

OR-2212

O-83061

Miljøgifter i ål fra Frierfjorden

En statistisk analyse av data fra 1985

NIVA - RAPPORT

Norsk institutt for vannforskning



NIVA

Hovedkontor
Postboks 33, Blindern
0313 Oslo 3
Telefon (02) 23 52 80
Telefax (02) 39 41 29

Sørlandsavdelingen
Grooseveien 36
4890 Grimstad
Telefon (041) 43 033
Telefax (041) 42 709

Østlandsavdelingen
Rute 866
2312 Ottestad
Telefon (065) 76 752

Vestlandsavdelingen
Breiviken 5
5035 Bergen - Sandviken
Telefon (05) 95 17 00
Telefax (05) 25 78 90

Prosjektnr. O-83061
Undernummer:
Løpenummer: 2212
Begrenset distribusjon:

Rapportens tittel: Miljøgifter i ål fra Frierfjorden. En statistisk analyse av data fra 1985.	Date: 1/3-1988
	Prosjektnummer: O-83061
Forfatter (e): Birger Bjerkgeng.	Faggruppe: Statistikk.
	Geografisk område: Grenland.
	Antall sider (inkl. bilag): 24

Oppdragsgiver: Statens forurensningstilsyn.	Oppdragsg. ref. (evt. NTNF-nr.):
--	----------------------------------

Ekstrakt
Innholdet av polyklorerte hydrokarboner og kvikksølv i ål fanget i Frierfjorden i 1985 er analysert statistisk mot data for lengde, vekt, samt fett-prosent i lever og muskel. Hensikten var å finne ut om individuell variasjon i miljøgift-konsentrasjoner kan kobles til fysiologiske forskjeller. Isåfall kan måle-verdier korrigeres, slik at en får frem sikrere mål på miljøbelastningen, og lettere kan sammenligne forskjell mellom steder eller påvise utvikling over t.d. I dette tilfelle var det ingenting å vinne på en enkel omregning til fettbasis, noe som ellers er vanlig å anta. Ved å bruke alle fire fysiologiske variable kan en forklare ca. 70% av variasjonen for noen av miljøgiftene. Prinsippal komponent analyse antyder en gruppering av materialet. Det er også holdepunkter for at en mer nøyaktig fysiologisk bestemmelse (bl.a. alder, kjønn) kunne forklare denne grupperingen, og gi vesentlig bedre forklaringsgrad. Datasettet er for lite til at det kan trekkes sikre konklusjoner.

4 emneord, norske.

1. Miljøgifter
2. Ål
3. Frierfjorden
4. Statistikk.

4 emneord, engelske:

- 1.
- 2.
- 3.
- 4.

Prosjektleder:

Birger Bjerkgeng

For administrasjonen:

Tor Bokn

ISBN - 82-577-1505-0

NORSK INSTITUTT FOR VANNFORSKNING

0-83061

MILJØGIFTER I ÅL FRA FRIERFJORDEN.
En statistisk analyse av data fra 1985.

Prosjektleder: Birger Bjerken

Dato: 01.03.1989

FORORD

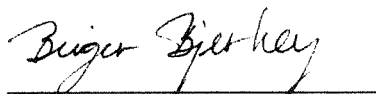
NIVA og Norsk Regnesentral (NR) hadde i årene 1983-1987 et samarbeidsprosjekt på oppdrag fra Statens forurensningstilsyn for å utvikle bruken av statistiske metoder i overvåkningen av vannressurser. Hensikten med prosjektet var bl.a. å finne ut hvordan prøvetakingsopplegg må utformes for å gi svar på spørsmål m.h.t. tidsutvikling.

Den statistiske analysen som presenteres i denne rapporten inngår som et ledd i dette arbeidet. Rapporten gir en grundigere analyse av miljøgiftdata for ål fra Frierfjorden, tidligere rapportert av Rygg et.al. (1985). Hensikten er å vurdere hvordan utsagnskraften kan bedres for slike data, og gi en vurdering av prøveomfang og variabelvalg i datamaterialet.

Arbeidet med rapporten er delvis betalt av SFT innenfor samarbeidsprosjektet med NR (NIVA prosjektnummer O-83061, Tittel: "Bruk av statistikk i overvåkningen"), og delvis også av NIVA's interne forskningsmidler.

Arbeidet er gjort ved NIVA, og rapporten er skrevet av undertegnede, men Sigmund Kalvenes og Kim Esbensen ved NR har gjennomgått rapporten, og kommet med vesentlige forslag til forbedringer.

Forfatteren har også fått verdifulle kommentarer til utkastet fra Gunnar Norheim ved Veterinærinstituttet i Oslo og fra Brage Rygg og John Arthur Berge på NIVA mht. de biologiske tolkningene.



Birger Bjerkeng
Oslo, 01.03.1989

I N N H O L D S F O R T E G N E L S E

SAMMENDRAG OG KONKLUSJON	1
1 INNLEDNING	3
2 DATA-MATERIALET	4
2.1 Rådata	4
2.2 Log-transformering av data	5
2.3 Sammenheng lengde/vekt, eliminering av sannsynlig feil i data	5
2.4 Statistikk på enkeltvariable	7
3 INNLEDENDE ANALYSE AV SAMMENHENDER	8
3.1 Sammenheng mellom enkeltvariable - korrelasjonsmatrise	8
3.2 Bør data regnes om til fettbasis?	9
4 PRINSIPAL KOMPONENT ANALYSE (PCA)	11
4.1 Kortfattet beskrivelse av prinsipal komponent analyse	11
4.1.1 Prinsipal komponent analyse av alle variable sammen	13
4.1.2 Prinsipal komponent analyse av fysiologiske variable	18
4.1.3 Prinsipal komponent analyse på miljøgift-konsentrasjoner	20
4.1.4 Sammenheng mellom miljøgifter og fysiologi	20
4.1.5 Gruppering av materialet - en avsluttende spekulasjon	21
5 Konklusjoner	25

SAMMENDRAG OG KONKLUSJON

Konsentrasjoner av miljøgifter i fisk og blåskjell kan variere med størrelse, alder og andre fysiologiske forhold. For å kunne vurdere miljøbelastning mest mulig sikkert ut fra data for miljøgiftinnhold i organismene, og til en lavest mulig kostnad, er det ønskelig å korrigere for slike fysiologiske forskjeller, slik at miljøeffekten kommer klarere frem. Det vil gjøre det lettere å sammenligne forskjell mellom steder eller påvise utvikling over tid.

I denne rapporten er et gjort en statistisk analyse av et lite datasett for miljøgift-konsentrasjoner i ål som ble fanget i Frierfjorden i 1985. Innholdet av polyklorerte hydrokarboner og kvikksølv er analysert statistisk mot data for lengde, vekt, samt fett-prosent i lever og muskel.

Hensikten med den statistiske analysen er å finne frem til sammenhenger som kan brukes til å korrigere data, og vurdere om de variable som er med i datasettet er tilstrekkelige for dette formålet.

En innledende analyse basert på parvis korrelasjon mellom variablene viser liten grad av åpenbar sammenheng mellom miljøgiftkonsentrasjoner og fysiologiske variable. Det viser seg også at det i dette tilfelle ikke er noe å vinne på en direkte omregning av konsentrasjoner på fettbasis, noe som ellers er vanlig å anta. En slik omregning vil bare kamuflere reelle sammenhenger, ved å innføre tilsynelatende sammenhenger i materialet.

Prinsippal komponent analyse viser tendens til gruppering av materialet i to grupper av individer, med ulik størrelse. Fettinnholdet er bare delvis korrelert med størrelse. Det viser seg at variasjon i miljøgiftkonsentrasjoner i hovedsak er koblet til variasjon i fettinnhold for gitt størrelse, slik at de individene som har høyest fettinnhold sett i forhold til størrelsen inneholder mest miljøgifter. Miljøgiftinnholdet er bare sekundært avhengig av størrelse.

Ved å bruke alle fire fysiologiske variable i sammenheng kan en forklare ca. 60-75% av variasjonen mellom individer for 3-4 miljøgiftene. For de andre variable er det bare 15-40% av variansen som lar seg forklare.

Dersom den observerte tendensen til gruppering av materialet er reell, så ser det ut til at variasjonen i alle miljøgiftkonsentrasjonene vil kunne forklares med 60-80%. Det er indikasjoner på at en mer nøyaktig fysiologisk bestemmelse kunne underbygge en slik gruppering, og gjøre det mulig å oppnå denne forklaringsgraden.

