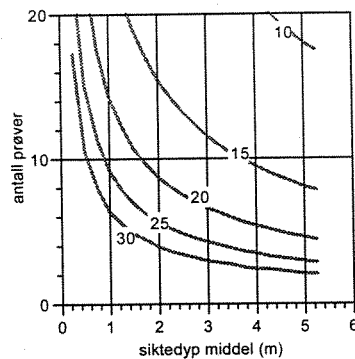
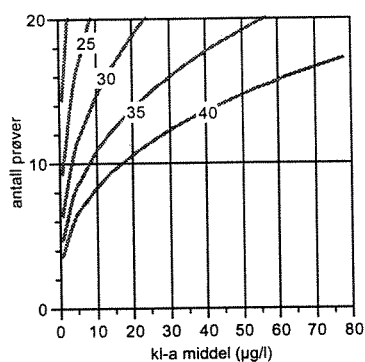
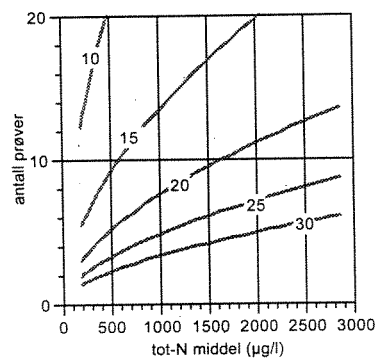
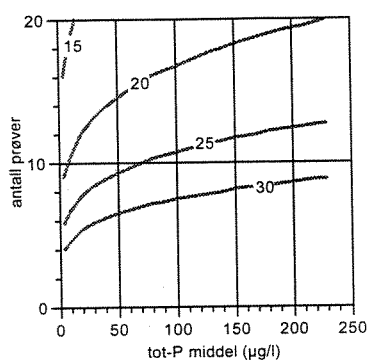


Landsomfattende trofiundersøkelse av norske innsjøer

Statistisk analyse av usikkerhet i sesongmiddelverdier



Konturplott av relativ nøyaktighet som funksjon av antall prøver og innsjøenes midlere totP, Tot-N, kl-a og siktedyb.

Hovedkontor

Postboks 173, Kjelsås
0411 Oslo
Telefon (47) 22 18 51 00
Telefax (47) 22 18 52 00

Sørlandsavdelingen

Televeien 1
4890 Grimstad
Telefon (47) 37 04 30 33
Telefax (47) 37 04 45 13

Østlandsavdelingen

Rute 866
2312 Ottestad
Telefon (47) 62 57 64 00
Telefax (47) 62 57 66 53

Vestlandsavdelingen

Thormøhlensgt 55
5008 Bergen
Telefon (47) 55 32 56 40
Telefax (47) 55 32 88 33

Akvaplan-NIVA A/S

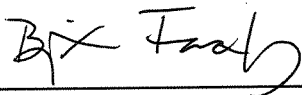
Søndre Tollbugate 3
9000 Tromsø
Telefon (47) 77 68 52 80
Telefax (47) 77 68 05 09

Tittel Landsomfattende trofiundersøkelse av norske innsjøer Statistisk analyse av usikkerhet i sesongmiddelverdier	Løpenr. (for bestilling) 3427-96	Dato
	Prosjektnr. Undernr. 91050	Sider Pris 21
Forfatter(e) Bjørn Faafeng Eirik Fjeld	Fagområde VASSDRAG	Distribusjon FRI
	Geografisk område NORGE	Trykket NIVA

Oppdragsgiver(e) Statens Forurensningstilsyn	Oppdragsreferanse
---	-------------------

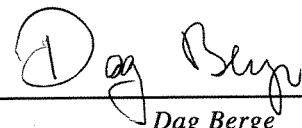
<p>Sammendrag</p> <p>Hensikten med denne rapporten har vært å anslå hvor mange prøver pr. sesong som er nødvendig for å beregne middelveidien av de 4 mest aktuelle parametrene for klassifisering av eutrofiering i innsjøer: total fosfor, klorofyll-a, siktedyp og total nitrogen, med en ønsket nøyaktighet (i % av middelveidien). Analysen omfatter også en vurdering av hvordan denne nøyaktigheten varierer over aktuelle konsentrasjonsnivåer. Resultatene viser at usikkerheten er størst for middelveidien av klorofyll, og avtar suksessivt for fosfor, siktedyp og nitrogen. Det er utarbeidet tabeller og figurer der antall nødvendige prøver pr. sesong kan leses ut.</p>

