

Denne rapport
tilhører



L&U DOK. SENTER

L. NR. 20084340014

KODE Well 31/3-1 nr 43

Returneres etter bruk

KONF

KONFIDENSIELL

MULTIVARIAT BEHANDLING AV
PALYNOFACIES DATA FRA
TROLLFELTET
BRØNN 31/3 - 1

Olav H.J. Christie
ROGALANDSFORSKNING

10. August 1984

SAMMENDRAG

Palynologiske observasjoner fra prøvene 1322.50 - 1542.00 i brønn 31/3 - 1 er databehandlet ved multivariate teknikker.

Analysen påviser eksistensen av seks lag og overgangssoner, som kan henføres til sedimentologiske prosesser, geografisk beliggenhet og økologiske forhold (side 4 og 5).

INNLEDNING

I brev av 6. juli 1984, ref. U&P LAB-ME/abf, ble det forespurt om en multivariat dataanalyse av kerogendata på Troll. Etter avtale ble data utlevert 31 juli. På grunn av den forholdsvis korte tidsfristen og datamengden ble arbeidet konsentrert om de 49 første prøvene fra 1322.50 til 1542.00 meter. Det ble også foretatt en kort undersøkelse av prøver i det underliggende intervallet 1543.43 til 1567.98 meter.

Data-analysen er foretatt ved hjelp av programpakken SIMCA. En kortfattet innføring i filosofien bak denne programpakken er gitt i Appendiks 1.

METODE

1. Gruppering av prøver. Analysen er basert på den arbeidshypotese at det finnes grupperinger av prøver med sukessiv nummerrekkefølge som representerer segmenter med likeartet palynofascies. For å komme på spor av disse grupperingene ble det først foretatt en egenvektoranalyse basert på det komplette sett med variabler.

2. Karakterisering av geologiske prosesser. Ved ladningsanalyse av hele datamaterialet under ett ble de tre viktigste prosessenes karakteristika kartlagt. Det er prosessenes resultater som fremlegges her, og det er ikke foretatt noen palynologisk/sedimentologisk vurdering av prosessenes art.

3. Test av gruppens relevans. Etter at grupperingene var funnet ble det foretatt en undersøkelse av egenvektorladningene hos hver enkelt, antatt disjunkte, klasse for å finne om det var reelle forskjeller i korrelasjonen mellom hver av gruppenes variabler. Dette er

en multivariat form som er fri for de sterke begrensninger man har i konvensjonell korrelasjonsanalyse.

4. **Kartlegging av gruppeforskjell.** Deretter ble det dannet disjunkte prinsipalkomponentmodeller for hver av de distinkte klassene. Det optimale antallet av prinsipalkomponenter ble fastlagt ved kryssvalidering. Distansen mellom hver av klassene ble beregnet og hver variabels bidrag til distansen (discrimination power) ble listet ut. Denne discrimination power illustrerer resultatet av geologiske og biologiske endringer i de aktive prosesser.

DATA

De anvendte data er semikvantitative og avhenger av mikroskopørens skjønn og faglige erfaring. For å tillempe observasjonene til databehandling må de digitaliseres, og dette må nødvendigvis gjøres til et lite antall heltall. Den tallmessige oppløsningsevnen er altså lav, og en feilvurdering i den semikvantitative mikroskopiske undersøkelsen vil kunne gi store numeriske utslag.

Dataene har derfor sannsynligvis et forholdsvis høyt støynivå og vil da være dårlig egnet til de konvensjonelle plottemetoder. Dette gjør at multivariat databehandling synes særlig attraktiv.

De utleverte data ble digitalisert på følgende måte:

Abundance: Ikke registret = 0, Present = 1, Common = 2, Abundant = 3.

Shape ble gitt som et mål for avrundethet: Ikke registrert = 0, Angular = 1, Sub. Ang. = 2, Sub. rounded = 3, Rounded = 4.

Diversity: Ikke registrert = 0, Low = 1, Medium = 2, High = 3.

Degradation, physical and biological: Ikke registrert = 0, Low = 1, Medium = 2, High = 3.

Av hensyn til begrensninger i det benyttede programmet er 96 variabler benyttet. Variablene Indet. Amorphous, Fungal spores/hyphae og Foraminiferal test linings, som alle er sparsomt registrert i materialet, er utelatt. Dette gjelder også Unstructured vitrinite. De benyttede data er gitt i Appendix 2. En oversikt over de benyttede tallkodene er gitt i Tabell 1.

En del variabler er strøket på grunn av manglende rapportering. Disse er:

- 20 Amorphous vitrinite, shape
- 32 Amorphous vitrinite, physical degradation
- 38 Vitrinite, small, biodegradation
- 39 Vitrinite, medium, biodegradation
- 40 Vitrinite, large, biodegradation
- 44 Amorphous vitrinite, biodegradation
- 46 Inertinite, small, biodegradation
- 47 Inertinite, medium, biodegradation
- 48 Inertinite, large, biodegradation
- 72 Amorphous sapropel, diversity
- 84 Amorphous sapropel, physical degradation
- 87 Spores, medium/large, biodegradation
- 90-96 Biodegradering for marint plankton, ferskvannsalger og amorphous sapropel