Konkret forslås å utvide måleprogrammet ved å bestemme alder, kjønn og hodebredde for hvert individ i slike prøver, eventuelt også registrere mer direkte fysiologiske indikasjoner på stadium i livssyklusen.

Dette er bare tentative konklusjoner, basert på et nokså spinkelt materiale. For å få frem eventuelle grupperinger sikrere, bør materialet for et år omfatte flere individer, iallfall det dobbelte av det som finnes for 1985. Det vil si at 40-50 individer bør undersøkes hvert år.

1 INNLEDNING

Konsentrasjoner av miljøgifter i fisk og blåskjell kan variere med størrelse, alder og andre fysiologiske forhold. Dette har betydning når en skal overvåke forurensningsstiasjonen i et område over tid, eller kartlegge forskjeller mellom geografiske områder. Enkle gjennomsnittsverdier for ulike prøvesett vil variere både med miljøbelastning og med forskjeller i fordelingen av fysiologiske egenskaper for innsamlede individer. Den fysiologisk betingede "støyen" i datamaterialet vil da gjøre det vanskeligere å påvise endringer i miljøbelastningen, enten det gjelder utvikling over tid eller geografiske forskjeller. Hvis de fysiologiske egenskapene varierer rent stokastisk, dvs. ved tilfeldig utvalg, kan en kompensere for dette ved å ta større prøveserier, altså øke omkostningene. Hvis variasjonene er knyttet til reelle forskjeller i fordeling, f.eks. ut fra populasjonsdynamiske svingninger fra år til år, så vil det ikke hjelpe å ta flere prøver i hvert datasett.

For å kunne vurdere miljøbelastning ut fra data for miljøgiftinnhold i organismene mest mulig sikkert, og til lavest mulig kostnad, er det derfor ønskelig å korrigere for slike fysiologiske forskjeller, slik at miljøeffekten kommer klarere frem.

I denne rapporten er et gjort en statistisk analyse av et lite datasett for miljøgift-konsentrasjoner i ål som ble fanget i Frierfjorden i 1985. Hensikten er å finne frem til sammenhenger som kan brukes til å korrigere data, og vurdere om de variable som er med i datasettet er tilstrekkelige for dette formålet.

En foreløpig statistisk analyse av de samme data er tidligere gitt i en overvåkningsrapport fra NIVA (Rygg et.al. 1985). Her ble det undersøkt om størrelse (lengde og vekt) kunne forklare endel av variasjonene i materialet, slik at en kunne korrigere data og få frem sikrere gjennomsnitt for sammenligning med andre tidspunkter og steder. Analysen viste at det var lite å vinne på dette. Veterinærinstituttet hadde imidlertid også registrert fett-innholdet i prøvene for egen regning, og på initiativ fra Gunnar Norheim ved Veterinærinstituttet er det nå gjort en ny analyse med det utvidede variabelsettet.

2 DATA-MATERIALET

2.1 Rådata

Som et ledd i NIVA's overvåkning av Frierfjorden på oppdrag fra SFT, ble det i 1985 fanget inn noen individer av ål. Prøvene ble analysert for miljøgifter (klor-forbindelser og kvikksølv) ved Veterinær-instituttet i Oslo. Data er tidligere rapportert i Rygg et.al. (1985), men er nå utvidet med data for fettinnhold i lever og muskel, som også fantes tilgjengelig ved Veterinærinstituttet.

Tabell 1 gir en oversikt over data-materialet. Fysiologiske variable er lengde, vekt og fett-innhold i lever og muskel. Konsentrasjoner av miljøgifter er målt både i lever og muskel, for kvikksølv og for de tre klororganiske forbindelsene Hexaklorbenzen (HCB), Oktaklorstyren (OCS) og Dekaklorbifenyl (DCB).

Tabell 1: Data for ål fanget i Frierfjorden høsten 1985.

Fisk nr.	Lengde (cm)	Vekt (g)	Fett-innhold (vekt-%)		Miljøgift-konsentrasjoner							
			Lever	Muskel	LEVER (ug/g)				MUSKEL (ug/g)			
			Hg	HCB	OCS	DCB	Hg	HCB	OCS	DCB		
1(A)	56	453	3.4	23.2	0.69	0.58	0.11	0.02	0.31	4.7	1.0	0.23
2(B)	58	412	6.2	30.0	0.77	0.99	0.32	0.04	0.77	4.4	1.1	0.18
3(C)	50	196	5.5	9.2	0.29	0.85	0.25	0.05	0.45	1.3	0.48	0.12
4(D)	57	398	4.2	24.0	0.47	0.53	0.09	0.02	0.49	4.0	0.78	0.18
5(E)	60	440	6.3	29.7	0.54	1.28	0.29	0.08	0.57	6.1	1.2	0.38
6(F)	52	255	4.8	21.9	0.36	0.90	0.13	0.01	0.60	5.2	1.2	0.23
7(G)	50	200	5.4	11.4	0.27	0.99	0.28	0.09	0.50	2.4	0.81	0.34
8(H)	51	255	13.2	21.8	0.45	1.93	0.48	0.05	0.48	5.1	0.98	0.26
9(I)	71	820	8.2	21.3	0.48	0.21	0.16	0.01	0.17	0.85	0.46	0.05
10(J)	51	306	5.7	26.2	0.44	1.02	0.28	0.04	0.53	5.7	1.3	0.35
11(K)	62	484	8.5	21.5	1.11	0.57	0.14	0.03	0.65	4.8	1.0	0.28
12(L)	67	555	16.8	35.6	2.36	1.46	0.96	0.06	0.69	9.4	5.4	0.42
13(M)	52	306	9.1	40.0	1.88	1.85	0.57	0.08	0.75	7.6	1.9	0.38
14(N)	42	355	5.2	24.6	1.63	1.11	0.89	0.04	0.64	6.3	3.5	0.22
15(O)	56	310	9.2	27.0	0.93	1.38	0.57	0.14	0.80	6.3	1.5	0.52
16(P)	43	125	2.0	3.4	0.27	0.64	0.25	0.03	0.29	0.63	0.26	0.06
17(Q)	41	140	2.3	16.2	3.85	2.10	1.76	0.09	1.45	4.6	2.7	0.19
18(R)	41	122	2.6	13.7	3.22	3.04	2.27	0.04	1.43	11.0	6.4	0.16
19(S)	47	164	1.6	5.6	1.18	0.91	0.85	0.02	0.51	1.4	1.1	0.07
20(T)	43	128	2.3	9.2	0.23	0.38	0.13	<0.01	0.18	4.2	3.5	0.20
21(U)	35	60	2.6	3.8	1.18	1.40	1.53	0.05	0.49	2.0	1.4	0.15
22(V)	34	58	2.2	4.0	1.77	1.12	1.12	0.06	0.63	1.5	1.1	0.06

Prøvene er fra laboratoriet identifisert med nummer 85/035116-035137.

Hg: Kvikksølv HCB: Hexaklorbenzen OCS: Oktaklorstyren DCB: Dekaklorbifenyl

Individ nr. 1-8 er oppgitt å være hanner, for de andre er ikke kjønn angitt. Kjønn kan derfor ikke brukes som forklaringsvariabel i analysen.

2.2 Log-transformering av data

Alle variabel har variasjonsområde fra 0 til $+\infty$, og med ganske stor spredning (se tabell 2 nedenfor), slik at fordelingene blir skjeve. Det er da naturlig å log-transformere data for å få en mer symmetrisk fordeling.

Slik transformering gir for de fleste variablene i dette datasettet også bedre resultater når data testes for normalitet med en test beskrevet i MINITAB (1986) p.49-50. Denne testen er essensielt ekvivalent med Shapiro/Wilks test. Se også Ryan og Joiner, eller Shapiro og Wilks for en beskrivelse av metoden. For dette datamaterialet vil det derfor være bedre å bruke logtransformerte data enn de originale måleverdiene, når en skal benytte tester som forutsetter normalfordeling.

En log-transformering av variablene er også riktigst dersom f.eks. fysiologiske forhold har samme virkning med hensyn til relative konsentrasjons-forskjeller, uavhengig av absolutt konsentrasjonsnivå.

Den videre analysen benytter derfor bare log-transformerte data.