<p>Fire norske emneord</p> <ol style="list-style-type: none"> Eutrofiering Innsjøovervåking Statistisk analyse Prøveinnsamling 	<p>Fire engelske emneord</p> <ol style="list-style-type: none"> Eutrophication Lake monitoring Statistical analysis Sampling strategy
--	---



Bjørn Faafeng
Prosjektleder

ISBN 82-577-2961-2



Dag Berge
Forskningsjef

Norsk institutt for vannforskning

O-91050

Landsomfattende trofiundersøkelse av norske innsjøer

Statistisk analyse av
usikkerhet i sesongmiddelverdier

dato: 24.1.96

prosjektleder: Bjørn Faafeng

medarbeider: Eirik Fjeld

for administrasjonen: Dag Berge

FORORD

Landsomfattende trofiundersøkelse av norske innsjøer har pågått siden 1988, med innsamling av data i ialt 5 år. Undersøkelsen er finansiert av SFT. Vannprøvene er samlet inn og analysert på en enhetlig og kvalifisert måte både mhp. kjemiske og biologiske parametre, og resultatene egner seg derfor godt til videre statistiske analyser av generelle forhold i norske innsjøer. I "Revidert program for 1995" datert 13.2.95, ble det planlagt å bearbeide og rapportere visse data fra undersøkelsen i løpet av 1995.

Den foreliggende rapporten fokuserer på en faktor som ofte blir ignorert ved undersøkelser av miljøforhold, nemlig den usikkerheten i de beregnede sesongmiddelverdiene som er forårsaket av variasjoner av stoffkonsentrasjon, siktedyp ol. over tid. Denne sesongmiddelverdien er viktig fordi den er grunnlaget for klassifisering av innsjøenes vannkvalitet (SFT 1992). Det er derfor av stor interesse å få et bedre innblikk i hvor mange prøver som er nødvendig pr. sesong for å oppnå en ønsket eller akseptabel (u)sikkerhet. I denne forbindelse blir annen usikkerhet i resultatene, som skyldes f.eks. usikkerhet i den kjemiske analysen, forurensning av prøveflasker, valg av prøvetakingsdyp og vannprøvetaker, ikke vurdert. Analysen omfatter sentrale parametre for klassifisering av eutrofi i innsjøer.

Det ble utarbeidet et notat i 1987 (Faafeng og medarb. 1987) med en foreløpig statistisk analyse av prøvetakingsfrekvens i innsjøer basert på data som NIVA hadde samlet inn ved tidligere undersøkelser.

En stor takk rettes til Peter Kristensen ved Miljø- og energiministeriet i Roskilde og Erik Jeppesen ved Danmarks Miljøundersøgelser i Silkeborg for at vi fikk anledning til å bruke data fra dansk overvåking av innsjøer i denne statistiske analysen.

Det er vårt håp at rapporten kan stimulere til en diskusjon av de omtalte forholdene og en mer bevisst holdning både fra fagfolk og forvaltning ved planlegging og gjennomføring av vassdragsundersøkelser.

Rapporten er utarbeidet av Eirik Fjeld og Bjørn Faafeng. Førstnevnte har gjennomført de statistiske beregningene.

INNHold

	side
FORORD	1
INNHold	2
1. KONKLUSJONER	3
2. INNLEDNING	4
2.1 Generelt	4
2.2 Mål for undersøkelsen	5
2.3 Datagrunnlag	6
3. STATISTISKE METODER	7
3.1 Utvikling av en modell	7
3.2 Validering av modellen	8
4. RESULTATER	10
4.1 Sammenhengen mellom varians og middelværdien	10
4.2 Validering	16
4.2.1 Kovariansanalyser	16
4.2.2 Bootstrapping	17
REFERANSER	21

1. KONKLUSJONER

SFTs system for klassifisering av vannkvalitet i ferskvann (SFT 1992) bygger på beregnede middelveier for produksjonssesongen. Gjennomsnittsverdier vil falle innenfor en av 5 intervaller som avgjør tilstandsklassen for den aktuelle parameteren (klasse I - V). I denne rapporten har vi foretatt en statistisk analyse av variansen av middelveien av de 4 mest aktuelle parametrene for klassifisering av eutrofiering i innsjøer: total fosfor, klorofyll-a, siktedyp og total nitrogen.