Tabell 1. Tallkode for de benyttede variablene

| | | PRESENT | COMMON | ABUNDANT | SHAPE | | | DEGRADATION | | | |
|-----------|------------------------|---------------------|--------|----------|--------|--------|-----|-------------|-----|------|--|
| | | | | | IRREG. | MEDIUM | LOW | PHY. | | BIO. | |
| | | | | | | | | MEDIUM | LOW | | |
| VITRINITE | STRUCTURED | LATHS | 1 | 73 | 25 | 33 | | | | | |
| | | EQUIDIM. | SMALL | 2 | 14 | 26 | 38 | | | | |
| | | | MEDIUM | 3 | 15 | 27 | 39 | | | | |
| | | | LARGE | 4 | 16 | 28 | 40 | | | | |
| | | CORTEX | SMALL | 5 | 17 | 29 | 41 | | | | |
| | | | MEDIUM | 6 | 18 | 30 | 42 | | | | |
| | | | LARGE | 7 | 19 | 31 | 43 | | | | |
| | UNSTRUCTURED VITRINITE | — | — | — | — | | | | | | |
| | INERT. | AMORPHOUS VITRINITE | 8 | 20 | 32 | 44 | | | | | |
| | | LATHS | 9 | 21 | 33 | 45 | | | | | |
| EQUIDIM. | | SMALL | 10 | 22 | 34 | 46 | | | | | |
| | | MEDIUM | 11 | 23 | 35 | 47 | | | | | |
| | | LARGE | 12 | 24 | 36 | 48 | | | | | |

| | | PRESENT | COMMON | ABUNDANT | DIV. | | | DEGRADATION | | |
|----------|-----------------------|------------------|--------|----------|--------|--------|-----|-------------|-----|------|
| | | | | | IRREG. | MEDIUM | LOW | PHYS. | | BIO. |
| | | | | | | | | MEDIUM | LOW | |
| SAPROPEL | PLANT CUTICLE | 49 | 61 | 73 | 83 | | | | | |
| | SPORES | SMALL | 50 | 62 | 74 | 86 | | | | |
| | | MED-LARGE | 51 | 63 | 75 | 87 | | | | |
| | POLLEN | BISACCATES | 52 | 64 | 76 | 88 | | | | |
| | | GYMNOPTERIS | 53 | 65 | 77 | 89 | | | | |
| | | ANGIOSPERMS | 54 | 66 | 78 | 90 | | | | |
| | MEGASPORES | — | — | — | — | | | | | |
| | MARINE MICRO-PLANKTON | LECOSPHERES | 55 | 67 | 79 | 91 | | | | |
| | | ACRITARCHS | 56 | 68 | 80 | 92 | | | | |
| | | PRASIN. | 57 | 69 | 81 | 93 | | | | |
| | | DINOCYSTE | 58 | 70 | 82 | 94 | | | | |
| | | FRESHWATER ALGAE | 59 | 71 | 83 | 95 | | | | |
| | AMORPHOUS SAPROPEL | 60 | 72 | 84 | 96 | | | | | |

RESULTATER

Grupperinger

Grupperingen av 49 prøver (1322.50 - 1542.00) ble studert ved hjelp av alle data, og resultatene er gitt i Figur 1 og Tabell 2. De viser at det foreligger seks veldefinerte palynologiske klasser og noen klare overgangssoner.

Tabell 2. Distinkte grupper.

| Dybde | Prøve nummer | Klasse |
|-------------------|---------------------|---------------------------------|
| 1322.50 - 1351.48 | 1 - 6 | 1 |
| 1356.20 - 1425.86 | 7 - 10, 12, 14 - 18 | 2 |
| 1429.00 - 1447.20 | 19-23 | 3 |
| 1455.60 | 24 | Overgangssone |
| 1461.35 - 1475.66 | 25 - 28 | 4, Antagelig overgangsklasse |
| 1482.45 - 1483.62 | 29 - 30 | Overgangssone |
| 1491.00 - 1517.40 | 31 - 39 | 5 |
| 1519.50 - 1529.33 | 40 - 44 | 6 |
| 1533.62 - 1542.00 | 45 - 49 | Overgangssone |

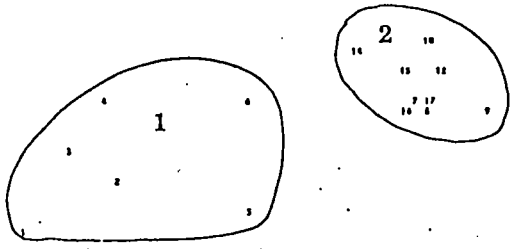
Enkelte prøver innen gruppene er erratiske (outliers, Tabell 3). Det er uvisst om dette avspeiler et reelt forhold, eller om det skyldes unøyaktigheter eller feil i observasjoner eller rapportering.

Tabell 3. Erratiske prøver.

| Dybde | Prøve | Dybde | Prøve |
|---------|-------|---------|-------|
| 1389.78 | 11 | 1469.80 | 26 |
| 1399.80 | 13 | 1551.68 | 53 |

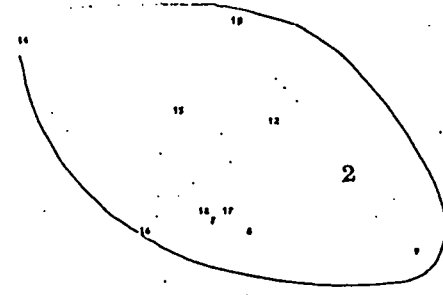
Figur 1. Egenvektorplott som viser grupperingen av prøver.