Log-transformering av data innebærer for de statistiske metodene som brukes at sammenhengene beskrives av funksjoner på formen:

$$x_1^\alpha x_2^\beta \dots x_3^\zeta$$

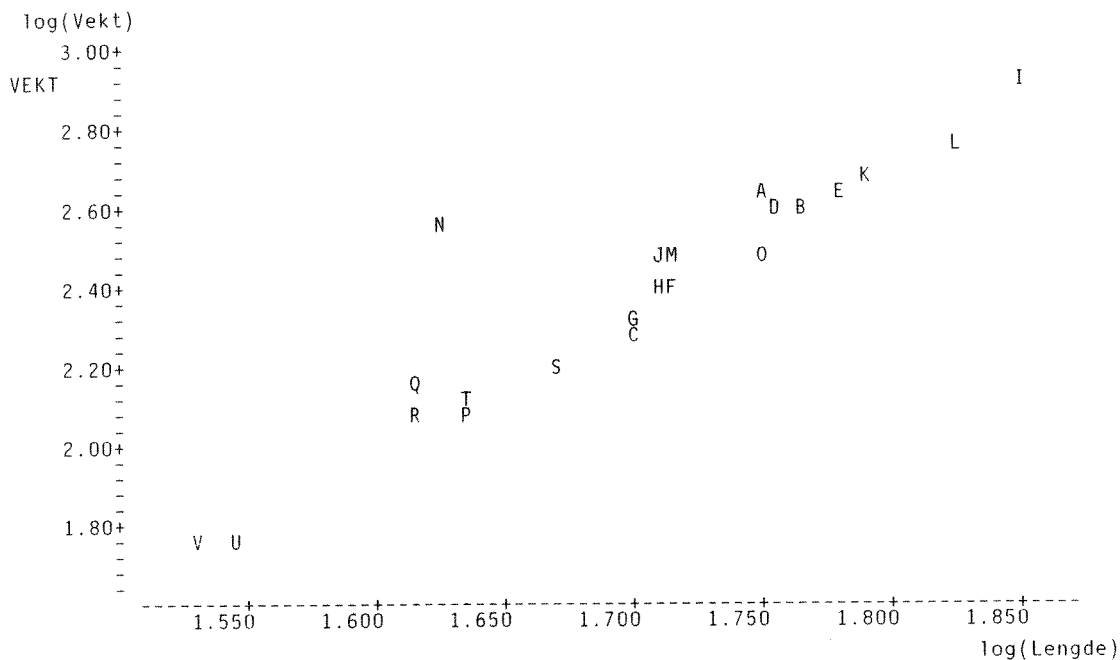
som går over til lineære kombinasjoner ved log-transformering.

2.3 Sammenheng lengde/vekt, eliminering av sannsynlig feil i data

Forsøk på parvis plotting av variable mot hverandre gir ikke noe interessant informasjon, med unntak av en kombinasjon: lengde og vekt.

Av figur 1. ser vi at det er god sammenheng mellom disse to variablene. Figuren viser imidlertid at individ nr. 14, merket N i figur 1, har et unormalt stort vekt/lengde-forhold i forhold til resten av materialet. Ifølge Veterinær-instituttet, som har analysert

prøvene, er det ut fra analyse-arket grunn til å tro at lengden her er feilskrevet på laboratoriet, og dette individet kuttet derfor fra den videre analysen.



Figur 1. Sammenheng mellom lengde og vekt for ål i Frierfjorden, 1985. Lengde i cm, vekt i gram.

Regresjons-linjen for sammenhengen mellom vekt og lengde for de restende 21 individene blir:

$$\text{Log}_{10}(W) = -3.68 + 3.56 \log_{10}(L) ; \text{Students } t(19)=24.17, r^2=0.97$$

Når individ nr.14(N) holdes utenfor er det et maksimalt avvik på 0.116 i $\log_{10}(\text{vekt})$ fra gjennomsnittet for gitt lengde. Det tilsvarer en faktor $10^{0.116} = 1.3$ for vekten. Standardavvik for enkeltindivider rundt regresjonslinjen er 0.056, tilsvarende en faktor 1.14 for vekt. Individ nr. 14 (N) har til sammenligning et avvik fra regresjonslinjen på ca. 8 ganger standardavviket for de punktene som er med å bestemme linjen.

Durbin-Watson koeffisienten er 1.97, meget nær forventet verdi 2 for tilfeldige variasjoner rundt regresjonslinjen. For en beskrivelse av denne testen, se Durbin og Watson (1971).

Den etterfølgende analysen bygger altså på log-transformerte variable, og individ nr. 14 i tabell 1 er kuttet ut.

2.4 Statistikk på enkeltvariable

Tabell 2 gir en enkel statistisk beskrivelse av hver enkelt variabel. Individ nr. 14 i tabell 1 er kuttet ut for sikkerhets skyld, selv om det bare er lengde/vekt data som er påvist å være feilaktig. Dette innvirker ikke mye på verdiene i tabell 2.

Tabell 2. Statistisk karakterisering av data for ål fra Frierfjorden fanget høsten 1985. Individ nr. 14 er ikke med.

	Median (Ca. 95% konfidens intervall)	Middel-verdier		95%-konf.intervall log-normal fordeling		Variasjons- område	
		Aritm.	Geom.	Geometrisk middel	Enkelt- verdi	Min.	Max.
Lengde	51 (45-57)	51	50	(46-55)	(33-76)	34	71
Vekt	255 (150-400)	295	240	(170-330)	(54-1050)	58	820
Miljøgifter i lever:							
Hg	0.7 (0.44-1.2)	1.08	0.76	(0.51-1.12)	(0.13-4.5)	0.23	3.85
HCB	1.0 (0.75-1.4)	1.15	1.0	(0.73-1.3)	(0.26-3.6)	0.21	3.04
OCS	0.3 (0.2 -0.7)	0.60	0.38	(0.25-0.6)	(0.05-2.9)	0.09	2.27
DCB	0.04 (0.02-0.06)	0.049	0.038	(.027-.054)	(.008-.19)	0.01	0.14
i muskel:							
Hg	0.53 (0.48-0.67)	0.61	0.53	(0.42-0.68)	(0.17-1.6)	0.17	1.45
HCB	4.6 (2.2-5.5)	4.44	3.5	(2.4-5.0)	(0.67-18)	0.63	11.0
OCS	1.1 (1.0-1.5)	1.69	1.25	(0.9-1.8)	(0.25-6.2)	0.26	6.4
DCB	0.2 (0.15-0.32)	0.23	0.19	(0.14-0.26)	(0.04-0.8)	0.05	0.52
Fett-%							
Lever:	5.4 (2.6-7.5)	5.8	4.7	(3.5-6.4)	(1.2-19)	1.6	16.8
Muskel:	21.5 (10-25)	19.0	15.2	(10-22)	(3.1-75)	3.4	40.0

3 INNLEDENDE ANALYSE AV SAMMENHENGER

3.1 Sammenheng mellom enkeltvariable - korrelasjonsmatrise

Tabell 3 viser korrelasjons-matrisen for alle variabel-par, beregnet for data omregnet til log-skala. Også her influeres tallene lite av at vi har kuttet ut en observasjon.

Tabell 3. Korrelasjonskoeffisienter mellom log-transformerte variable.
Individ nr. 14 er ikke med.

Koeffisienter som er signifikante med separat sign.nivå 1% er understreket. Kombinasjoner hvor det ikke er naturlig å se noen primær sammenheng er ikke tatt med. (Suffiks -L: i lever, -M: i muskel.)

	LENGDE	VEKT	Hg-L	HCB-L	OCS-L	DCB-L	Hg-M	HCB-M	OCS-M	DCB-M	FETT%-L
VEKT	<u>0.984</u>										
Hg-L	-0.199	-0.152									
HCB-L	-0.372	-0.351	<u>0.607</u>								
OCS-L	-0.537	-0.541	<u>0.771</u>	<u>0.780</u>							
DCB-L	-0.150	-0.160	0.411	<u>0.718</u>	<u>0.578</u>						
Hg-M	-0.190	-0.158	<u>0.730</u>	<u>0.840</u>	<u>0.646</u>	<u>0.629</u>					
HCB-M	0.231	0.289	0.399	0.535	0.146	0.272	<u>0.566</u>				
OCS-M	-0.145	-0.120	<u>0.627</u>	0.540	0.523	0.186	0.513	<u>0.764</u>			
DCB-M	0.352	0.368	0.071	0.406	-0.053	0.458	0.375	<u>0.825</u>	0.459		
FETT%-L	<u>0.747</u>	<u>0.719</u>	-0.042	0.071	-0.187	0.267					
FETT%-M	<u>0.770</u>	<u>0.826</u>					0.239	<u>0.719</u>	0.283	<u>0.694</u>	<u>0.755</u>

Korrelasjons-matrisen viser et sammensatt og tildels tilsynelatende motstridende bilde.

