI MATRIISI)
```

Tulostiedosto on nimeltään 'EDIT.OUT'. Kirjoittamalla käskyn "AVERAGES" tilalle "CONSTANCY" saadaan runsauskeskiarvojen sijasta suhteelliset frekvenssit. Käskyllä "NAME" annetaan uusille yhdistelmänäytealoille vapaamuotoinen 8 merkin mittainen nimi, "TITLE" taas tarkoittaa uuden tiedoston otsikkoa.

DATAEDIT varaa muistitilaa automaattisesti riittävän määrän. Havaintomatriisin nollaa suurempien alkioiden määrä ei kuitenkaan saa ylittää 150000.

HUOM!

Kaikki DATAEDIT -käskyt kirjoitetaan ISOILLA kirjaimilla!

### 1.2.3 CONVERT - CEP-TIEDOSTON MUUNTO PERUSMUOTOON

Tiivistettyyn muotoon tallennettu aineisto on muutettavissa perusmuotoiseksi matriisiksi ohjelmalla CONVERT. Merkkitiedon (nimien) käsittely ei kuitenkaan ole mahdollista, joten uusi matriisi sisältää vain numeerista tietoa. Lajit ovat uudessa tiedostossa riveinä ja alat sarakkeina. Lajinimet voi tietysti kirjoittaa käsin jälkeinpäin, jos niitä olettaa tarvitsevänsä.

Muunnettavan CEP-tiedoston otsikkorivin on oltava ehdottomasti sivulla 1-7 kuvatun muotoinen, koska ohjelma lukee alkuparametrinsä ja lukuformaatin juuri näiltä riveiltä.

Ohjelma käynnistyy komennolla:

```
$ RUN CEP:CONVERT
```

Tämän jälkeen ohjelma kyselee syöttötiedoston nimeä ja tulosformaattia. Formaatti annetaan reaalinuotoisena ja kenttien lukumäärän on täsmättävä alojen määrän kanssa. Jos tietueesta tulee liian pitkä (yli 80 merkkiä) se on syytä jakaa / - merkeillä useammille riveille.

Kapasiteetti on 200 alaa ja 300 lajia.

## LUKU 2

### ORDINAATIO- JA LUOKITTELUOHJELMAT

#### 2.1 DECORANA

DECORANA on CEP-ohjelmistoon kuuluva pitkälle kehitetty ordinaatio-ohjelma. Eräät varhaisempia menetelmiä vaivanneet puutteet kuten lajien ja ympäristögradienttien epälineaarisista suhteista johtuva pisteparven kaareutuminen on pyritty korjaamaan. Gauchin (1982) mukaan DECORANA toimii varsinkin pitkillä gradienteilla ja heterogeenisissä aineistoissa muita ordinaatiomenetelmiä paremmin. Ohjelman toimintaperiaate: ks. Hill (1979a), Mikkola ym. (1984).

Syöttötiedoston on oltava CEP-tyyppinen.

Käynnistys:

\$ RUN CEP:DECO

Aluksi ohjelma pyytää syöttötiedoston nimeä. Seuraavat kysymykset ovat englanninkielisiä.

Viesti päätteellä:

```
TYPE 1 IF LOG TRANSFORMATION REQUIRED
TYPE 0 IF NO NO LOG TRANSFORMATION WILL BE MADE
```

Jos tähän vastataan ykkösellä saadaan muunnos  $\ln(X+1)$ . Tällöin lajit tulevat suhteellisesti tasa-arvoisiksi, dominanttien painotus vähenee ja pienipeittävyyksiset lajit korostuvat.

Logaritmuunnos selventää varsinkin prosenttipeittävyysasteikolla tallennetun aineiston tulosten tulkittavuutta. Jos aineistossa on käytetty jo sinänsä korkeita arvoja vaimentavaa luokka-asteikkoa (esim. 1-9 -asteikko), ei muunnosta ole syytä käyttää.

Viesti päätteellä:

```
IS DOWNWEIGHTING OF RARE SPECIES REQUIRED?
TYPE 1 IF YES, TYPE 0 IF NO
```

Joskus aineistossa voi olla näytealoja, joilla on runsaasti muihin aloihin nähden harvinaisia lajeja. Tällöin ordinaatio voi häiriintyä ellei harvinaisten lajien painoa vähennetä. Laskutapa on likimain seuraavanlainen: Olkoon AMAX yleisimmän lajin frekvenssi. Tällöin niiden lajien, joiden frekvenssi on pienempi kuin (AMAX/5) painoa pudotetaan suhteessa niiden frekvensseihin.

Viesti päätteellä:

TYPE 0 IF DETRENDED CORRESPONDENCE ANALYSIS REQUIRED  
TYPE 1 IF BASIC RECIPROCAL AVERAGING

Tällä parametrilla voi valita RA:n tai DCA:n.  
RA eli reciprocal averaging -ordinaatio on Hillin (1973) kasviekologiseen käyttöön kehittelemä metodi, josta DECORANA (DCA) on parannettu versio.

Viesti päätteellä:

SPECIFY NUMBER OF SEGMENTS - OR TYPE 0 FOR DEFAULT VALUE

Oletusarvo on 26 lohkoa. Jos lajeja tai aloja on vähän, esim alle 15 voi seurata virheilmoitus. Tällöin lohkoja voi vähentää neljääntoista. Alle 10 alan tai lajin analysoinneissa voi silti tulla vaikeuksia.

Viesti päätteellä:

ENTER NUMBER OF AXES TO BE PLOTTED

Tällä parametrilla voi määritellä tulostuvien kuvien määrän. Kakkosella saa kaksi kuvaa, 1. akseli vastaan 2. akseli sekä lajeille että aloille. Kolmosella tulostuu 6 kuvaa ja antamalla nelosen saa 12 kuvaa, joka onkin maksimimäärä. Alussa onärkevintä tarkastella vain kahden ensimmäisen akselin muodostamia kuvia, koska yleensä tärkein informaatio sisältyy juuri niihin.

Viesti päätteellä:

ENTER NUMBERS (NOT NAMES) OF SAMPLES TO BE OMITTED  
ONE AT A TIME, ENDING LIST WITH A ZERO

Jos esimerkiksi alat 4, 7 ja 10 pitäisi pudottaa pois analyysistä, annettaisiin seuraavat luvut:

4  
7  
10  
0

Erikoisen poikkeavien näytealojen pudotus analyysistä voi selvittää tuloksen tulkintaa. Tämän option käyttö sekoittaa kuitenkin ordinaatiopisteiden numerointia, joten on mielekkäämpää

poistaa näytealoja käsittelemällä syöttötiedostoa DATAEDIT-ohjelmalla.