↑ 2 EV

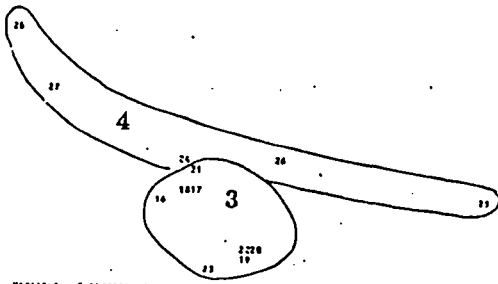


VARIABLE 2 PLOTTED AGAINST VARIABLE 1
FILE: TYPALIN
NUMBER OF VARIABLES: 3
NUMBER OF OBJECTS: 49

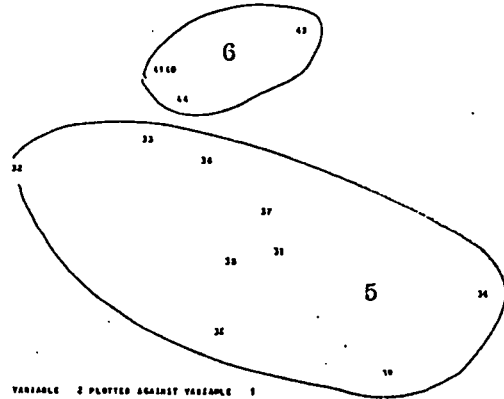
→ 1 EV



VARIABLE 2 PLOTTED AGAINST VARIABLE 1



VARIABLE 2 PLOTTED AGAINST VARIABLE 1



VARIABLE 2 PLOTTED AGAINST VARIABLE 1

Grupperingen er ytterligere undersøkt ved hjelp av disjunkt prinsipalkomponentanalyse. Ved denne analysen lages det en egenvektormodell av hver enkelt klasse, og avstandene mellom hver klasse beregnes som avstand mellom modeller i det multivariate rom der hver variabel er skalert til middelerdi = 1 og standard avvik = 1. Det er for få prøver i klasse 4 (prøve 25 - 28) til å danne egen klassemodell. Den er derfor utelatt fra den disjunkte prinsipalkomponentanalysen som belyser relasjonene mellom klassemodellene.

Klassedistansematrisen er gitt i Tabell 4. En klassedistanse høyere enn 2.0 viser at klassene er godt skilt fra hverandre. Det er særlig diagonalen i matrisen som er interessant, fordi den viser avstanden mellom hvert på hverandre liggende palynofacies stratum. Disse data viser at hver klasse er godt skilt fra hverandre og danner konsistente enheter. En slik struktur vil ikke kunne oppnås ved plotting i konvensjonelle diagrammer.

Tatt i betraktning at tallgrunnlaget er semikvantitativt og at klassedistanseene er basert på hele tallmaterialet, gir klassedistansematrisen en sterk indikasjon på at de beregnede klassene er relle.

Såfremt de økologiske og sedimentologiske forhold hadde gjentatt seg i visse preioder ville dette ført til at klassedistansen mellom de berørte klassene hadde vært lav, det vil si under 1.7. Vi kan derfor slutte at de registrerte observasjoner ikke antyder at historien har gjentatt seg innen det studerte geologiske tidsavsnitt i denne brønnen.

Tabell 4. Klassedistansematrise.

| | 1 | 2 | 3 | 5 |
|---|-----|-----|-----|-----|
| 2 | 4.5 | | | |
| 3 | 6.2 | 3.2 | | |
| 5 | 2.7 | 3.5 | 4.4 | |
| 6 | 6.5 | 3.8 | 3.9 | 5.0 |

De variabler som særlig bidrar til den multivariate forskjellen ved overgang fra en klasse til den neste er gitt i Figur 2.

Figur 2. Variabler som endres vesentlig ved overgangen fra en klasse til den neste. Pluss og minus betyr at variabelens verdi stiger, henholdsvis synker, ved overgang til den underliggende klasse (økende klassenummer).

| | | PRESENT | | SHAPE | | | | | DEGRADATION | | | | |
|-------------------------|----------|---------|----------|---------|----------|-----------|---------|------|-------------|-----|-----------|--------|-----|
| | | COMMON | ABUNDANT | ANGULAR | SUB-ANG. | SUB-ROUND | ROUNDED | HIGH | MEDIUM | LOW | PHY. BIO. | | |
| | | | | | | | | | | | HIGH | MEDIUM | LOW |
| VITRINITE STRUCTURED | LATHS | | | | | | | | | | | | |
| | EQUIDIM. | | | | | | | | | | | | |
| | CORTEX | | | | | | | | | | | | |
| INERT | LATHS | | | | | | | | | | | | |
| | EQUIDIM. | | | | | | | | | | | | |
| | LARGE | | | | | | | | | | | | |

| | | PRESENT | | SHAPE | | | | | DEGRADATION | | | | |
|-------------------------|----------|---------|----------|---------|----------|-----------|---------|------|-------------|-----|-----------|--------|-----|
| | | COMMON | ABUNDANT | ANGULAR | SUB-ANG. | SUB-ROUND | ROUNDED | HIGH | MEDIUM | LOW | PHY. BIO. | | |
| | | | | | | | | | | | HIGH | MEDIUM | LOW |
| VITRINITE STRUCTURED | LATHS | | | | | | | | | | | | |
| | EQUIDIM. | | | | | | | | | | | | |
| | CORTEX | | | | | | | | | | | | |
| INERT | LATHS | | | | | | | | | | | | |
| | EQUIDIM. | | | | | | | | | | | | |
| | LARGE | | | | | | | | | | | | |