Hensikten har vært å anslå hvor mange prøver pr. sesong som er nødvendig for å beregne middelveien med en ønsket nøyaktighet (i % av middelveien). Analysen omfatter også en vurdering av hvordan denne nøyaktigheten varierer over aktuelle konsentrasjonsnivåer.

Analysen viser at variansen av middelveien øker systematisk med økende middelveier for de fire parametrene. Den matematiske sammenhengen som beskriver dette forholdet synes å ha generell gyldighet, siden vi finner det samme forholdet i et dansk datamateriale. Dette gir grunnlag for å beregne det nødvendige antallet prøver for å oppnå ønsket sikkerhet i sesongmiddelveien.

Som eksempler på de beregnede resultatene viser vi under hvor stor sikkerhet en kan vente å få i sesongmiddelveiene ved forskjellig antall prøver pr. sesong:

Usikkerhet i sesongmiddelveien for en innsjø i tilstandsklasse II:

ant. prøver -->	4	6	8	10	12	15
totalP	30	25	22	20	18	15
kl-a	40	35	32	27	25	22
siktedyp	25	20	17	15	14	13
totalN	20	17	15	13	12	10

Usikkerhet i sesongmiddelveien for en innsjø i tilstandsklasse IV:

ant. prøver -->	4	6	8	10	12	15
totalP	35	30	27	23	22	20
kl-a	>50	50	45	40	37	33
siktedyp	35	27	23	20	18	16
totalN	25	20	18	16	15	13

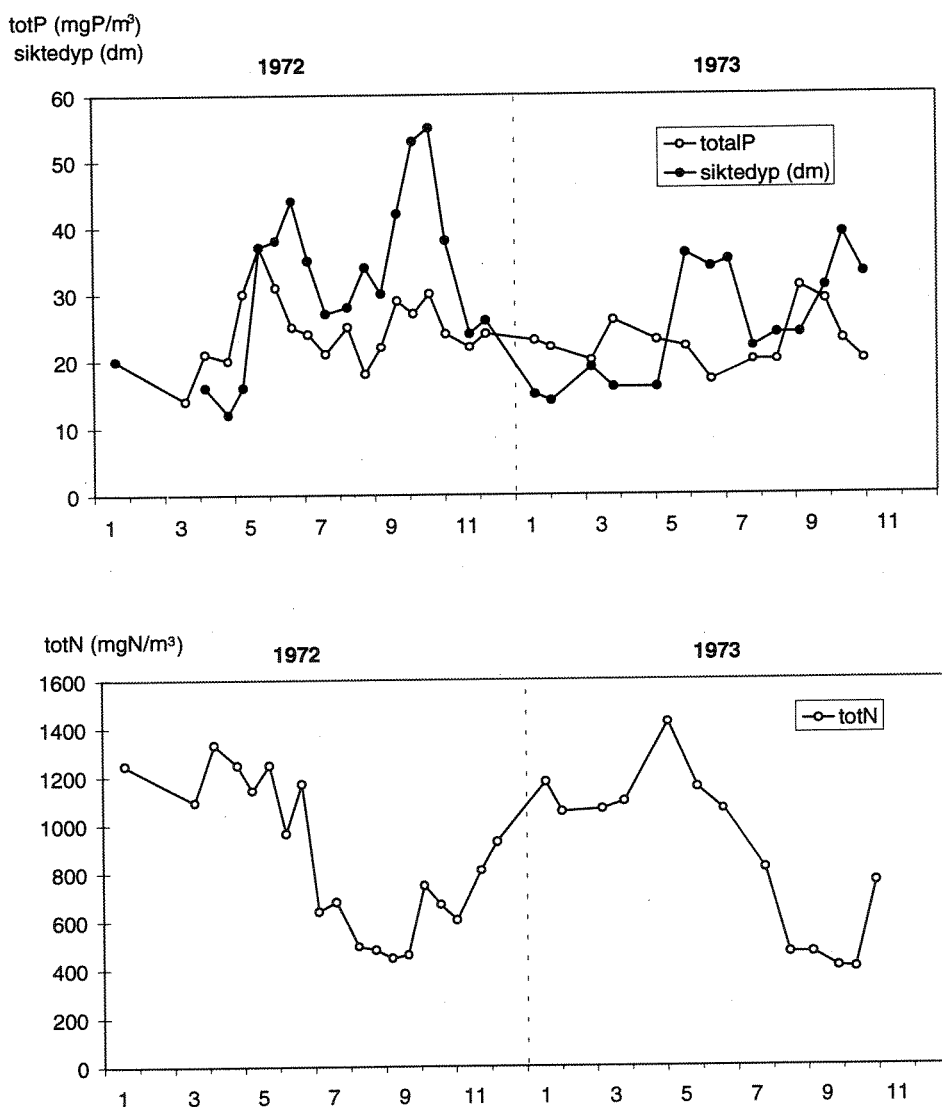
Usikkerheten er størst for middelveien av klorofyll, og avtar suksessivt for fosfor, siktedyp og nitrogen. Beregningene viser f.eks. at en innsjø i tilstandsklasse II vil ved 6 prøver pr. sesong få følgende usikkerhet i middelveiene for hhv. totalP, klorofyll-a, siktedyp og totalN: 25%, 35%, 20% og 17%. Tilsvarende for en innsjø i klasse IV vil være: 30%, 50%, 27% og 20%.