Tämän jälkeen laskenta käynnistyy ja ruutuun ilmestyy erilaista "väliaikatieta". Tulostiedosto kirjoittuu levyille käyttäjän hakemistoon nimellä DEC.OUT.

Tuloslistauksen alussa on toistettu alkukyselyt vastauksineen. Ordinaatiopisteet on tulostettu neljälle akselille sekä aloille että lajeille. "RANKED" -sarakkeissa alat ja lajit on järjestetty ordinaatiopisteiden mukaan. EIGENVALUE tarkoittaa akselin ominisarvoa. Mitä korkeampi arvo, sitä paremmin akseli selittää aineiston vaihtelua. Ominaisarvoiltaan selvästi muita pienemmiltä akseleilta ei kannattane etsiä ekologisesti merkittävää informaatiota.

Ordinaatiokuvien reunoilta löytyvät mittayksiköt ovat SD -yksiköitä tuhannella kerrottuna (ks. mm. Gauch 1982).

OVERPRINT TABLE ilmoittaa kuvassa olevista pisteiden päällekkäisyyksistä. Suurten aineistojen, esim. yli 300 alaa, kuvat kannattaa tehdä käsityönä koska päällekkäisyyksien määrä nousee häiritseväksi valmiina tulostuvassa kuvassa.

DECORANAN kapasiteetti on 1000 lajia ja 2000 alaa. Kapasiteettia voi tarvittaessa kasvattaa n. 3000 alaan.

## 2.2 TWINSPAN

TWINSPAN on vuorovaikutteinen, kyselevä ohjelma joka selittää päätteelle tulostuvilla teksteillä hyvin itseään. Syöttötiedoston on oltava CEP-tyyppinen.

Toimintaperiaate: ks. Gauch (1982), Hill (1979b), Mikkola ym. (1984).

Käynnistys:

§ RUN CEP:TWIN

Suurten aineistojen käsittelyt kuluttavat CPU -aikaa niin paljon, että virka-aikana vain eräajot ovat suositeltavia.

Aluksi päätteelle ilmestyy tavanomainen tiedustelu syöttötiedoston nimestä. Loput kysymykset ovat englanninkielisiä.



Viesti päätteellä:

DO YOU WISH TO OMIT SOME SAMPLES?  
 ENTER NUMBERS (NOT NAMES) OF ITEMS TO BE OMITTED  
 ONE PER CARD, ENDING LIST WITH A -1.  
 OTHER NEGATIVE NUMBERS DENOTE SEQUENCES, FOR EXAMPLE  
 A 4 FOLLOWED BY A 8 OMTS ITEMS 4 THROUGH 8.

Vastaus: -1 = ei pudoteta aloja pois  
 Vastaus: 5  
 -29  
 35  
 -1 = pudotetaan alat 5-29 ja 35.

Tätä optiota voi käyttää, jos haluaa analysoida vain osa-  
 aineistoa.

Viesti päätteellä:

ENTER NUMBER (NOT EXCEEDING 9) OF PSEUDOSPECIES CUT LEVELS  
 OR TYPE -1 FOR DEFAULT CUT LEVELS, WHICH ARE 0 2 5 10 20

Jos annetaan -1, ohjelma asettaa pseudolajien muodostuksen  
 kynnysarvot sisäisten oletusarvojen mukaan. Nämä on sovitettu  
 antamaan yleensä käyttökelpoisen tuloksen tavanomaisilla 0-100  
 prosenttipeittävyyksillä. Jos annetaan esim. luku 6, seuraa  
 viesti:

NOW ENTER 6 CUT LEVELS

Kynnysarvot annetaan kukin erikseen. Kun määrä tulee täyteen,  
 arvot kirjoittuvat vielä näkyviin tarkistuksen vuoksi. 1-9  
 -asteikolla tallennetuille runsausarvoille voi kokeilla seu-  
 raavia kynnysarvoja: 0, 1.1, 2.1, 3.1, 4.1, 5.1, 6.1, 7.1,  
 8.1 (9 kpl) tai 0, 1.1, 2.1, 3.1, 4.1, 6.1 (6 kpl). Kynnysar-  
 vojen sopivalla valinnalla voi vaikuttaa erilaisten runsausar-  
 vojen painotukseen.

Viesti päätteellä:

ENTER MINIMUM GROUP SIZE FOR DIVISION  
 TYPE -1 FOR DEFAULT VALUE (=5); OTHERWISE TYPE  
 VALUE REQUIRED WHICH MUST NOT LIE OUTSIDE LIMITS  
 2 10000

Tällä parametrilla asetetaan jaettavaksi kelpaavan  
 ryhmän pienin koko. Oletusarvo on 5.

Viesti päätteellä:

ENTER MAXIMUM NUMBER OF INDICATORS PER DIVISION  
 TYPE -1 FOR DEFAULT VALUE (=7); OTHERWISE TYPE VALUE REQUIRED  
 WHICH MUST NOT LIE OUTSIDE LIMITS 0 15

Pienissä aineistoissa voi indikaattorien määrän vähentää  
 vaikkapa neljään.

Viesti päätteellä:

ENTER MAXIMUM NUMBER OF SPECIES IN FINAL TABULATION  
 TYPE -1 FOR DEFAULT VALUE (=100); OTHERWISE TYPE VALUE REQUIRED  
 WHICH MUST NOT LIE OUTSIDE LIMITS 0 1000

Annetaan luku N; - vain N yleisintä lajia tulostetaan listauksen loppuun ryhmittelyn mukaan järjestettyyn kasvillisuustaulukkoon. Oletusarvo on 100.

Viesti päätteellä:

ENTER MAXIMUM LEVEL OF DIVISIONS  
 TYPE -1 FOR DEFAULT VALUE (=6); OTHERWISE TYPE VALUE REQUIRED  
 WHICH MUST NOT LIE OUTSIDE LIMITS 0 15

Antamalla -1 saadaan korkeintaan  $2^6 = 64$  ryhmää. Aluksi kannattaa kokeilla ryhmittelyä neljällä jaolla (= 16 ryhmää.) Myös jaettavaksi kelpaavan ryhmän minimikoko vaikuttaa lopulliseen lukumäärään.