- Mellom de fysiologiske variablene er det endel nokså naturlige sammenhenger:
 - Det er stor positiv korrelasjon mellom lengde og vekt.
 - Det er positiv korrelasjon mellom fett-% i lever og i muskel.
 - Fett-% er positivt korrelert med størrelse. Det kan være at dette egentlig er uttrykk for sammenheng med alder.

- Det er også tegn på innbyrdes sammenhenger mellom mange av miljøgift-konsentrasjonene:
 - Hg i muskel er positivt korrelert med konsentrasjoner i lever generelt.
 - HCB i muskel er korrelert med de andre klorerte forbindelsene i muskel.
 - Konsentrasjonene i lever er i stor grad korrelert med hverandre.
- Sammenheng mellom de observerte fysiologiske variable og miljøgiftkonsentrasjonene er langt svakere:
 - Det er generelt ingen tegn til signifikant sammenheng mellom miljøgift-konsentrasjon og størrelse, selv for HCB og DCB i muskel, som viser økende innhold med økende fett-%.
 - For lever er det ingen tegn til signifikante sammenhenger mellom miljøgift-innhold og fett-%.
 - For muskel viser konsentrasjonen av HCB og DCB positiv korrelasjon med fett-%, mens det ikke er tilfelle for Hg og OCS.

3.2 Bør data regnes om til fettbasis?

Det hevdes av og til at miljøgift-konsentrasjonene bør beregnes på fett-basis, dvs. multipliseres med $100/(\text{fett-\%})$, ut fra den tankegang at miljøgiftene finnes i fett, og at konsentrasjonen i fett derfor vil være det beste uttrykk for forurensningsnivået. Ut fra en slik tankegang burde vi vente mindre spredning i verdiene omregnet til fettbasis, dersom alle individene har levd i det samme miljøet.

I dette tilfelle ser det imidlertid ikke ut til å være noe å vinne på å korrigere konsentrasjonene ved å regne dem om på fett-basis.

For de fleste miljøgift-konsentrasjonene i datasettet ovenfor vil en slik omregning gi større relativ spredning i verdiene enn for de opprinnelige data.

Konsentrasjoner omregnet på fettbasis vil dessuten bli tilsynelatende korrelert med hverandre, fordi en vesentlig del av variasjonen vil skyldes variasjon i fett-%. Sammenhengen mellom fett-% og størrelse

(lengde og vekt), vil også medføre at konsentrasjoner på fett-basis vil vise en ganske sterk negativ korrelasjon med størrelsen (korrelasjon -0.7 til -0.8), som ikke finnes i de opprinnelige data (se tabell 3).

Bare for to variable, nemlig HCB-M og DCB-M, gir korrelasjonsmatrisen grunn til å anta en virkelig sammenheng med fett-%. Disse variablene er imidlertid vel så sterkt korrelert med andre miljøgift-konsentrasjoner, som ikke er bundet til fett-%.

En direkte omregning til fett-basis vil derfor alt i alt bidra til å kamuflere sammenhenger fremfor å klargjøre dem, og den etterfølgende analyse vil ikke bli basert på en slik omregning. I stedet skal vi se på hvordan vi kan kombinere de fysiologiske variable for å forklare variasjonen i miljøgiftkonsentrasjonene.

4 PRINSIPAL KOMPONENT ANALYSE (PCA)

I dette kapitlet skal vi se om det er mulig å få mer ut av materialet ved å bruke Prinsipal Komponent Analyse (PCA). Dette er en multi-variabel analyse-teknikk som prøver å kartlegge de sammenhenger som finnes i et observasjonsmateriale med flere variable som kan være korrelert innbyrdes, uten å gjøre sterke statistiske antagelser.

En slik analyse ser på samvariasjonen mellom flere variable, og finner innbyrdes ortogonale lineære kombinasjoner av variablene, slik at så mye som mulig av variasjonen i materialet fanges opp i så få kombinasjoner som mulig. På den måten kan en ofte redusere antall dimensjoner til et mindre antall underliggende faktorer, og en kan også få fram grupperinger i materialet. Ofte kan slike reduksjoner bidra til å få fram underliggende biologiske og/eller kjemiske trender, som ellers ville være sterkt kamuflert av "støy" i data.

Analysen gjøres på log-transformerte data fra tabell 1, normalisert til standardavvik = 1. (Prinsipal koordinat analyse). Individ nr. 14 fra tabell 1 er fjernet også i den multivariable analysen.

4.1 Kortfattet beskrivelse av prinsipal komponent analyse

Vi tenker oss observasjonene som en punktsky i et flerdimensjonalt rom med de observerte variable langs akse-retningene. Vi finner så et sett av retninger i dette rommet, vinkelrett på hverandre, som er slik at mest mulig av variasjonen i materialet kan beskrives i så få dimensjoner som mulig.

Vi søker først en retning som er slik at punktene i størst mulig grad ligger langs denne linjen, dvs. ligger i punktskyens lengde-retning. Det er 1. hovedkomponent. Deretter søkes den retningen vinkelrett på den første som oppfanger mest mulig av den gjenværende (residuale) variasjonen rundt den første linjen, det er 2. hovedkomponent. Slik fortsetter vi til alle retningene er bestemt.

Matematisk kan dette uttrykkes som en transformasjon (dreining) av aksene i variabelrommet. Med N variable, dvs. akseretninger i variabelrommet har vi:

$$C_i = \sum_k p_{ik} \cdot X_k$$

hvor C_i = hoved-komponent nr. i ($i=1,..N$) (skåring)

X_k = normalisert obs.variabel nr. k ($k=1,..N$)

p_{ik} = komponent-vektorer (ladning)

En kan også snu disse sammenhengene:

$$X_k = \sum_i \alpha_{ki} \cdot C_i \quad ; \quad K=1..N$$

og det kan vises at $\alpha_{ki} = p_{ik}$.

C-verdiene kan da tolkes som underliggende faktorer som innvirker lineært på flere målbare størrelser samtidig. Målet er å finne om det er et mindre antall C-faktorer som forklarer variasjonen i X-variablene.

Mer omfattende introduksjon til PCA er gitt av Afizi & Aspen (1979), av Wold, Esbensen og Geladi (1987), og i Esbensen og Kalvenes (1988).

I praksis gjøres PCA-analysen ved å beregne korrelasjonsmatrisen for de observerte variablene, og så finne egenvektorer med tilhørende egenverdier til denne matrisen.

Egenvektor nr. i inneholder da verdiene $p_{i,k}$ og tilhørende egenverdi uttrykker hvor mye av den totale variansen som ligger langs denne retningen.

4.1.1 Prinsipal komponent analyse av alle variable sammen

Vi vil først se på hele materialet samlet, og utfører PCA-analyse på alle 12 variable. Dette gir følgende resultat:

Tabell 4. Resultat av PCA analyse på alle variable.

Individ nr.14 er ikke tatt med.