1-2

2-3

| | | PRESENT | | DIV. | | DEGRADATION | | |
|----------|-----------------------|---------|----------|------|--------|-------------|--------|------|
| | | COMMON | ABUNDANT | HIGH | MEDIUM | PHY. | | BIO. |
| | | | | | | HIGH | MEDIUM | |
| SAPROPEL | PLANT CUTICLE | | | | | | | |
| | SPORES | | | | | | | |
| | POLLEN | | | | | | | |
| | MEGASPORES | | | | | | | |
| | MARINE MICRO-PLANKTON | | | | | | | |
| | FRESHWATER ALGAE | | | | | | | |
| | AMORPHOUS SAPROPEL | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

| | | PRESENT | | DIV. | | DEGRADATION | | |
|----------|-----------------------|---------|----------|------|--------|-------------|--------|------|
| | | COMMON | ABUNDANT | HIGH | MEDIUM | PHY. | | BIO. |
| | | | | | | HIGH | MEDIUM | |
| SAPROPEL | PLANT CUTICLE | | | | | | | |
| | SPORES | | | | | | | |
| | POLLEN | | | | | | | |
| | MEGASPORES | | | | | | | |
| | MARINE MICRO-PLANKTON | | | | | | | |
| | FRESHWATER ALGAE | | | | | | | |
| | AMORPHOUS SAPROPEL | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

| | | PRESENT | | SHAPE | | | | | DEGRADATION | | | |
|-------------------------|----------|---------|----------|---------|----------|-----------|---------|------|-------------|-----|-----------|--------|
| | | COMMON | ABUNDANT | ANGULAR | SUB-ANG. | SUB-ROUND | ROUNDED | HIGH | MEDIUM | LOW | PHY. BIO. | |
| | | | | | | | | | | | HIGH | MEDIUM |
| VITRINITE STRUCTURED | LATHS | | | | | | | | | | | |
| | EQUIDIM. | | | | | | | | | | | |
| | CORTEX | | | | | | | | | | | |
| INERT | LATHS | | | | | | | | | | | |
| | EQUIDIM. | | | | | | | | | | | |
| | LARGE | | | | | | | | | | | |

| | | PRESENT | | SHAPE | | | | | DEGRADATION | | | |
|-------------------------|----------|---------|----------|---------|----------|-----------|---------|------|-------------|-----|-----------|--------|
| | | COMMON | ABUNDANT | ANGULAR | SUB-ANG. | SUB-ROUND | ROUNDED | HIGH | MEDIUM | LOW | PHY. BIO. | |
| | | | | | | | | | | | HIGH | MEDIUM |
| VITRINITE STRUCTURED | LATHS | | | | | | | | | | | |
| | EQUIDIM. | | | | | | | | | | | |
| | CORTEX | | | | | | | | | | | |
| INERT | LATHS | | | | | | | | | | | |
| | EQUIDIM. | | | | | | | | | | | |
| | LARGE | | | | | | | | | | | |

3-5

5-6

| | | PRESENT | | DIV. | | DEGRADATION | | |
|----------|-----------------------|---------|----------|------|--------|-------------|--------|------|
| | | COMMON | ABUNDANT | HIGH | MEDIUM | PHY. | | BIO. |
| | | | | | | HIGH | MEDIUM | |
| SAPROPEL | PLANT CUTICLE | | | | | | | |
| | SPORES | | | | | | | |
| | POLLEN | | | | | | | |
| | MEGASPORES | | | | | | | |
| | MARINE MICRO-PLANKTON | | | | | | | |
| | FRESHWATER ALGAE | | | | | | | |
| | AMORPHOUS SAPROPEL | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

| | | PRESENT | | DIV. | | DEGRADATION | | |
|----------|-----------------------|---------|----------|------|--------|-------------|--------|------|
| | | COMMON | ABUNDANT | HIGH | MEDIUM | PHY. | | BIO. |
| | | | | | | HIGH | MEDIUM | |
| SAPROPEL | PLANT CUTICLE | | | | | | | |
| | SPORES | | | | | | | |
| | POLLEN | | | | | | | |
| | MEGASPORES | | | | | | | |
| | MARINE MICRO-PLANKTON | | | | | | | |
| | FRESHWATER ALGAE | | | | | | | |
| | AMORPHOUS SAPROPEL | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