Viesti päätellään:

TYPE 1 IF DIAGRAMS OF DIVISIONS ARE WANTED  
 TYPE -1 FOR DEFAULT VALUE (=0); OTHERWISE TYPE  
 (INTEGER) VALUE REQUIRED, WHICH MUST NOT LIE OUTSIDE  
 LIMITS 0 1

Jos näytealojen jakoa halutaan tarkastella yksityiskohtaisesti eräänlaisena pienimuotoisena ordinaatiodiagrammina, annetaan ykkönen.

Viesti päätteellä:

TYPE 1 IF MACHINE-READABLE COPY OF SOLUTION TO BE  
 WRITTEN TO DEVICE 7. TYPE -1 FOR DEFAULT VALUE (=0);  
 OTHERWISE TYPE VALUE REQUIRED, WHICH MUST NOT LIE  
 OUTSIDE LIMITS 0 1

Tällä optiolla on käyttöä haluttaessa jaon tuloksista erillinen, vektorimuotoinen tiedosto. DATAEDIT -ohjelman avulla voi sitten muodostaa sen avulla halutun jakotason ryhmistä yhdistelmiä ja analysoida niitä edelleen. Alkuperäisessä englanninkielisessä käsikirjassa on tarkemmat ohjeet menettelyn yksityiskohdista.

Viesti päätteellä:

ENTER WEIGHTS FOR LEVELS OF PSEUDOSPECIES. FOR EXAMPLE  
 WEIGHTS 1 2 2 2 SIGNIFY THAT PSEUDOSPECIES CORRESPONDING  
 TO 3 HIGHER CUT LEVELS ARE TO BE GIVEN TWICE THE WEIGHT  
 OF PSEUDOSPECIES AT THE LOWEST LEVEL.  
 TYPE -1 FOR DEFAULT VALUES (I.E. IF ALL VALUES ARE TO BE  
 SET TO 1) OR TYPE 0 IF NON-DEFAULT VALUES ARE TO BE ENTERED

Jos nyt annetaan 0, saadaan viesti:

```
NOW n INTEGER VALUES MUST BE SUPPLIED
NONE OF WHICH MUST LIE OUTSIDE LIMITS 0 1000000
```

Tällä parametrilla voidaan määritellä painotus eri tasoille pseudolajeille. Jos esim. korkeita peittävyysarvoja haluttaisiin painottaa, voitaisiin 5 kynnyksarvon tapauksessa antaa painot 1 1 1 1 3.

Useimmiten -1 on sopivin vastaus. Jos näytealakoko on hyvin suuri, voi olla järkevää pudottaa 1. tason painotusta, jolloin lajin läsnäolon/poissaolon vaikutus vaimenee.

Viesti päätteellä:

```
ENTER INDICATOR POTENTIALS FOR CUT LEVELS. FOR EXAMPLE
POTENTIALS 1 0 0 1 0 SIGNIFY THAT PSEUDOSPECIES AT
LEVELS 1 AND 4 CAN BE USED AS INDICATORS, BUT THAT
THOSE AT OTHER LEVELS CANNOT. IN THE DEFAULT CASE, ALL
PSEUDOSPECIES ARE AVAILABLE AS INDICATORS. TYPE -1
FOR DEFAULT VALUES (I.E. IF ALL VALUES ARE TO BE SET TO 1)
OR TYPE 0 IF NON-DEFAULT VALUES ARE TO BE ENTERED.
```

Jos annat nollan, seuraa viesti:

```
NOW n INTEGER VALUES MUST BE SUPPLIED
WHICH MUST NOT LIE OUTSIDE LIMITS 0 1
```

Jos lajin läsnäolo halutaan indikaattoriksi, voidaan antaa 1 0 0 0 0. Tämä voi selkeyttää tulosten tulkittavuutta, mutta kvantitatiivisuus osittain poistuu.

Viesti päätteellä:

```
DO YOU WISH TO OMIT SOME SPECIES FROM LIST OF POTENTIAL
INDICATORS? SPECIES OMITTED FROM THIS LIST ARE USED
IN THE CALCULATION, BUT CANNOT APPEAR AS INDICATORS.
ENTER NUMBERS (NOT NAMES) OF ITEMS TO BE OMITTED ONE
PER CARD, ENDING LIST WITH A -1. OTHER NEGATIVE
NUMBERS DENOTE SEQUENCES. FOR EXAMPLE A 4 FOLLOWED
BY A -8 OMITTS ITEMS 4 THROUGH -8.
```

Jos lajeissa on mukana hankalia taksoneita (=epävarma määrittely), voidaan ne pudottaa pois mahdollisten indikaattorilajien joukosta. Jos kaikki otetaan mukaan, vastaus on -1.

Tämän jälkeen laskenta alkaa. Analyysin kulkua voi seurata päätteeltä, koska "väliaikatietoja" kirjoittuu näkyviin.

Tulostiedosto ohjautuu levyllä käyttäjän hakemistoon nimelle TWIN.OUT.

Tuloslistaus voi vaikuttaa alussa sekavalta, mutta tulkinta onnistuu kyllä ohjeita seuraamalla.