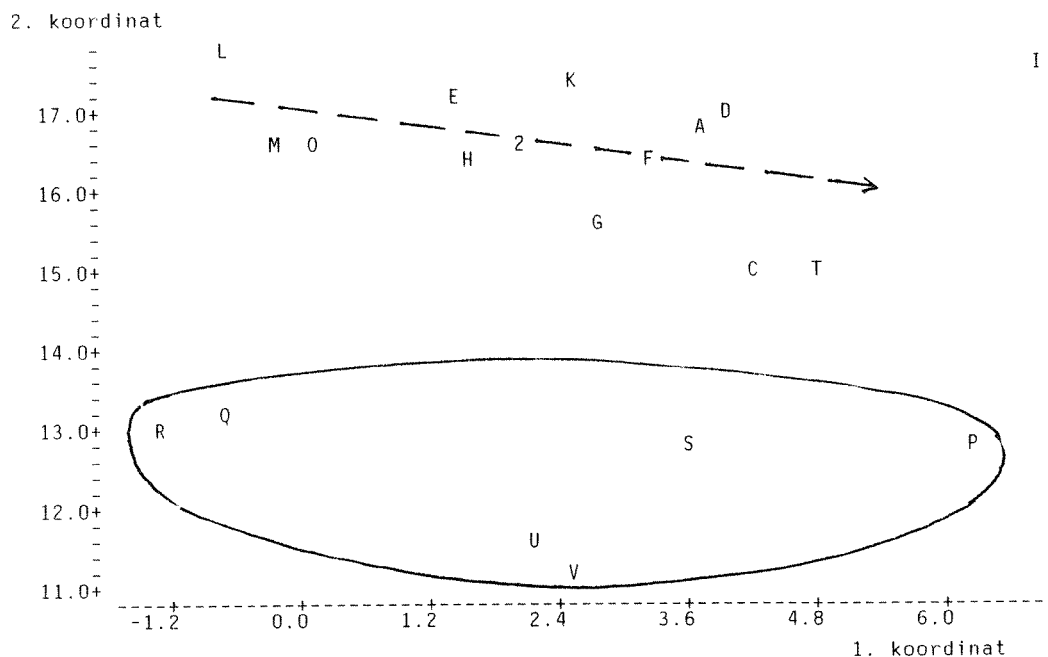
(Suffiks (l): i lever,
(m): i muskel)

Variable (log ₁₀ -transf.) v(i); i=1,...,12	Komponent-vekter for	
	1.koordinat p(1,i)	2.koordinat p(2,i)
Lengde	0.043	0.452
Vekt	0.028	0.456
Hg(l)	-0.338	-0.131
HCB(l)	-0.401	-0.141
OCS(l)	-0.310	-0.295
DCB(l)	-0.314	-0.056
Hg(m)	-0.396	-0.079
HCB(m)	-0.354	0.212
OCS(m)	-0.342	-0.017
DCB(m)	-0.282	0.279
Fett-%(l)	-0.125	0.389
Fett-%(m)	-0.183	0.424
Forklart varians	40 %	36%

Variasjonen langs første koordinat-akse forklarer 40% av den totale variansen i materialet, mens den andre aksene forklarer 36%. Tilsammen beskriver de to ligningene altså 76% av variansen i materialet, og begge er omtrent like viktige. Det er altså ikke mulig å finne en enkelt lineær kombinasjon av log-variable som beskriver det meste av variasjonen.

Hvis en tenker seg datapunktene i et 12-dimensjonalt rom, så ligger dataene rundt det "planet" som er beskrevet av disse to aksene i dette rommet. Variansen rundt dette planet er bare 24% av den totale variansen i dataene, og det kan derfor mesteparten oppfattes som stokastisk støy.

Hvis vi plotter observasjonene projisert inn i det planet som de to koordinat-retningene C_1 og C_2 beskriver, får vi et bilde som vist i figur 2. Bokstavene identifiserer prøvene, se tabell 1.



Figur 2. PCA-analyse av alle 12 variable. Observasjoner projisert inn i hovedplanet i variabelrommet.

I figuren skiller prøvene R,Q,P,S, U og V seg ut i en gruppe med lave verdier i 2. koordinat. Av tabell 1 ser vi at denne gruppen omfatter nesten alle de minste individene, med vekt fra 60 til 160 gram. Prøve T hører hjemme i den samme gruppen ut fra vekten, men plasserer seg sammen med hovedgruppen i PCA-plottet, og avviker altså fra de andre små individene.

Prøve I, som er det klart største individet, skiller seg ut i motsatt retning.

De andre observasjonene kan se ut til å samle seg rundt en linje, hvor 1.komponent avtar noe med økende 2.komponent.

De beregnede komponent-vektene i tabell 4 viser at lav verdi i 1.komponent i hovedsak er et uttrykk for høye konsentrasjoner av miljøgifter, til en viss grad også høyt fettinnhold, og uavhengig av størrelsen (både vekt og lengde).

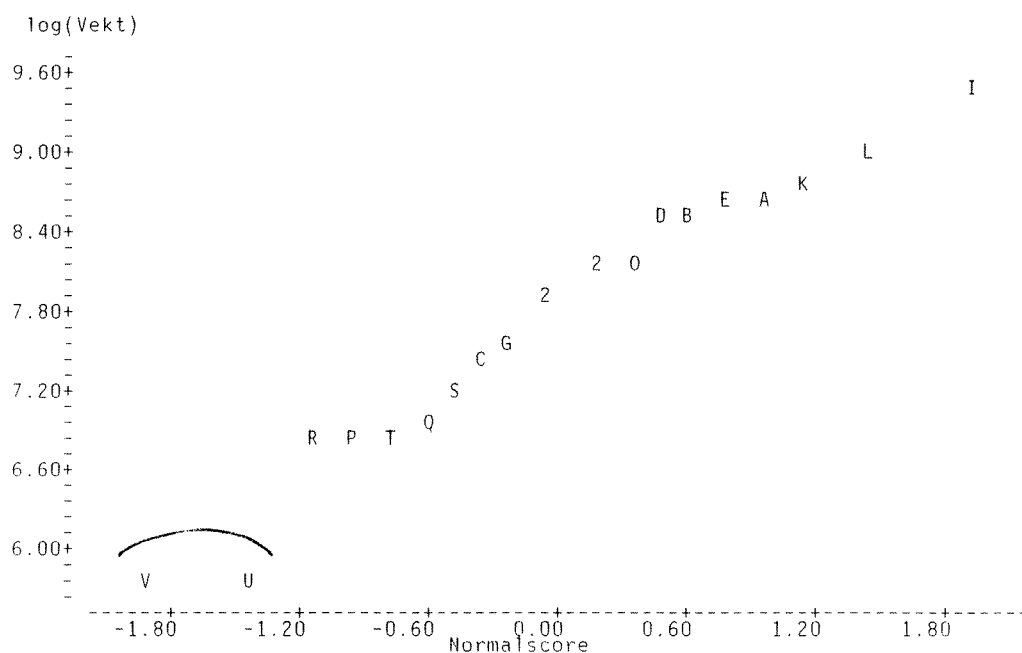
2.komponent er derimot i stor grad et uttrykk for størrelsen, i kombinasjon med fettinnhold i positiv retning. Innenfor hovedgruppen av punkter får vi altså frem en antydning til forventet positiv

korrelasjon mellom størrelse og konsentrasjon, men det ser ikke ut til å være noen forskjell i miljøgift-konsentrasjoner mellom de to gruppene som skilles ut av 2. komponent, med eventuelt unntak for OCS(lever).

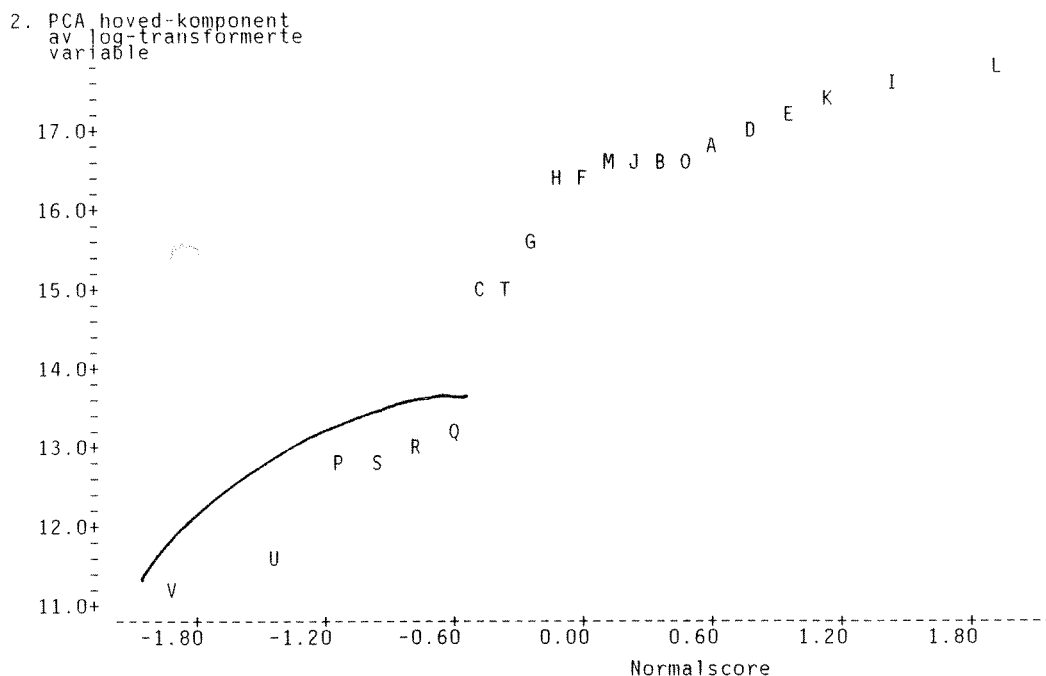
Siden det er relativt få punkter, skal vi ikke legge for stor vekt på tendensen til gruppering i PCA-plottet, det skal ikke så mange punkter fra eller til for å forandre bildet. Tendensen til gruppering kan på den andre siden ikke fullstendig avvises.