De beregnede verdiene som vist i denne rapporten gjelder for et gjennomsnitt av et stort antall innsjøer. Variansen i en tilfeldig valgt ny innsjø kan være noe større eller noe mindre enn gjennomsnittet. Dersom det er kritisk å oppnå en ønsket relativ nøyaktighet bør en derfor vurdere å øke antallet prøver noe.

2. INNLEDNING

2.1 Generelt

"Landsomfattende trofundersøkelse av norske innsjøer" er gjennomført med rutinemessig innsamling av vannprøver 4 ganger pr. sesong fra et utvalg innsjøer. Antall prøver pr. sesong ble valgt som et kompromiss mellom det faglig ønskelige og det økonomisk mulige. Typisk nok var det faglige ønsket fra NIVA om "flere" prøver pr. år ikke begrunnet i en statistisk analyse av hvor mange prøver som er nødvendig for å oppnå en ønsket sikkerhet i de beregnede middelverdiene. Tilsvarende lå det vel heller ikke en bevisst faglig vurdering til grunn for at oppdragsgiveren ville "nøye seg med" 4 prøver pr. sesong. Overvåking av innsjøer for å karakterisere vannets kvalitet gjennomføres oftest med 4 - 10 prøver pr. sesong alt etter tilgang på ressurser, kostnadsnivå og forhandlingssuksess. Ved andre typer undersøkelser med andre mål vil det selvsagt kunne være behov for flere eller færre prøver.



Figur 2.1 Total fosfor og siktedyb (A) og total nitrogen (B) i Gjersjøen i sjiktet 0-10 meters dyp gjennom årene 1972 og '73. Data fra NIVAs overvåkingsprogram for Oppegård kommune.

Over er vist et eksempel på analyseresultater fra Gjersjøen i Akershus gjennom 1972 og '73 (Figur 2.1), med hhv. 12 og 9 prøver i produksjonssesongen. Verdier utenom produksjonssesongen er også tatt med for å illustrere at forholdene normalt er mer stabile i denne perioden. Verdiene fra Gjersjøen viser ifølge figuren store variasjoner, og det er påfallende at nitrogen i denne innsjøen har et ganske regelmessig, sinusformet forløp, mens fosfor og siktedyp varierer mer tilfeldig. I tillegg til innsjøenes trofinivå vil faktorer som vannets oppholdstid, innsjøens dybdeforhold og type forurensningskilder og avstand til innsjøen være av stor betydning. Det foreliggende datamaterialet gir ikke grunnlag for diskusjon av de sistnevnte faktorene.