\*\*\*\*\*

① DIVISION 1 (N= 77) I.E. GROUP \*  
 EIGENVALUE 0.320 AT ITERATION 2  
 INDICATORS, TOGETHER WITH THEIR SIGN

② CLAD UNCL1(+), DICR ELON1(+), JUNI COMM1(-) CALA LAPP1(+)  
 MAXIMUM INDICATOR SCORE FOR NEGATIVE GROUP 1 STER SPPL1(+)  
 MINIMUM INDICATOR SCORE FOR POSITIVE GROUP 2

③ ITEMS IN NEGATIVE GROUP 2 (N= 48) I.E. GROUP \*0  
 1COEM 2SEM 3LEHT 5LEHT 6COEM 7SEM 8SEM 10COEM 11COEM 12SEM 14VYSO 15VYSO  
 16VYSO 18VYSO 19VYSO 20PAVY 20PAVY 21COEM 22COEM 23COEM 24COEM 25LEHT 26LEHT 27LEHT  
 28LEHT 29LEHT 30VYSO 31PAVY 32COEM 33SEM 37COEM 39COEM 40SEM 41VEM 47 UEV 48 UEV  
 49 UEM 50SEM 51SEM 52COEM 53COEM 54 UVE 59COEM 60ELIP 66COEM 67SEM 73EVT 74 EVT

⑤ BORDERLINE NEGATIVES (N= 6) 47 UEV 73EVT  
 8SEM 14VYSO 19VYSO 20PAVY

⑤ MISCLASSIFIED NEGATIVES (N= 1)  
 59COEM

④ ITEMS IN POSITIVE GROUP 3 (N= 29) I.E. GROUP \*1  
 4SEM 9PALJ 13SEM 34ELIT 35ELIT 36PALJ 38SEM 42PALJ 43ELIT 44ELIP 45ELIT 46PALJ  
 56ELIP 56PALJ 57SEM 58SEM 61ELIP 62ELIT 63ELIT 64ELIT 65PALJ 68ELIT 69ELIT 70ELIT  
 71PALJ 72 EVT 75ELIT 76ELIT 77PALJ

⑤ BORDERLINE POSITIVES (N= 3) 72 EVT  
 4SEM 57SEM

⑥ NEGATIVE PREFERENTIALS  
 JUNI COMM1 ( 36, 2) CARE VAGI1 ( 18, 2) LUZU FLOL1 ( 10, 0) CORN SUEC1 ( 21, 1) GERA SYLV1 ( 10, 0) LYCO ANNO1 ( 20, 6)  
 MELA PRAT1 ( 12, 0) PEDI LAPP1 ( 22, 5) HIER SPPL1 ( 11, 0) TRIE EUROL ( 32, 7) BRAC REFL1 ( 23, 2) JUNI COMM2 ( 28, 0)  
 FEST OVIN2 ( 23, 3) DESC FLEX2 ( 26, 4) CORN SUEG2 ( 10, 1) LINN BORE2 ( 12, 0) HYLO SPLE2 ( 27, 3) JUNI COMM3 ( 19, 0)  
 DESC FLEX3 ( 13, 1) PLEU SCHR3 ( 27, 7) HYLO SPLE3 ( 22, 1) JUNI COMM4 ( 13, 0) HYLO SPLE4 ( 17, 1) HYLO SPLE5 ( 11, 0)  
 BARB COLL5 ( 10, 2)

⑥ POSITIVE PREFERENTIALS  
 ARCT ALP1 ( 3, 6) CALA LAPP1 ( 6, 21) JUNC TRIF1 ( 0, 9) POLY JUNI1 ( 21, 27) POLY PILI1 ( 1, 7) DICR ELON1 ( 4, 20)  
 CHAN SET1 ( 0, 14) CETR NIVAL ( 1, 15) CETR CUCU1 ( 5, 15) CLAD UNCL1 ( 5, 26) CLAD BELI1 ( 4, 17) CLAD PYX11 ( 18, 23) STER SPPL1 ( 7, 19)  
 CLAD STELL ( 2, 17) CLAD AMAU1 ( 6, 18) CLAD UNCL1 ( 0, 13) LEGI GRAN1 ( 4, 20) PTIL CILI1 ( 15, 22) CLAD ALPH1 ( 1, 6) BETU NANA2 ( 13, 17)  
 OCHR FRIG1 ( 1, 17) SPHA GLOB1 ( 0, 3) DICR ELON2 ( 10, 14) DICR FUSC2 ( 3, 6) DICR ELON2 ( 10, 14) DICR FUSC3 ( 6, 11) STER SPPL1 ( 0, 15)  
 POLY JUNI2 ( 3, 6) DICR FUSC2 ( 10, 14) DICR FUSC3 ( 6, 11) STER SPPL1 ( 0, 15) STER SPPL2 ( 0, 15)  
 NEPH ARCT2 ( 3, 6) PTIL CILI1 ( 1, 6) DICR FUSC3 ( 6, 11) STER SPPL1 ( 0, 8) DICR FUSC4 ( 3, 7) STER SPPL4 ( 0, 7)

⑦ NON-PREFERENTIALS  
 BETU TORT1 ( 10, 7) BETU NANA1 ( 18, 21) VACC MYRT1 ( 43, 20) VACC VIT11 ( 44, 29) EMPE HERM1 ( 46, 29) VACC ULIG1 ( 40, 19)  
 PHYL COER1 ( 11, 8) FEST OVIN1 ( 39, 12) DESC FLEX1 ( 36, 12) LINN BORE1 ( 39, 15) SOLI VITR1 ( 34, 11) PLEU SCHR1 ( 43, 19)  
 HYLO SPLE1 ( 38, 13) POLY COMM1 ( 30, 13) DICR SCOP1 ( 26, 16) DICR FUSC1 ( 41, 27) POHL NUT11 ( 35, 20) CETR ISLA1 ( 17, 14)  
 CLAD OCC1 ( 12, 12) CLAD CORN1 ( 19, 11) CLAD CRAC1 ( 31, 26) CLAD RANG1 ( 34, 26) CLAD MIT11 ( 31, 28) CLAD BLEW1 ( 29, 25)  
 NEPH ARCT1 ( 20, 20) FELT APHT1 ( 22, 7) BARB COLL1 ( 40, 22) ORTH ATTE1 ( 19, 15) HEFA SPPL1 ( 9, 6) VACC MYRT2 ( 22, 9)  
 VACC VIT12 ( 16, 14) EMPE HERM2 ( 39, 29) VACC ULIG2 ( 24, 8) PLEU SCHR2 ( 30, 10) DICR SCOP2 ( 7, 6) BARB COLL2 ( 29, 13)  
 BETU NANA3 ( 10, 12) VACC MYRT3 ( 8, 6) VACC VIT13 ( 8, 7) EMPE HERM3 ( 34, 28) VACC ULIG3 ( 10, 4) BARB COLL3 ( 16, 6)  
 BETU NANA4 ( 7, 8) EMPE HERM4 ( 23, 26) PLEU SCHR4 ( 23, 7) BARB COLL4 ( 12, 5) EMPE HERM5 ( 14, 16) PLEU SCHR5 ( 19, 6)

END OF LEVEL 1

\*\*\*\*\*

Taulukko 3. TWINSPAN-ryhmittelyanalyysin 1. jakotason tulostus.