Korrelasjoner mellom variabler

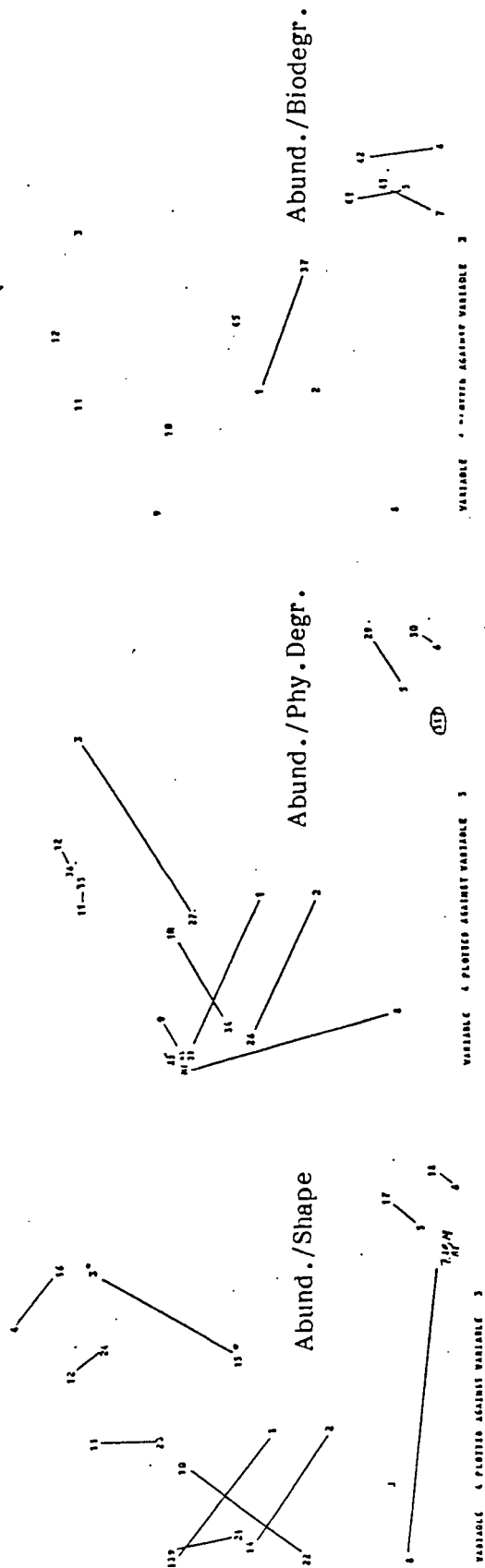
Det er genererte korrelasjoner i data, noe som ligger i observasjonsmaterialets natur. Hver prøve er dels karakterisert ved mikro-fossilers tilstedeværelse, dels ved attributter, som diversitet, form og degradering. Hvis fossilet ikke er registrert, vil heller ikke dets attributter være det. Det er altså en klar korrelasjon mellom abundance og attributter. Likevel er forholdet mellom disse korrelerte variablene meget relevante for karakterisering av både prøvegrupper og av de aktive geologiske prosesser. Der hvor variasjonene i korrelerte variabler er høy, vil de plotte likt i diagram for egenvektorladninger. Der hvor en attributt også er avhengig av andre faktorer enn fossilets tilstedeværelse, vil de plotte et stykke fra hverandre i egenvektordiagrammer.

I disse diagrammene ligger det også mulighet til å trekke ulike og uavhengige påvirkningsfaktorer. De blir her først vurdert kvalitativt, og i det neste kapitlet kvantitativt for å beskrive hver påvirkningsfaktors natur.

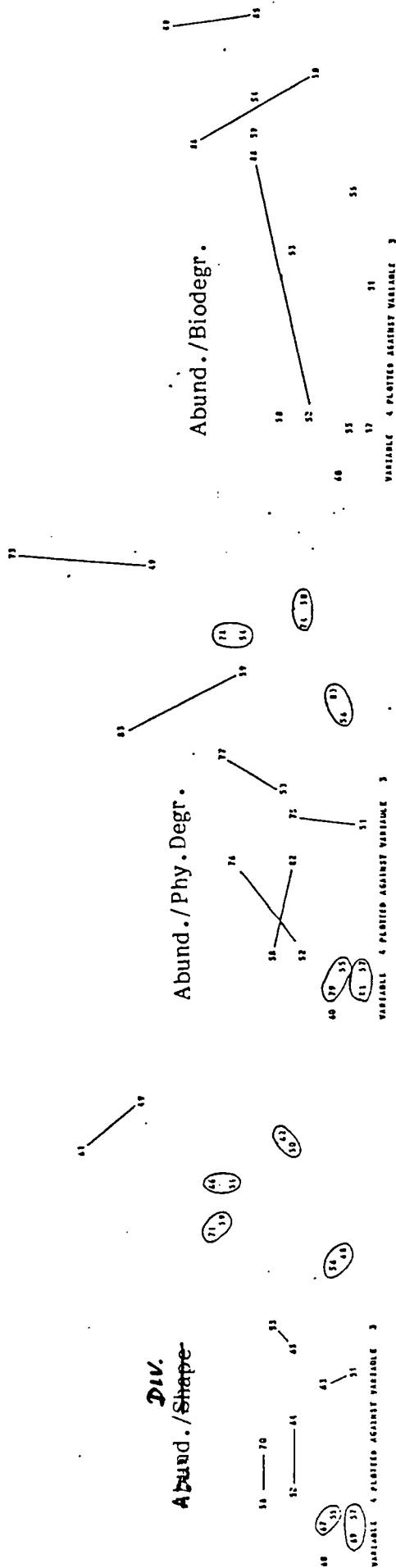
Figur 3 - 6 viser vitrinitt-, inertinitt, og sapropel-variabler for henholdsvis 1. og 2. faktorladninger og 1. og 3. egenvektorladninger.

Rent kvalitativt kan man fastslå at korrelerte variabler ikke plotter på samme sted for vitrinitt/inertinittdata. Dette viser at det ligger mye informasjon langs første og annen egenvektor for vitrinitt-/inertinittdata (lange linjer i diagrammene, Figur 3a - c, 5a - c), mens korrelerte variabler plotter nær hverandre for sapropeldata (nært beliggende punkter, bortsett fra data for fysisk degradering, 4a - c, 6a - c). Vitrinitt og inertinitt synes altså å ha vært utsatt for prosesser som virket sterkere på attributtene (form og fysisk degradering) enn hva tilfellet har vært for sapropel.