Det vil sjelden eller aldri være mulig å måle seg fram til en "sann" middelvei gjennom en sesong, fordi de fleste parametrene varierer mer eller mindre systematisk med årstid og døgn; mens nedbør, forurensningsepisoder ol. også vil ha stor innflytelse. Visse parametre holder omtrent samme konsentrasjonsnivå gjennom sesongen, f.eks. enkelte oppløste, lite reaktive ioner, mens parametre som inngår i biologiske prosesser, som f.eks. klorofyll og nitrogen kan variere mye. Antallet prøver må derfor være tilstrekkelig høyt til at det representerer den aktuelle variasjonen og kan gi en tilfredsstillende middelvei. Et viktig forhold er at antallet bør tilpasses de viktigste parametrene for formålet med undersøkelsen. Overvåkingsundersøkelser gjennomføres sjelden slik at en kan få et detaljert bilde av variasjonen f.eks. i algesammensetning gjennom sesongen, da mange biologiske prosesser er svært raske og kan gi store endringer over få dager.

Det blir sjelden gjort en seriøs vurdering av sikkerheten i tilstands-klassifiseringen pga. variasjonen i datagrunnlaget. Dette er av stor betydning f.eks. ved vurdering av trender over tid.

2.2 Mål for denne undersøkelsen

Hensikten med denne rapporten er å analysere hvor mange prøver som trengs for å gi en ønsket sikkerhet i middelveien over produksjonssesongen for parametrene total fosfor, klorofyll, siktedyp og total nitrogen.

Beregningene forutsetter at prøvene tas med faste intervaller i perioden fra ca. 1. mai til ca. 1. oktober. Lengden av den reelle produksjonssesongen vil selvsagt variere fra sør mot nord og fra kyst mot høyfjell og mer kontinentalt klima.

2.4 Datagrunnlaget

Vi har benyttet oss av to databaser i denne studien: hhv. fra et norsk og et dansk landsomfattende overvåkingsprogram for eutrofiering.

De norske datene er hentet fra "Landsomfattende trofiundersøkelse av norske innsjøer" som NIVA gjennomfører for SFT. I dette programmet tas normalt 4 prøver pr. sesong i de aktuelle innsjøene. Som et første grunnlag for å kunne foreta de følgende analysene ble prøveinnsamlingen i 1992 gjennomført med 8 prøver gjennom sesongen fra hver innsjø istedenfor 4. Antall innsjøer som ble undersøkt ble tilsvarende redusert for å opprettholde samme aktivitetsnivå som foregående år. Alle vannprøvene er blandprøver fra overflaten ned til 2 ganger siktedypet (minimum 0-1 m, maksimum 0-10m) og er analysert på NIVAs laboratorium i Oslo etter gjeldene analysemetodikk.

De norske dataene som er brukt i analysen er fra disse 26 innsjøene, med omlag 12 observasjoner pr. innsjø. Prøvetakingshyppigheten var 8 eller 4 prøver pr. år, innsamlet i vegetasjonsperioden mai–september. Prøvene er tatt med samme tidsintervall innen en sesong. Majoriteten av prøvene (8 pr. innsjø) var tatt i 1993, resten var samlet inn i perioden 1988–1992. Innledende undersøkelser indikerer ikke noen betydelige forskjeller i eutrofieringsnivået eller variasjonen imellom årene i disse innsjøene, og vi ser det derfor som uproblematisk å behandle alle prøvene fra en innsjø samlet, dvs. som om de var samlet inn i samme sesong.

Danmarks Miljøundersøgelser har ansvaret for organiseringen av et overvåkingsprogram for innsjøer i Danmark (se f.eks. Jensen og medarb. 1995). Den danske databasen har vi benyttet til å validere resultatene. Da de danske innsjøene var tildels betydelig mer eutrofe enn de norske, måtte vi ekskludere de mest eutrofe for å gjøre dataene bedre sammenliknbare med de norske. Av i alt 40 innsjøer har vi benyttet data fra 22 innsjøer. I likhet med den norske undersøkelsen har vi benyttet observasjonene fra månedene mai–september. Fra hver innsjø har vi omlag 20 observasjoner fra årene 1990 og 1992, dvs. omlag 2 pr. måned. Heller ikke for dette datautvalget fant vi noen betydningsfulle forskjeller mellom årene.