Listauksen alussa ovat mm. alkukyselyt vastauksineen tarkistusta varten. Rengastetut numerot viittaavat taulukkoon 3, jossa on esitetty TWINSPAN -ryhmittelyn ensimmäisen jakotason tulostus. Tulkinta aloitetaan kohdasta DIVISION 1... ①, jossa luetellaan ensimmäisessä jaossa määritetyt indikaattorilajit ② ja positiivinen ④ ja negatiivinen ③ koealaryhmä tunnuksineen. "+" ja "-" indikaattorilajin yhteydessä viittaa ao. ryhmään. Nimen jäljessä oleva numero taas tarkoittaa pseudolajimuodostuksen kynnsarvoa, esim. CLAD UNC1(+), tarkoittaa tämän lajin indikoivan positiivista ryhmää esiintyessään yli 0 % peittävyysinä jos käytetään oletusarvoisia kynnystasoja. Jos nimen perässä oleva luku olisi esim. 3, indikaattoriarvoa ko. lajilla olisi vasta yli 5 % runsauksilla.

"BORDERLINE" ⑤ ja "MISCLASSIFIED" -ryhmät ovat hankalasti luokiteltavia rajatapauksia.

"PREFERENTIALS" -ryhmissä ⑥ ovat yleisimmät lajit (ja pseudolajit) ryhmiteltynä sen mukaan, miten ne esiintyvät ko. jaossa syntyneissä ryhmissä. Suluissa olevat numerot viittaavat esiintymisfrekvensseihin. Esim. LUZU PILO1( 10, 0) tarkoittaa lajin esiintyvän kymmenesti negatiivisessa ryhmässä eikä lainkaan positiivisessa.

"NON-PREFERENTIALS" -luettelossa ⑦ ovat jaon kannalta indifferentit lajit.

Jaot suoritetaan seuraavasti:

DIVISION 1 - koko aineisto kahtia osiin 0 ja 1  
 DIVISION 2 - ryhmä 0 kahtia; tulos 00 ja 01  
 DIVISION 3 - "- 1 -" -" 10 ja 11  
 DIVISION 4 - "- 00 -" -" 000 ja 001  
 DIVISION 5 - "- 01 -" -" 010 ja 011  
 jne...

Alojen jaon tulostus päättyy tekstiin:

THIS IS THE END OF THE DIVISIONS REQUESTED

Tämän jälkeen seuraa listassa lajien jako ryhmiin vastaavasti. Indikaattoreita ei tulosteta.

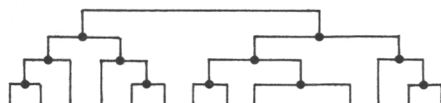
Listauksen lopussa on jaon mukaan järjestetty kasvillisuustaulukko. Ryhmät ovat päävaihtelusuunnan mukaisessa järjestyksessä.

Dendrogrammit lajeille ja aloille voidaan piirtää taulukon oikeassa reunassa ja alareunassa olevien 1/0 koodien mukaan. Malli selittää asian parhaiten:

```

1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
1 1 1 0 0 0 1 1 1 1 1 1 0 0 0
1 1 0 1 0 0 1 1 0 0 0 0 1 0 0

```



TWINSKAN vaatima muistitila riippuu pseudolajien jakotasojen määrästä. Tilantarve riippuu myös aineiston sisäisestä rakenteesta, joten tarkkoja lukuja on mahdotonta antaa. Yläraja lienee n. 1000 lajia + pseudolajia ja 2000 näytealaa.

### 2.3 CLUSTAN - RYHMITTELYANALYYSI

CLUSTAN on polyteettinen, kokoava ryhmittelyanalyysiohjelma, joka tulostaa 2-suuntaisen havaintomatriisin pohjalta havaintojen tai muuttujien hierarkkista samankaltaisuutta kuvaavan dendrogrammin.

Ryhmittelyanalyysien teoriasta on runsaasti tietoa seuraavissa julkaisuissa: Clifford & Stephenson (1975), Gauch (1982), Sneath & Sokal (1973), Virtala (1983), Mikkola ym. (1984).

Syöttötiedoston on oltava perusmuotoinen lajit-näytealat-havaintomatriisi.

Käynnistys:

\$ RUN CEP:CLUSTAN

Aluksi ohjelma kysyy syöttö- ja tulostustiedostojen nimiä sekä rivien (havaintojen) ja muuttujien (sarakeiden) määriä. Kysely jatkuu:

ANNA SYÖTTÖTIEDOSTON FORMAATTI

Formaatti annetaan reaalinuotoisena (F).

HALUATKO RYHMITTELYN HAVAINNOILLE (=0) VAI MUUTTUJILLE (=1)?

Jos vastataan nollalla, saadaan ryhmittely havaintomatriisin riveille ja ykkösellä saa ryhmittelyn sarakkeille. Jos kasvillisuusaineistoille halutaan lajien ryhmittely, on syytä

muistaa, että lajien erilaiset frekvenssit voivat vaikuttaa tulokseen vääristävästi.

HALUATKO STANDARDISOINNIN? KYLLÄ=1, EI=0

Standardisointi tarkoittaa muuttujien arvojen muokkaamista siten, että varianssi, keskiarvo, vaihteluväli tms. vakioidaan. Tässä analyysissä toimenpide suoritetaan siten, että muuttujan keskiarvoksi asetetaan 0 ja varianssiksi yksi (keskistys + jako keskihajonnalla). Jos muuttujat on mitattu eri asteikoilla (esim. pH, korkeus mpy.), on standardisointi suositeltavaa.

ANNA ETÄISYYSINDEKSIN TUNNUS

1=EUKLIDINEN ETÄISYYS

2=LAJISTOETÄISYYS

14=PROSENTTIETÄISYYS ("RENKOSEN INDEKSI")

Tällä parametrilla määritellään kahden havainnon tai ryhmän välinen etäisyysmitta. Yleisimmin käytetty on euklidinen etäisyys, jonka arvo riippuu suuresti käytetystä mitta-asteikosta ja standardisointi on usein tarpeen. Lajistoetäisyys (Sorensen -indeksi) perustuu lajien läsnä- tai poissaoloon. Prosenttietäisyys (Czekanowskin indeksi) antaa usein hyviä tuloksia kasvillisuusaineistojen analyyseissä.