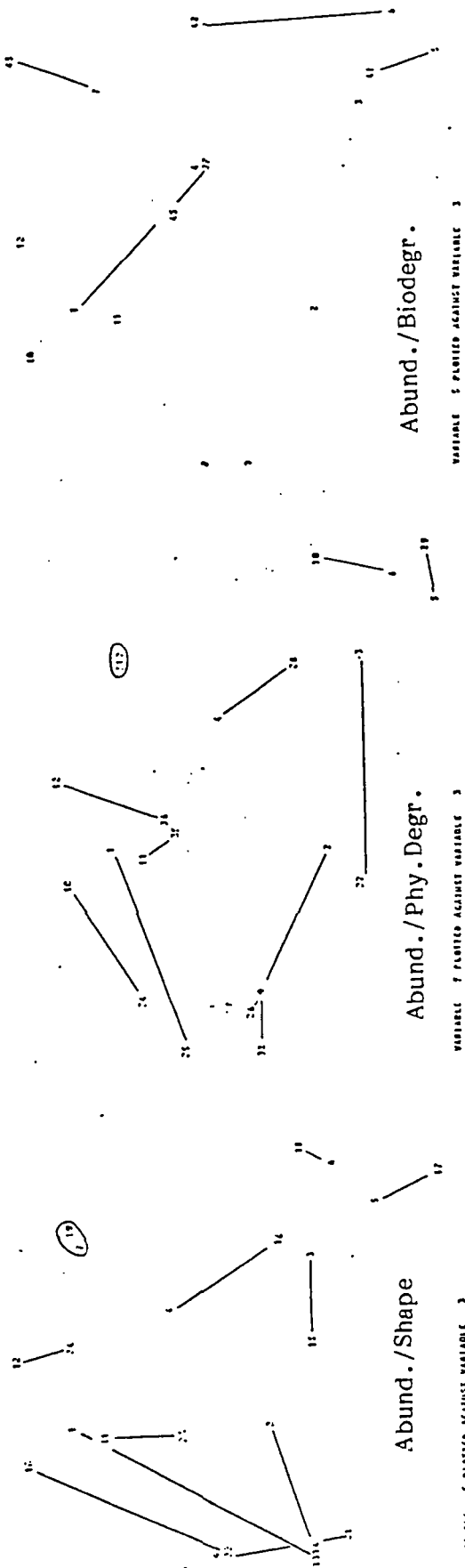
Variablenes informasjonsstruktur vil bli nærmere behandlet i neste kapitel.



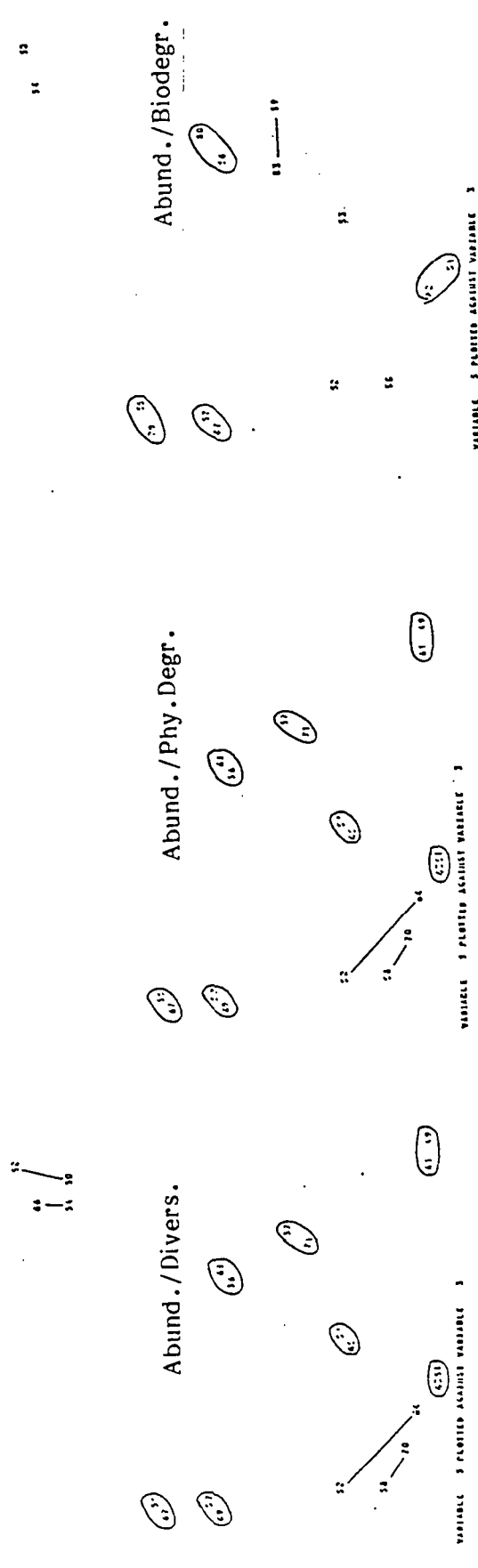
Figur 3a - c. Vitrinitt- og intertinit-variabler for 1. og 2. egenvektor faktorladninger, 1. egenvektor horisontal, 2. egenvektor vertikal. Såfremt sammenhørende variabler ligger omtrent på samme sted i diagrammet, vil de bære lik informasjon. Hvis de ligger langt fra hverandre, som for eksempel variabel 3* og variabel 15* (Abundance og Shape for Medium Equidim. Vitrinite), bærer de informasjon om ulike prosesser.