Hvis det er en reell gruppeinndeling, så er det imidlertid nærliggende å se det som en oppdeling i årsklasser.

I figur 3 og 4 er vist normalscore-plott av henholdsvis vekt og av 2. hovedkomponent beregnet ovenfor. Dersom materialet er normal-fordelt skal verdiene ligge på en rett linje. For vekten er det bare de to minste individene (U og V) som skiller seg klart ut fra en normal-fordeling. For 2. hovedkomponent, som i stor grad kombinerer alle fysiologiske variable, vises den oppdelingen som er drøftet ovenfor, idet også prøvene P,Q,R,S,U og V nå skiller seg ut fra hovedgruppen.



Figur 3. Kumulativt normalfordelingsplott, log(vekt).

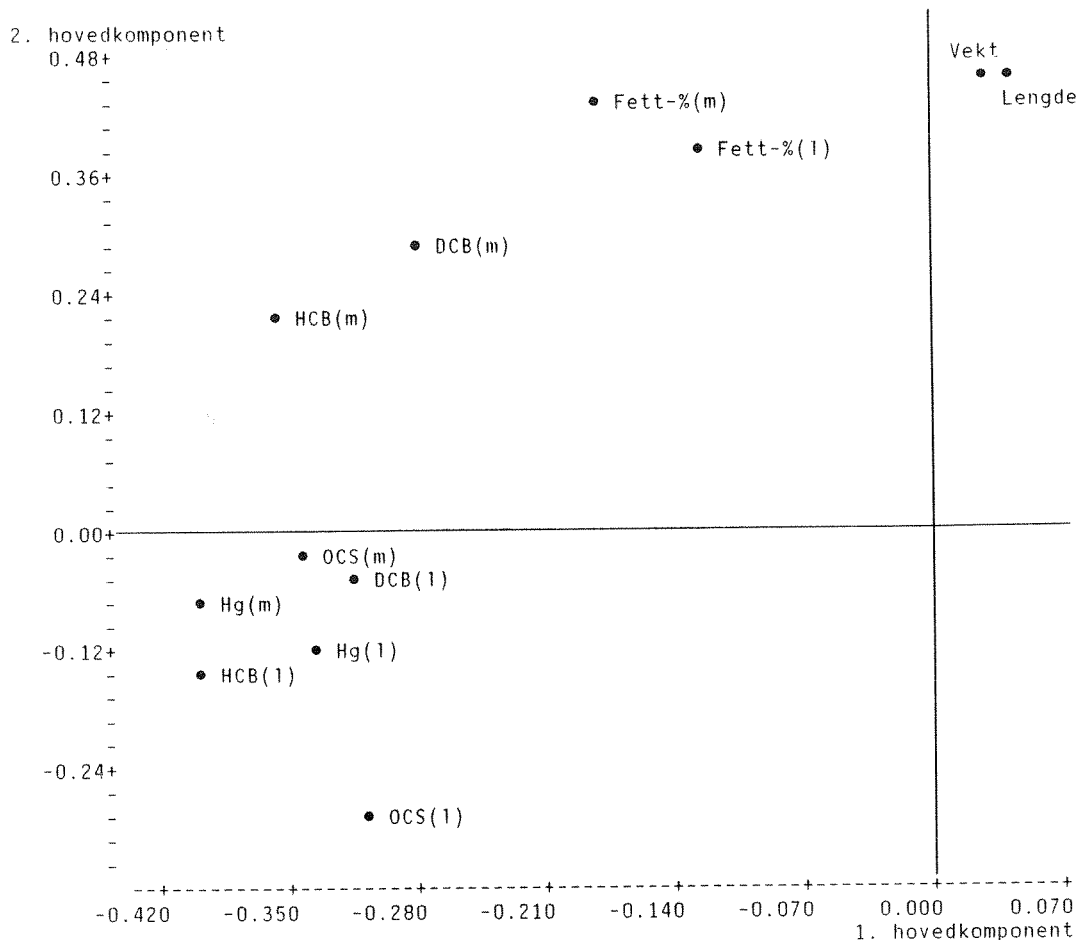


Figur 4. Kumulativt normalfordelingsplott, 2.PCA komponent.

Det må understrekes at dette er tendenser i et materiale på 21 individer. En reell aldersgruppering kan være en annen, og delt inn i flere klasser, det kan for eksempel godt være at U og V kommer fra en egen årsklasse. Det kan også være andre grupperinger som ligger bak, f.eks. kjønn.

For å se nærmere på koblingen mellom variablene er det også tatt ut et ladnings-plott. Her er hver variabel tegnet som et punkt mot de to hoved-komponentene, og koordinat-verdiene er de komponent-vektorer, eller (ladninger) som variabelen har i de to hovedkomponent-retningene. Hvis to punkter ligger i omtrent samme retning fra origo viser det at de to variablene i stor grad varierer likt, dvs. er sterkt positivt korrelerte.

I dette tilfelle fås figuren nedenfor. Origo er markert med et kryss, og hver variabel er markert med et punkt sammen med variabelnavnet.



Figur 5. PCA-analyse på alle variable. Variablene projisert inn i hovedplanet i observasjonsrommet. (l): lever, (m): muskel.

Diagrammet bekrefter grafisk det som er sagt foran:

Økning i størrelse, dvs. lengde og vekt, er tildels koblet til økende fett-innhold, og dette er vesentlig uttrykt i hovedkomponent 2.

Miljøgift-konsentrasjonene er i stor grad uavhengig av variasjon i størrelse, og knyttet til hovedkomponent 1.

Delvis unntak fra dette er HCB(muskel) og DCB(muskel), som til en viss grad er positivt korrelert med fettinnhold, og OCS(lever), som er noe negativt korrelert med størrelse.

Generelt ser det ut til at de forskjellige miljøgift-konsentrasjonene varierer på ganske forskjellige måter. Det kan derfor være interessant

å se på dem enkeltvis, eller i mindre grupper, og se om de kan beskrives som funksjon av de fysiologiske variable (vekt, lengde, fett-% i lever og fett-% i muskel).

Først skal vi imidlertid studere de fysiologiske variable og miljøgiftkonsentrasjonene hver for seg.

4.1.2 Prinsipal komponent analyse av fysiologiske variable

Tabell 3 viste at det er signifikant sammenheng mellom alle de fysiologiske variable:

lengde,
vekt,
fett-% i lever
fett-% i muskel.

Ved hjelp av Prinsipal komponent analyse kan vi finne innbyrdes uavhengige kombinasjoner av de fire variablene. Ved å bruke slike kombinasjoner som nye variable kan vi kanskje få et bedre bilde av hva det er som innvirker på konsentrasjonene.

En PCA-analyse på disse 4 standardiserte variable, med observasjon nr. 14 kuttet ut, gir følgende resultat:

Tabell 5. PCA komponentvekter for de fire fysiologiske variable.

Individ nr. 14 er holdt utenfor.
Usikkerhet som 95% konfidensintervaller.