Muita mahdollisuuksia: 2 = Bray-Curtis -indeksi  
 3 = Canberra metric  
 4 = Jaccardin indeksi  
 6 = Simple matching coefficient  
 7 = Ochiain indeksi  
 8 = Fager-McGowan -indeksi  
 9 = Canberra metric  
 12 = Arcus cos theta  
 13 = Cos theta

ANNA YHDISTELYALGORITMIN TUNNUS

NN = LÄHIMMÄN NAAPURIN MENETELMÄ

FN = KAUIIMMAN -"- -"-

GA = RYHMÄKESKIAARVO -"-

C = SENTROIDI -"-

M = MEDIAANI -"-

WE = MINIMIVARIANSSI -"-

FLEX = "FLEXIBLE" -"-

NN (=single linkage) toimii hyvin, jos ryhmät ovat riittävän erillisiä. Mikäli ryhmien välissä on ns. kohinahavaintoja, dendrogrammista tulee sekava. Menetelmä on "heikosti ryhmittelevä", eli havainto liitetään mieluummin johonkin olemassaolevaan ryhmään kuin muodostetaan kokonaan uusi ryhmä.

FN (=complete linkage) on vahvasti ryhmittelevä, mutta tuloksena voi olla sangen heterogeenisia ryhmiä.

GA (=group average, average linkage) antaa usein hyviä tuloksia. Se lienee eniten käytetty yhdistelyalgoritmi.

C (=centroid) on lähinnä NN:n ja FN:n välimuoto. Haittana on dendrogrammin inversioalttius; tällöin jokin ryhmien yhdistäminen tapahtuukin alemmalla tasolla ts. pienemmällä etäisyyshypoteesin arvolla kuin edellinen yhdistäminen. Tällaista dendrogrammia ei ohjelma osaa tulostaa.

M (mediaani) on pitkälle samankaltainen sentroidimenetelmän kanssa.

WE (=Wards error method) on vahvasti ryhmittelevä metodi. Jokaiseen ryhmäpariin liitetään ns. virheneliösumma laskemalla muodostettavan ryhmän keskipiste ja summaamalla yhdistetyn ryhmän havaintojen neliöidyt etäisyydet tästä keskipisteestä. Ne ryhmät yhdistetään, joiden yhdistämisestä aiheutunut varianssin kasvu on pienin. WE antaa usein kohtalaisen selkeitä tuloksia.

FLEX on "joustava" menetelmä jonka ominaisuudet riippuvat erikseen annettavan beta-kertoimen arvosta. Valitsemalla arvo sopivasti väliltä  $-1 \dots +1$  on menetelmä joko havaintoja ketjuttava ( $\beta < 0$ ) tai vahvasti ryhmittelevä ( $\beta > 0$ ).

HALUATKO MUUNNOKSEN?

1 = LN(X+1)

7 = NELIÖJUURI

8 = ARCUS SINI NELIÖJUURI

0 = EI MUUNNOSTA

Eräät etäisyyshypoteesit (euklidinen et., Bray-Curtis -indeksi) ovat herkkiä suurille muuttujan arvoille. Tällöin sopiva muunnos voi selkiyttää tulosta huomattavasti.

ANNA OTSIKKO, MAX. 80 MERKKIÄ

Otsikko on vapaamuotoinen.

Jos pyydetty yhdistelyalgoritmi oli FLEX, saadaan seuraava kysymys:

ANNA BETA-KERROIN FLEXIBLE-ALGORITMIA VARTEN. VOIT KOKEILLA ARVOA  $-0.25$ .

Annetaan luku väliltä  $-1 - +1$ .

ALOITETAAN LASKENTA...

Tulos ohjautuu käyttäjän hakemistoon alussa annetulle nimelle. Tulostuksen alussa on tietoja annetuista ohjausparametreista ym. Dendrogrammi tulostuu joko yhdessä tai kahdessa osassa aineiston koosta riippuen. Jos listauksessa on luetelo yhdistelyetäisyyksistä, muttei dendrogrammia, on syytä vaihtaa yhdistelyalgoritmia.



## LUKU 3

### DIVERSITEETTI-INDEKSIT

Diversiteetti -käsite voidaan jakaa kolmeen pääryhmään (Kosonen 1976):

- (1)  $\alpha$ -diversiteetti, jolla yleensä kuvataan näytealakohtaista lajiston monimuotoisuutta. Sama asia voidaan esittää graafisesti dominanssi-diversiteettikäyrillä tai numeerisesti diversiteetti-indekseillä.
- (2)  $\beta$ -diversiteetti, joka kuvaa lajiston esiintymistä ympäristögradientilla. Esim. yhdeltä metsätyypiltä kuvatun aineiston  $\beta$ -diversiteetti on yleensä alhainen, mutta useita tyyppisiä sisältävän aineiston vastaava arvo on korkeampi. Mm. DCA -ordinaation akseleiden pituudet SD-yksikköinä ovat eräänlaisia  $\beta$ -diversiteetin mittayksiköitä.
- (3)  $\gamma$ -diversiteetillä voidaan kuvata kokonaisen tutkimusalueen ympäristökijöistä johtuvaa kasvillisuuden monimuotoisuutta. Tätä indeksiä voidaan pitää  $\alpha$ - ja  $\beta$ -diversiteettien resultanttina.

DIVER -ohjelma laskee numeerisia  $\alpha$ -diversiteetti-indeksejä. Laskukaavoja on kehitelty lähes lukemattomia, ohjelman tulokset ovat usein käytettyjä ja ominaisuuksiltaan hyvin tunnettuja. Indeksien kaavat löytyvät esim. julkaisusta Jukola-Suonen (1983).

DIVER käynnistetään käskyllä:

§ RUN CEP:DIVER

Syöttötiedoston on oltava perusmuotoinen, lajit riveinä, alat sarakkeina.

Ohjelma on vuorovaikutteinen. Lajien määrä, alojen määrä sekä formaatti on oltava selvillä ennen ajoa. Runsausarvojen skaala on myös annettava päätteeltä. Jos ao. kysymykseen vastaa ykkösellä (=1-9 -asteikko), suoritetaan seuraava si-

säinen muunnos:

Alkup. arvo:	1	2	3	4	5	6	7	8	9
Muunn. -"- :	0.25	0.75	1.5	3.0	6.0	12.0	24.0	48.0	82.0

Indeksit on alunperin tarkoitettu yksilömäärille tai absoluutisille runsausarvoille. Tällaisen muunnoksen vaikutus on kiistanalainen joten tulokseen on syytä suhtautua tietyin varauksin.

Kapasiteetti on 300 lajia ja 200 alaa.

## LÄHTEET

- Clifford, H. T. Stephenson, W. 1975: An introduction to numerical classification. - Academic Press, New York-London.
- Gauch, H.G. 1982: Multivariate analysis in community ecology. Cambridge University Press, New York. 298 pp.