Figur 4a - c. Sapropelvariabler for 1. og 2. egenvektor faktorladninger.



Figur 5a - c. Vitrinitt- og inertinittvariabler for 1. og 3. egenvektor faktorladninger.



Figur 6a - c. Sapropelvariabler for 1. og 3. egenvektor faktorladninger.

E g e n v e k t o r e n e s b e t y d n i n g

Såfremt to uavhengige prosesser (for eksempel økologisk endring og transport-energi) genererer hvert sitt sett med korrelasjoner, vil disse bli projisert oppå hverandre i det endelige tallresultat.

En av grunnsetningene i moderne multivariat databehandling er at hvis et system er utsatt for ulike, uavhengige prosesser, vil disse bli fanget opp langs ulike egenvektorer. Ved et nærmere studium av egenvektroladningene vil man derfor ha mulighetene til å kartlegge disse prosessene, samtidig som man vil kunne klarlegge hvilke variabel-korrelasjoner som er knyttet til hver enkelt prosess.

Den prosess som gir det største tallmessige utslaget representeres ved første egenvektor. De øvrige prosesser representeres ved egenvektorer av suksessivt lavere rang ettersom hvor stort tallmessig utslag prosessene har forårsaket.

Styrken i den anvendte data-analysen er at den skiller prosessene fra hverandre på en orthogonal måte som gjør at man kan studere hver enkelt prosess' korrelasjoner uavhengig av hverandre. De variabler som er særlig følsomme for en prosess vil fremtre som høye ladninger i den egenvektor som representerer prosessen.

Variabler med høye egenvektorladninger er gitt i Tabell 5. Tabellen antyder følgende muligheter:

Første egenvektor synes å ha tilknytning til en marin faktor, der også den fysiske degradering har noe med prosesser i sedimentet eller under sedimentasjonen å gjøre.

Annen egenvektor synes å ha tilknytning til forholdet marin/terrestrisk påvirkning.

Tredje egenvektor antyder at det er mer enn én prosess (de oppfanget i første egenvektor) som fører til fysisk degradering. Den prosess som er oppfanget i tredje egenvektor fører til mindre markante resultater enn den oppfanget i første egenvektor.

Tabell 5. Variabler med høye egenvektorladninger.

| Nummer | Variabler |
|------------------------------|--|
| Første egenvektor | |
| 6, 18, 30 | Abundance, form og fysisk degradering hos medium cortex |
| 25, 26, 29, 30 33, 34, 79 | Fysisk degradering |
| 55, 57, 58, 60 | Abundance for marint plankton og amorft sapropel |
| Annen egenvektor | |
| 3, 4, 6, 7 | Abundance, medium og large vitrinitt |
| 11, 12 | Abundance, medium og large inertinitt |
| 16, 19, 24 | Form hos stor vitrinitt, cortex og inertinitt |
| Tredje egenvektor | |
| 28, 31, 36 | Fysisk degradering av stor vitrinitt, cortex og inertinitt |

Appendix 1

THE MULTIVARIATE, EIGENVECTOR APPROACH

The vast amount of data in science and technology has raised the need for improved methods for extraction of pertinent information. It has been known for long that some conventional methods, like those based upon correlation coefficients or frequency distributions and probabilities, come short in many problems in technology and science, the major reason being that such methods were based upon preconditions rarely met with in the real life.

A change in this situation was promoted by the introduction of computers, and with every new microprocessor generation new and more general methods became available. In the 1970ies pattern recognition methods were used for the first time in many branches of technology and science, but for historical reasons many of them had the same severe limitations as those of the slide-rule age. For instance, cluster analytical methods so successfully used for classification of plant species sometimes gave disastrous results in geochemistry and somehow discredited statistics as a discipline in geological sciences for many years.

Eigenvector based methods opened for new, generalized solutions some ten years ago, when fast computers of large capacity became available. The more general the method, the larger is the range of validity, and applications in many fields, as different as chemistry and archaeology, gave results being no less than revolutionary.

The basic philosophy of multivariate pattern recognition is that numerical information can be transformed to geometric information without loss of information. If one has three measurements per sample, each of the samples can be uniquely represented by a sample plot in a three-dimensional diagram. The sample point configuration contains all the numerical information, and if the points form, say, two clusters in the diagram, this demonstrates that there are two groups of samples. This is pattern recognition, a notation which unfortunately has been used outside its range of validity for inferior statistical approaches.

However, if we have, say, thirty measurements per sample, we need a thirty-dimensional space for not losing information upon the transformation from numbers to geometric configurations. Therefore, we need methods for handling such spaces, and this is the crucial point in

pattern recognition. Poor methods may work decently in some cases, but suddenly, they give a result being in conflict with the reality without warning us. This is why the choice of methods is crucial.

Forgetting about the academic subtleties of the game, we state that the largest density of information is along the direction of largest point scatter. And this is the direction of the first eigenvector.

We can even find additional eigenvectors at right angles to the first one. The direction of the largest point scatter at right angle to the first eigenvector is the direction of the second eigenvector. We stated that large point scatter is consistent with large information density, and the first-to-second eigenvector plane is, therefore, the plane of projection containing the largest portion of information that can ever be placed in a multidimensional diagram.

Sometimes the information is so complex that even more eigenvectors are needed to extract it. This is no obstacle, since there are as many orthogonal directions (eigenvectors are by definition orthogonal) as there are dimensions.