Tabell 2.1. Median-, max- og min-verdier for vannkvalitetsvariablene benyttet i denne undersøkelsen. Medianen er verdien som deler prøveutvalget i to like store deler (50% persentilen).

variabel	enhet	Norge			Danmark		
		median	max	min	median	max	min
tot-P	mg/m ³	22	245	4	72	202	15
tot-N	mg/m ³	916	3398	254	1112	4157	333
kl-a	mg/m ³	11	80	2	28	93	4
siktedyp	m	2.7	5.7	0.6	1.7	5.5	0.4

3. STATISTISKE METODER

3.1 Utvikling av en modell

For å kunne besvare hvor mange prøver man bør ta, må man først definere seg et ønsket presisjonsnivå. Det vil si at man ønsker at unøyaktigheten i det estimerte gjennomsnittet ikke skal være større enn en viss prosent. Vi kan kalle dette presisjonsnivået for *relativ nøyaktighet* (d_{rel}). Eksempelvis kan vi ønske at differansen mellom konfidensgrensene og estimert gjennomsnitt ikke skal være større enn $\pm 20\%$, $\pm 10\%$ eller $\pm 5\%$ av gjennomsnittet. Med *konfidensgrenser* menes et intervall der man med en gitt sannsynlighet vet at det reelle gjennomsnittet til den statistiske populasjonen (dvs. alle mulige verdier i tidsintervallet) skal ligge. Om mange tilfeldige prøver trekkes fra den samme populasjonen og 95% konfidensintervall for hver prøve beregnes, kan man forvente at det sanne gjennomsnittet til populasjonen ligger innenfor det beregnede konfidensintervallet i 95 av 100 tilfeller.

De øvre og nedre 95% konfidensgrensene kan uttrykkes som en funksjon av middelveidien (\bar{x}), standardavviket til prøven (s), antall observasjoner den er basert på (n), og en tilhørende Students t -verdi ($p = 0.05$ og $n-1$ frihetsgrader):

likning 1:

$$CI = \bar{x} \pm t \cdot \left(\frac{s}{\sqrt{n}} \right)$$

En forutsetning for likning 1 er at middelveidien (\bar{x}) skal være normalfordelt. Det vil si at dersom man gjentatte ganger tar tilfeldige prøver med n observasjoner fra populasjonen, så skal frekvensfordelingen av \bar{x} være normalfordelt. Uavhengig om den underliggende fordelingen vil middelveidene av prøvene bli normalfordelt – gitt at prøvestørrelsen (n) er tilstrekkelig stor. Dette resultatet er fundamentalt innefor statistisk teori, og kalles *sentralgrenseteoremet*. Dersom prøvestørrelsen n er liten, og den underliggende populasjonen det tas prøver av avviker fra normalfordelingen, vil \bar{x} ikke nødvendigvis bli normalfordelt. Konfidensintervallene beregnet ut fra likning 1 vil da kunne bli gale. Vi kommer tilbake til dette under valideringen av modellene.

Ut fra likning 1 ser vi at relativ nøyaktighet kan uttrykkes som:

likning 2:

$$d_{rel} = t \cdot \frac{\left(\frac{s}{\sqrt{n}} \right)}{\bar{x}}$$

Vi har her sløffet \pm tegnet, og lar relativ nøyaktighet være oppgitt som absoluttverdien. Ut fra likning 2 kan vi utlede et uttrykk for antall prøver som må tas (n) for å oppnå en ønsket relativ nøyaktighet:

likning 3:

$$n = \frac{t^2 \cdot s^2}{(d_{rel} \cdot \bar{x})^2}$$

Kvadratet til standardavviket (s^2) kalles *varians*. Dersom variansen er uavhengig av gjennomsnittet er det forholdvis uproblematisk å beregne antallet nødvendige prøver (n) ut

fra en forventet middelværdi og varians. Erfaringsmessig øker imidlertid ofte variansen til vannkjemiske variabler med gjennomsnittet. Dette kan man illustrere ved å plote enkeltobservasjonene mot middelværdien for de respektive innsjøene. Vi har gjort dette for de aktuelle variablene i den norske regionale undersøkelsen, og fant at spredningen økte tilnærmet eksponentielt med middelværdien (figur 3.1). På et log-log plot vil enkeltobservasjonene da ligge som et jevnt bånd omkring middelværdiene. Vi kan derfor uttrykke variansen som en lineær funksjon av middelværdiene, når begge er log-transformert. Disse lineære forholdene er grunnleggende for modellene våre, da de kan benyttes til å predikere variansen ut fra en forventet middelværdi :