- Hill, M.O. 1973: Reciprocal averaging: An eigenvector method of ordination. - J. Ecol. 61: 237-249.
- Hill, M.O. 1979a: DECORANA - A FORTRAN Program for Detrended Correspondance Analysis and Reciprocal Averaging. Ithaca, New York: Cornell University.
- Hill, M.O. 1979b: TWINSPLAN - A FORTRAN Program for Arranging Multivariate Data in Ordered Two-Way Table by Classification of the Individuals and Attributes. Ithaca, New York: Cornell University.
- Jukola-Sulonen E.-L. 1983: Vegetation succession of abandoned hay fields in central Finland. A quantitative approach. - Comm. Inst. For. Fenn. 112: 1-85.
- Kosonen, M. 1976: A review of the origin, development and significance of ordination methods in present phytoecological research. - Publ. Univ. Joensuu B2 6: 1-24.
- Mikkola, K. J., Pakarinen, P. Oksanen, J. 1984: Ekologisia ATK-menetelmiä: ordinaatio, luokittelu ja taulukointi. - Helsingin yliopiston kasvitieteen laitoksen monisteita 91.
- Sneath, P. A. Sokal, R. R. 1973: Numerical taxonomy. - W. H. Freeman and Company, San Fransisco.
- Toivonen, J. Yrjölä, H. 1978: FORTRAN-77 ohjelmointiopas. OtaDATA ry. 121 s.
- Virtala, M. 1983: Kovarianssianalyysi, ryhmittelyanalyysi ja faktorianalyysi. Metsäntutkimuslaitoksen Rovaniemen tutkimusasema, matemaattinen osasto. Konekirjoite. 48 s.

## MUUTA AIHEESEEN LIITTYVÄÄ KIRJALLISUUTTA

- Austin, M.P. 1976: On non-linear species response models in ordination. - Vegetatio 33: 33-41.
- Austin, M.P. & Belbin, L. 1982: A new approach to the species classification problem in floristic analysis. - Aust. J. Ecol. 7: 75-89.
- Benzecri, J.P. 1973: L'analyse des donnees: II. L'analyse des correspondances. Paris: Dunod.
- Braun-Blanquet, J. 1964: Pflanzensoziologie. 3rd. ed. 865 pp. Wien-New York.
- Bray, J.R. & Curtis, J.T. 1957: An ordination of the upland forest communities of Southern Wisconsin. - Ecol. Monogr. 27: 325-349.
- Dagnelie, P. 1978: Factor analysis. In: Ordination of Plant Communities, ed. R.H. Whittaker, pp. 215-238. The Hague: Junk.
- Ellenberg, H. 1956: Aufgaben und Methoden der Vegetationskunde. - In: Einführung in die Phytologie IV (1). Ed. H. Walter, 136 pp. Stuttgart.

- Fewster, P.H. & Orloci, L. 1983: On choosing a resemblance measure for non-linear predictive ordination. - *Vegetatio* 54: 27-35.
- Feoli, E. & Orloci, L. 1979: Analysis of concentration and detection of underlying factors in structured tables. - *Vegetatio* 40: 49-54.
- Gauch, H.G. 1973: The relationship between sample similarity and ecological distance. - *Ecology* 54:618-622.
- Gauch, H.G. 1977: ORDIFLEX - A Flexible Computer Program for Four Ordination Techniques: Weighted Averages, Polar Ordination, Principal Components Analysis and Reciprocal Averaging, Release B, Ithaca, New York: Cornell University.
- Gauch, H.G. 1982: Multivariate analysis in community ecology. Cambridge University Press, New York. 298 pp.
- Gauch, H.G. & Whittaker, R.H. 1981: Hierarchical classification of community data. - *J. Ecol.* 69: 135-152.
- Goodall, D.W. 1953: Objective methods for the classification of vegetation. I. The use of positive interspecific correlation. *Aust. J. Bot.* 1: 39-63.
- Goodall, D.W. 1954: Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Aust. J. Bot.* 2: 304-204.
- Hajdu, L. 1981: Graphical comparison of resemblance measures in phytosociology. - *Vegetatio* 48: 47-59.
- Harman, H.H. 1967: Modern factor analysis. 2nd. ed. Chicago, University of Chicago Press.
- Hill, M.O. 1973: Reciprocal averaging: An eigenvector method of ordination. - *J. Ecol.* 61: 237-249.
- Hill, M.O. 1974: Correspondence analysis: A neglected multivariate method. - *J. of the Royal Statistical Society, Series C* 23: 340-354.
- Hill, M.O. & Gauch, H.G. 1980: Detrended correspondence analysis, an improved ordination technique. - *Vegetatio* 42: 47-58.
- Hotelling, H. 1933: Analysis of a complex of statistical variables into principal components. - *J. Educ. Psych.* 24: 417-441, 498-520.
- Jaccard, P. 1901: Distribution de la flora alpine dans le Bassin des Dranses et dans quelques regions voisines. - *Bull. Soc. vaud. Sci. nat.* 37: 241-272.
- Jensen, S. 1978: Influences of transformation of cover values on classification and ordination of lake vegetation. - *Vegetatio* 37: 19-31.
- Jukola-Sulonen E.-L. 1983: Vegetation succession of abandoned hay fields in central Finland. A quantitative approach. - *Comm. Inst. For. Fenn.* 112: 1-85.
- Kempton, R.A. & Taylor, L.R. 1976: Models and statistics for species diversity. - *Nature* 262: 818-820.
- Kempton, R.A. - Taylor, L.R. 1978: The Q-statistic and the diversity of the floras. - *Nature* 275: 252-253.
- Kempton, R.A. 1979: The Structure of Species Abundance and measurement of Diversity. - *Biometrics* 35: 307-321.
- Kurimo, H. & Kurimo, U. 1981: Distributional relations and homogeneous areas in aquatic macrophyte vegetation: a case study. - *Ann. Bot. Fenn.* 18: 292-312.
- Kärenlampi, L. 1972: Factor analytic studies on the vegetation of the surroundings of the Kevo Subarctic Station. - *Rep. Kevo Subarctic Res. Stat.* 9: 66-72.

- Lawley, D.N. & Maxwell, A.E. 1971: Factor analysis as a statistical method. 2nd. ed. New York, Elsevier.
- Maarel, E. van der 1979: Transformation of cover-abundance values in phytosociology and its effects on community similarity. - *Vegetatio* 39: 97-114.