A final remark before we look at an example: The direction of the eigenvector tells something about what information is important along that eigenvector. This is the key to the study of processes going on in the system we look at. Thus, eigenvector methods is not only used for classifying samples. They are also used for clarifying the correlation between the things measured (the variables), and this type of correlation study is vastly more informative than methods based upon correlation coefficients.

The histogram of Figure 1 is the result of a correlation based cluster analysis of crude oil samples from the Vienna Basin. This diagram, being based upon the content of aromatic compounds, tells clearly that all the studied samples are strikingly similar.

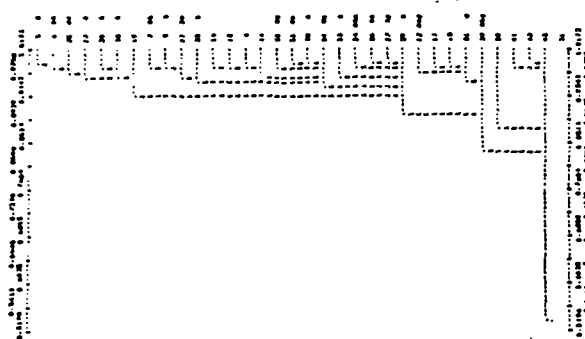


Figure 1. Correlation based cluster analysis fails to demonstrate the existence of four distinct petroleum provinces.

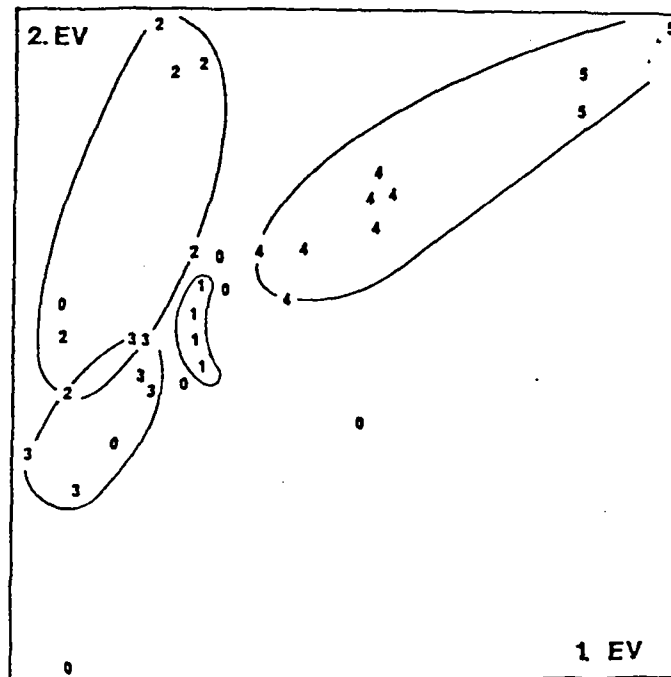


Figure 2. In a first-to-second eigenvector plot the four oil provinces of the Vienna Basin are clearly demonstrated (Christie and Rullkötter, 1984).

A completely different thing is told by the diagram of Figure 2, based upon the same data as that of Figure 1. This is a first-to-second eigenvector projection of the samples, demonstrating that the four studied geographical provinces (marked 1 to 4/5) are also geochemical provinces. In a similar manner variations within an oil field of composite reservoirs can easily be illustrated.

This is a case of a conventional method of narrow range of validity failing to tell a highly pertinent truth that an eigenvector based method tells right away.

There is a last chapter in the story: Figure 3 tells the correlation between selected compounds along the first and the second eigenvector. Along the horizontal direction there is a marked segregation in one group of less polar compounds to the left of the diagram, and another group of strongly polar compounds to the right. This probably reflects the effect of selective adsorption on clay mineral surfaces during migration.

Another, independent process is reflected along the vertical, second eigenvector direction of the diagram. This process leads to a segregation of the less polar group according to molecular weight. It is suggested that this is consistent with a sieving effect.

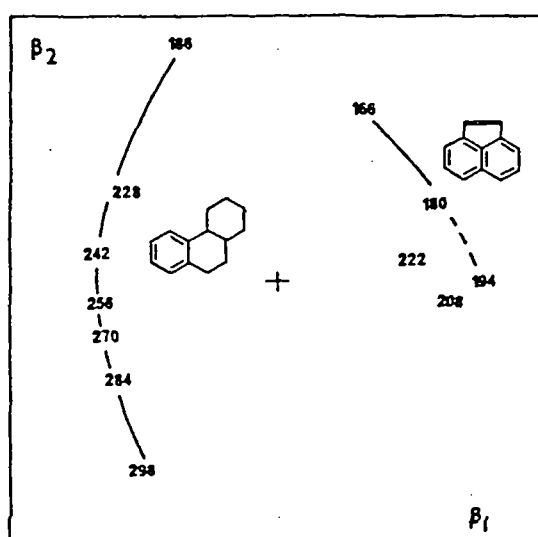


Figure 3. The correlation between two component groups indicate that the two processes adsorption and sieving have acted independently during migration.

Consequently, the eigenvector based analysis of the Vienna Basin aromatics demonstrates that there are four different petroleum provinces, and that the difference has primarily been promoted by two independent processes, probably selective adsorption and sieving during migration. The process along the first eigenvector accounts for the largest fractionation; consequently, selective adsorption is a process more important than sieving during the migration and emplacement of the oil provinces of the Vienna Basin.

At Rogalandsforskning powerful data programs are available for treating problems of multivariate nature. Extraction of complex numerical information has been a routine for several years, and we offer services at a high professional level in this field.