Komp.vekter p ik		PCA komponent nr. i=1,..4			
		1	2	3	4
Variabel	k				
Lengde	1	-0.516±0.014	0.426±0.50	0.327±0.62	-0.667±0.062
Vekt	2	-0.520±0.013	0.455±0.14	0.056±0.68	0.721±0.048
Fett%(l)	3	-0.470±0.040	-0.754±0.63	0.447±1.09	0.102±0.097
Fett%(m)	4	-0.491±0.034	-0.207±1.19	-0.830±0.28	-0.160±0.105
Prosent av total varians		85.1	8.8	5.9	0.2

Usikkerhetene i akse-retningene er bestemt ved den såkalte "jack-knife" metoden, se. f.eks. Cochran (1977). Usikkerheten vises som 95% variasjonsbredde, altså ca. 2. ganger standard-avvik for størrelsene ovenfor. Vi ser at 2. og 3.komponent er meget usikkert bestemt, mens 1. og 4.komponent er ganske sikre. Som det fremgår av tabellen er det tatt med flere desimaler enn hva som er signifikant.

Første komponent kan tolkes som en slags normal-retning for vekst, med en parallel variasjon i lengde, vekt og fett-prosent i lever og muskel. Den store andel av variansen som faller langs denne akse uttrykker at det i stor grad er samvariasjon mellom alle de fire variablene.

På tvers av dette karakteriseres materialet av to typer variasjon:

2. komponent representerer en slags "relativ magerhet" (høy verdi for 2.komponent betyr lav fett-% i forhold til lengde og vekt),
3. komponent uttrykker hovedsakelig en variasjon i forholdet mellom fett i lever og fett i muskel.

De to faktorene er imidlertid meget usikkert bestemt, og det eneste en kan si er at den sekundære variasjonen på tvers av hovedretningen ligger i det planet i variabel-rommet som komponent 2 og 3 beskriver, dvs. i variasjoner i fettinnhold for lever og muskel for en gitt individ-størrelse.

Faktor 4 er i hovedsak et vekt/lengde-forhold, og den lave variansen uttrykker bare det vi har sett før, at det er svært god korrelasjon mellom lengde og vekt. Det betyr imidlertid ikke nødvendigvis at forholdet vekt/lengde er uinteressant som forklaringsvariabel. Ifølge Tesch, 1977 vil vekt/lengde-forholdet være forbundet med alder, kjønn og andre fysiologiske forskjeller som ikke finnes separat i datamaterialet, men som kan ha gi forskjeller i miljøifftkonstrasjoner.

Hver PCA-komponent definerer et plan vinkelrett på akse-retningen. Når den siste komponenten har så lav andel av variansen, betyr det at alle observasjonene stort sett ligger i et slikt plan, som er omtrent det samme som det en vil få ved multippel-regresjon av en av variablene mot de andre.

4.1.3 Prinsipal komponent analyse på miljøgift-konsentrasjoner

Dersom vi analyserer miljøgift-konsentrasjonene for seg får vi følgende resultat:

Tabell 6. PCA-analyse av miljøgift-data, komponentvekter.

Individ nr. 14 er ikke med.

Variable v(i)	Komponentvekter for:	
	1.koordinat C1,m p(1,i)	2.koordinat C2,m p(2,i)
Hg(1)	-0.363	0.282
HCB(1)	-0.421	0.114
OCS(1)	-0.350	0.465
DCB(1)	-0.323	0.115
Hg(m)	-0.412	0.094
HCB(m)	-0.332	-0.513
OCS(m)	-0.349	-0.191
DCB(m)	-0.250	-0.608
Andel av var.	58%	20%

Variasjonen langs første koordinat-akse forklarer nå 58% av den totale variansen i materialet, mens den andre aksene forklarer 20%, tilsammen 78%.

4.1.4 Sammenheng mellom miljøgifter og fysiologi

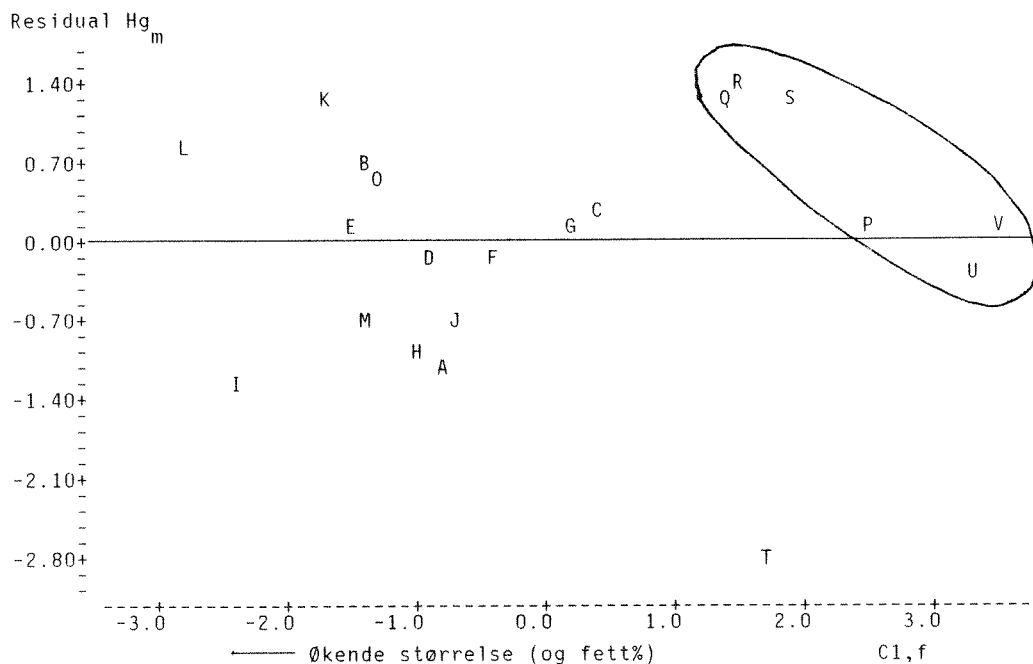
Vi kan nå se på de to hoved-komponentene for miljøgifter i sammenheng med hoved-komponentene for de fysiologiske variable.

En multippel regresjon mellom de to hovedretningene for miljøgifter C1,m og C2,m og hovedkomponentene C1,f og C2,f for de fysiologiske variable fra forrige avsnitt viser følgende:

C1,m er signifikant korrelert med C2,f og C3,f. Hovedtendensen i variasjonen i miljøgift-konsentrasjoner er altså først og fremst avhengig av fett-%, ikke så mye av variasjoner i størrelse. Regresjonen peker ut observasjon 20(T) som avvikende.

C2,m er signifikant korrelert med C1,f. Sekundært er altså miljøgift-konsentrasjonene avhengig av størrelse. Også her pekes observasjon 20(T) ut som avvikende.

Trinnvis multippelregresjon av enkelt-variable mot de fysiologiske hovedkomponentene viser at observasjon 19 pekes ut som avvikende for tre variable: HCB(lever), DCB(lever) og Hg(muskel). Det klareste tilfelle er Hg(muskel), hvor vi får residualer som vist i figur 6 etter å ha korrigert for signifikant regresjon med C2,f og C3,f. Vi ser at vi også her får frem den samme grupperingen som tidligere.



Figur 6. Residualer av $\log(\text{Hg}(\text{muskel}))$ etter regresjon mot C2,f og C3,f, fremstilt som funksjon av C1,f (størrelse).

4.1.5 Gruppering av materialet - en avsluttende spekulasjon

På basis av resultatene foran velger vi å se bort fra prøve 20(T), idet data for denne prøven ser ut til å ha et mønster som avviker fra alle de andre prøvene. I tillegg er prøve 14(N) utelatt som beskrevet tidligere.

Som et rent spekulativt eksperiment antar vi også at den tendens til gruppering vi har sett er reell, og deler opp materialet i to grupper:

Gruppe 1: Prøve 1-13 samt 15 - nr. 14 er kuttet ut

Gruppe 2: Prøve 16-19 samt 21-22 - nr. 20 er kuttet ut.

Gruppene skilles fra hverandre ved gruppe-nummer 1 eller 2, som legges inn som en tilleggsvariabel i datasettet.

Tabell 7 er et forsøk på å vise hvilken virkning en slik gruppering kan ha m.h.t. hvor godt vi kan beskrive og korrigere variasjonene i datasettet. Regresjonsberegningene er utført ved en baklengs eliminasjons-prosedyre, dvs. at en starter med å ha alle variable med i regresjonen, og så skrittvis eliminerer de variable som ikke gir signifikant bidrag til å forklare den avhengige variable ($F < 4.0$).

Tabell 7. Sammenligning av resultat fra trinnvis regresjon ved baklengs eliminasjon uten (a) og med (b) gruppevariabel.