- Maarel, E. van der, Janssen, J.G.M. & Louppen, J.M.W. 1978: TABORD, a program for structuring phytosociological tables. - *Vegetatio* 38: 143-156.
- Mikkola, K.J., Pakarinen, P., Oksanen, J. 1984: Ekologisia ATK-menetelmiä: ordinaatio, luokittelu ja taulukointi. - Helsingin yliopiston kasvitieteen laitoksen monisteita 91.
- Mueller-Dombois, D. & Ellenberg, H. 1974: Aims and methods of vegetation ecology. New York.
- Mäkinen, Y. 1974: Tilastotiedettä biologeille. 306 s. Turku.
- Noy-Meir, I. & Whittaker, R.H. 1978: Continuous multivariate methods in community analysis: Some problems and developments. - *Vegetatio* 33: 79-98.
- Oksanen, J. 1981: Reindeer lichen (*Cladina*) vegetation of rock outcrops. - *Ann. Bot. Fennici* 18: 133-154.
- Oksanen, J. 1983a: Vegetation of forested inland dunes in North Karelia, eastern Finland. - *Ann. Bot. Fenn.* 20: 281-295.
- Oksanen, J. 1983b: Ordination of boreal Heath-like vegetation with principal component analysis, correspondence analysis and multidimensional scaling. - *Vegetatio* 52: 181-189.
- Oksanen, L. 1976: On the use of the Scandinavian type class system in coverage estimation. - *Ann. Bot. Fenn.* 13: 149-153.
- Orloci, L. 1978a: Ordination by resemblance matrices. In: Ordination of plant communities. Ed. R.H. Whittaker, pp. 239-275. Junk, The Hague.
- Orloci, L. 1978b: Multivariate Analysis in Vegetation Research. 2nd.ed. Junk, The Hague.
- Pakarinen, P. 1976: Acclomerative clustering and factor analysis of south Finnish mire types. - *Annales Botanici Fennici* 13: 35-41.
- Pakarinen, P. 1982: Etelä-Suomen suo- ja metsätyyppien numeerisesta luokittelusta, - *Suo* 33(4-5): 97-103.
- Pakarinen, P. 1983: Cover estimation and sampling of boreal vegetation in northern Europe. - In: Sampling methods and taxon analysis in vegetation science. Ed. R. Knapp, pp. 35-44. The Hague.
- Pakarinen, P. & Ruuhijärvi, R. 1978: Ordination of northern Finnish peatland vegetation with factor analysis and reciprocal averaging. - *Annales Botanici Fennici* 15: 147-157.
- Podani, J. 1979: Association-analysis based on the use of mutual information. - *Acta Botanica Academiae Scientiarum Hungaricae* 25: 125-130.
- Pritchard, N.M. & Anderson, A.J.B. 1971: Observations on the use of cluster analysis in botany with an ecological example. - *Journal Ecol.* 59: 727-747.
- Shimwell, D.W. 1971: The Description and Classification of Vegetation. London: Sidgwick & Jackson.
- Singer, S.B. 1980: DATAEDIT - A FORTRAN program for Editing Data Matrices. Ithaca, New York: Cornell University.
- Sneath, P.H.A. & Sokal, R.R. 1973: Numerical Taxonomy. San Francisco: W.H.Freeman.

- Sykes, J.M., Horrill, A.D. & Mountford, M.D. 1983: Use of visual cover assessments as quantitative estimators of some British woodland taxa. - *J. Ecol.* 71: 437-450.
- Tuomikoski, R. 1942: Untersuchungen über die Untervegetation der Bruchmoore in Ostfinnland. I. Zur Methodik der Pflanzensoziologischen Systematik. - *Ann. Bot. Soc. Vanamo* 17(1): 1-203.
- Tuomikoski, R. 1949: Entomologian synekologisista tilastoista ja hyönteistuhojen typologiasta. - *Ann. Entomol. Fenn.* 14:
- Westhoff, V. & Maarel, E. van der 1978: The Braun-Blanquet approach. In: *Classification of plant communities*. Ed. R.H. Whittaker, 2nd. ed. The Hague. 101-115.
- Whittaker, R.H. 1956: Vegetation of the Great Smoky Mountains. - *Ecological Monographs* 26: 1-80.
- Whittaker, R.H. 1962: Classification of natural communities. - *Botanical Review* 28: 1-239.
- Whittaker, R.H. 1967: Gradient analysis of vegetation. - *Biological Reviews* 42: 207-264.
- Wildi, O. 1980: Management and multivariate analysis of large data sets in vegetation research. - *Vegetatio* 42: 175-180.
- Williams, W.T. & Lambert, J.M. 1959: Multivariate methods in plant ecology. I. Association-analysis in plant communities. - *Journal Ecol.* 47: 83-101.



**METSÄNHOIDON TUTKIMUSOSASTO**  
**Tutkijaluettelo 1.1.1985 toimipaikoittain**

**HELSINKI**

Lähde, Erkki, MMT, prof.  
Raulo, Jyrki, FT, erikoistutkija  
Jukola-Sulonen, Eeva-Liisa, FT  
Linnilä, Kimmo, FK, MMK  
Lyly, Olavi, FK, MMK  
Nieminen, Jarmo, MH  
Rummukainen, Ukko, MMM  
Savonen, Eira-Maija, FK

**PARKANO**

Laiho, Olavi, MMT  
Kinnunen, Kaarlo, MML  
Raitio, Hannu, FM

**MUHOS**

Valtanen, Jukka, MML  
Kubin, Eero, FT  
Oikarinen, Matti, MH

**ROVANIEMI**

Norokorpi, Yrjö, MMT, erikois-  
tutkija  
Mäkitalo, Kari, MH  
Sepponen, Pentti, FL  
Tikkanen, Eero, FK

**SUONENJOKI**

Lappi, Juha, FK  
Rikala, Risto, MH  
Rossi, Pekka, MH  
Saksa, Timo, MH

**JOENSUU**

Parviainen, Jari, MMT

**Sivulliset tutkijat**

Helle, Timo, FT	Rovaniemi
Huuri, Olavi, MMT, prof. h.c.	Helsinki
Jalkanen, Esko, MH	Helsinki
Leikola, Matti, MMT, prof.	Helsinki
Pelkonen, Paavo, MMT	Suonenjoki
Smolander, Heikki, MMK	Suonenjoki
Sutinen, Marja-Liisa, FK	Suonenjoki
Vapaavuori, Elina, MMK	Suonenjoki