- a: Bare de 4 fysiologiske PCA komponentene.
 b: De 4 fysiologiske PCA komponentene + den spekulative gruppe-variabelen som er innført ovenfor.

Variabel:	Alt:	Koeffisienter for forklaringsvariable:					Residualt st.avvik	R ² (justert)	
		-C1,f "størr."	C2,f "fett-%"	C3,f "Vekt .. Lengde"	C4,f "?"	Gruppe		a	b
Hg-L	a:			-0.326			0.328	15.5%	
	b:	+0.230	-0.224	-0.249		+1.16	0.200	68.7%	
HCB-L	a:	-0.036	-0.284	-0.245	-0.851		0.129	75.0%	
	b:		-0.299	-0.233	-0.827	+0.189	0.115	81.4%	
OCS-L	a:	-0.114	-0.313				0.325	39.6%	
	b:	+0.157	-0.408			+1.20	0.181	81.3%	
DCB-L	a:		-0.322				0.249	39.4%	
	b:	ingen signifikant endring							39.4%
Hg-M	a:		-0.148	-0.274	-1.07		0.113	71.8%	
	b:		-0.157	-0.267	-1.06	+0.111	0.101	77.5%	
HCB-M	a:	+0.089	-0.235	-0.459			0.1728	75.6%	
	b:	ingen signifikant endring							75.6%
OCS-M	a:		-0.210	-0.362			0.2535	38.8%	
	b:	+0.180	-0.268	-0.313	-1.07	+0.732	0.1769	70.2%	
DCB-M	a:	-0.092	-0.236	-0.219			0.1967	58.4%	
	b:		-0.206	-0.244		-0.369	0.1933	59.9%	

Siden vi nå har sett bort fra prøve nr. 14 og 20, vil PCA komponentvekter være litt forskjellige fra verdiene i tabell 5, men hovedbildet er det samme.

Vi ser at grupperingen bidrar vesentlig til å forklare variasjonen i miljøgiftvariablene Hg(lever), OCS(lever) og OCS(muskel). For de andre variablene har det mindre virkning.

Resultatet er altså ikke entydig, men det tyder iallfall på at det kan være verdt å se nærmere på om en slik gruppering kan være reell, og hva som i tilfelle ligger bak.

Tabell 1 viser at de individene som er plassert i gruppe 1 har lengder 50-71cm, mens individene i gruppe 2 har lengde 34-47cm.

Ifølge Tesch (1977) er det først og fremst lengde som er korrelert med alder for ål, mens vekten kan variere med en faktor 2 for gitt lengde. Figur 1 viser til sammenligning at $\log_{10}(\text{vekt})$ for gitt lengde varierer med 0.2, dvs. at vekten varierer med en faktor 1.6, når individ 14(N) unntas.

Det kan derfor godt være en form for aldersgruppering vi har sett antydninger til.

På basis av data i Tesch 1977, kapittel 3.3.2, ser det ut til at gjennomsnittslengden for ål øker med ca. 5-10% pr. år, og med en spredning på ± 3 til 7 cm innenfor en årsklasse. Isåfall skulle hver av gruppene 1 og 2 ovenfor omfatte flere årsklasser.

Resultatene ovenfor tyder på at det for de tre nevnte miljøgiftkomponentene, Hg(lever), OCS(lever) og OCS(muskel), er en kvalitativ forskjell mellom de laveste årsklassene i gruppe 2, og de høyere årsklassene i gruppe 1.

Det kunne være fristende å sette dette i sammenheng med en opplysning i Tesch (1977) om ålens livsløp: Ved en alder av 8-12 år for hunner, 6-9 år for hanner, går ålen inn i en vandringsfase, hvor den går ut i havet igjen for å gyte ("spawning migration stage"). Lengden er da 50-60cm for hunner, mindre for hanner. Ved denne overgangen skjer det en rekke fysiologiske forandringer, bl.a. vil tarmen tilbakedannes (Opplysninger fra John Arthur Berge, NIVA).

Gruppe 2 kunne da være ål som ikke er kommet inn i vandringsfasen, mens gruppe 1 kunne være ål som har nådd eller er i ferd med å nå denne fasen. Det ville være rimelig om et ulikt mønster for næringsopptak og metabolisme i de to fasene innvirket på konsentrasjonene av miljøgifter.

Det som taler mot denne spekulasjonen er kanskje at ålen i fase 2 vil forsvinne fra området, slik at den ikke fanges? Det må isåfall være fisk i et overgangs-stadium til fase 2 som markerer seg i gruppe 1.

En annen mulig forklaring på den indikerte grupperingen kan ligge i at hanner og hunner har forskjellige vekstkurver. Hanner er opptil 10 cm kortere enn hunner av samme alder, men har til gjengjeld større vekt i forhold til lengden.

Det er også slik at individer med bredt hode vokser raskere, dvs. får større lengde, enn individer med smalt hode, innenfor samme populasjon (Tesch, 1977). Det kan også finnes forskjeller i vekt/lengde-forhold på grunn av forskjeller i hodeform, fordi vekten ikke øker like raskt relativt sett.

Begge disse faktorer burde kanskje vært tatt i betraktning når en skulle vurdere en inndeling i grupper.

Det kan også tenkes at den tilsynelatende oppdelingen i to grupper som har ulikt forhold mellom miljøgiftkonsentrasjoner og fysiologiske forhold skyldes at det finnes ikke-lineære sammenhenger, dvs. at den log-lineære modellen beskrevet i avsnitt 2.3 er for enkel. Det er for få individer i prøvematerialet til å undersøke dette nærmere.

5 Konklusjoner

Resultatet av analysen viser at det er mulig at målte konsentrasjoner kan korrigeres vesentlig bedre enn nå, dersom en kjente alder, kjønn og hodebredde for hvert individ i materialet. Dette kunne gjøre det lettere å få frem miljøforskjeller.

Det anbefales derfor at måleprogrammet utvides til å inkludere disse tre variablene. Bestemmelse av alder for ål er noe problematisk (Tesch 1977), men det ser iallfall ut til at den kan bestemmes innenfor ± 1 år med en sikkerhet på 80% (Vøllestad 1985). Kanskje burde det også vurderes å inkludere mer direkte fysiologiske indikasjoner på hvilket stadium i livssyklusen hvert individ befinner seg.

Det presiseres at dette bare er tentative konklusjoner, basert på et nokså spinkelt materiale. For å få frem eventuelle grupperinger sikrere bør materialet for et år omfatte flere individer. Det bør iallfall være det dobbelt så stort som her, det vil si at det bør undersøkes 40-50 individer hvert år.

LITTERATUR

- Afizi, A.A og Azen, S.P. (1979): Statistical Analysis -
A computer oriented approach, Second Ed., Academic Press.
- Cochran, W.G. (1977): Sampling Techniques, 3.edition.
John Wiley & Sons.
- Durbin, J. og Watson, G.S. (1971): Testing for serial correlation in
least squares regression III. Biometrika, 58, 1-19.
- Esbensen, K. og Kalvenes, S. (1988): Statistikk-seminarer 1984.
NIVA-rapport O-83061, lnr. 2088, kapittel 3.
Bidrag fra Kim Esbensen og Sigmund Kalvenes,
Norsk Regnesentral.
- MINITAB (1986): Minitab Reference manual, Release 5, December 1986.
Minitab Inc., 3081 Enterprise Drive, State College, PA 16801,
U.S.A.
- Ryan, T.A., Jr. og Joiner, B.L.: Normal probability plots and tests
for normality. Technical Report, Statistics Department,
Pennsylvania State University.
- Rygg, B., Bjerkeng, B., Molvær., J. (1985): Grenlandsfjorden og
Skienselva 1984. Overvåkningsrapport 202/85, SFT/NIVA,
ISBN 82-577-0975-1.
- Shapiro, S.S. and Wilk, M.B.: An analysis of variance test for
normality (complete samples), Biometrika, 52, p.591.
- Tesch, F.-W., 1977: THE EEL.
Biology and Management of Anguillid Eels.
Chapmann and Hall Ltd.
- Vøllestad, L. A., 1985: Age determination and growth of yellow eels,
Anguilla anguilla (L.), from a brackish water, Norway.
J. Fish Biol. (1985) vol. 26, p.521-525.
- Wold, S., Esbensen, K., Geladi, P., 1987:
Principal Component Analysis. Chemometrics and Intelligent
Laboratory Systems, 2 (1987) 37-52.