likning 4:

$$\log(s^2) = \log a + b \cdot \log(\bar{x}) ,$$

$$\text{eller tilbaketransformert som } s^2 = a \bar{x}^b$$

der \bar{x} er middelværdien av tot-N, tot-P, kl-a eller siktedypet i innsjøen, og $\log(a)$ og b er konstanter (henholdsvis skjæringspunktet med y-aksen og stigningskoeffisienten).

Ved å sette likning 4 inn i likning 3 får vi et uttrykk som predikerer nødvendig prøvestørrelse ut fra middelværdi, relativ nøyaktighet og en t-verdi:

likning 5:

$$n = \frac{t^2 \cdot (a\bar{x}^b)}{(d_{rel} \cdot \bar{x})^2}$$

Ved å manipulere med likning 5 kan man så framstille konturplott av de ønskede variablene som funksjon av antall prøver pr. sesong.

Modellen forutsetter at det ikke er noen autokorrelasjon mellom observasjonene innen hver lokalitet. Med autokorrelasjon forstås at hver observasjon tenderer til å være ganske lik den foregående observasjonen. Arbeider man med vannkvalitetsdata fra innsjøer må nødvendigvis autokorrelasjon inntreffe ved en høy prøvetakingsfrekvens. Antallet prøver pr. sesong som modellen vår baserer seg på er såvidt lavt at autokorrelasjon ikke vil være noe stort problem (majoriteten av prøvene utgjøres av 8 prøvetakninger i løpet av en sesong). Øker man prøveantallet utover dette, vil autokorrelasjonen imidlertid kunne gjøre at modellene underestimerer nøyaktigheten.

3.2 Validering av modellen

For å validere modellen har vi brukt to metoder: kovariansanalyser og bootstrapping-metoder.

Ved *kovariansanalyser* testet vi om forholdet mellom varians og middelværdi (regresjonslinjene) var ulike i det norske og det danske datasettet. En kovariansanalyse gir svar på om stigningskoeffisientene eller konstantleddene i de to regresjonene er forskjellige. Dette blir å betrakte som test på generaliteten til regresjonsmodellen.

I tillegg sammenliknet vi modellprediksjonene fra det danske datasettet med resultatene fra bootstrapping-metoder. *Bootstrapping* er datamaskin-intensive metoder for å knytte mål på nøyaktighet til statistiske parametere. En innføring til metoden finner man i Diaconis og Efron (1983) og Efron og Tibshirani (1993). Metoden er i sin enkleste form fordelingsfri, dvs. at den stiller ingen krav om at parametrene skal følge en underliggende teoretisk

fordeling. Den grunnleggende ideen bak bootstrapping-metoden er enkel, og går ut på at en genererer store mengder med "syntetiske" datasett basert på tilfeldig trekking med tilbakelegging fra et originalt datasett. Ut fra disse nye og mange datasettene kan man så beregne de aktuelle statistiske parametrene. Konfidensintervaller kan for eksempel baseres på bootstrappede prosentiler. En anskueliggjøring av framgangsmåten er som følger:

1. Trekk en tilfeldig observasjon fra et datasett, notér verdien og legg observasjonen tilbake. Gjør dette like mange ganger som det er observasjoner i datasettet (n). Dette kalles tilfeldig trekking med tilbakelegging, og hver observasjon har da like stor sannsynlighet for å bli trukket hver gang. Beregn den statistiske parameteren av interesse, f. eks. middelverdien.
2. Gjenta prosedyren (1) et stort antall ganger, gjerne 1000 ganger. Vi sitter da med en lang tallrekke (vektor) med middelverdier som hver er beregnet på m tilfeldige trekninger fra et datasett med n observasjoner.
3. Lag en frekvensfordeling eller histogram av vektoren med de bootstrappede middelverdiene, og beregn prosentilene av interesse. Et 95% bootstrapped konfidensintervall vil f.eks. være lik intervallet mellom 2,5 og 97,5 prosentilene.

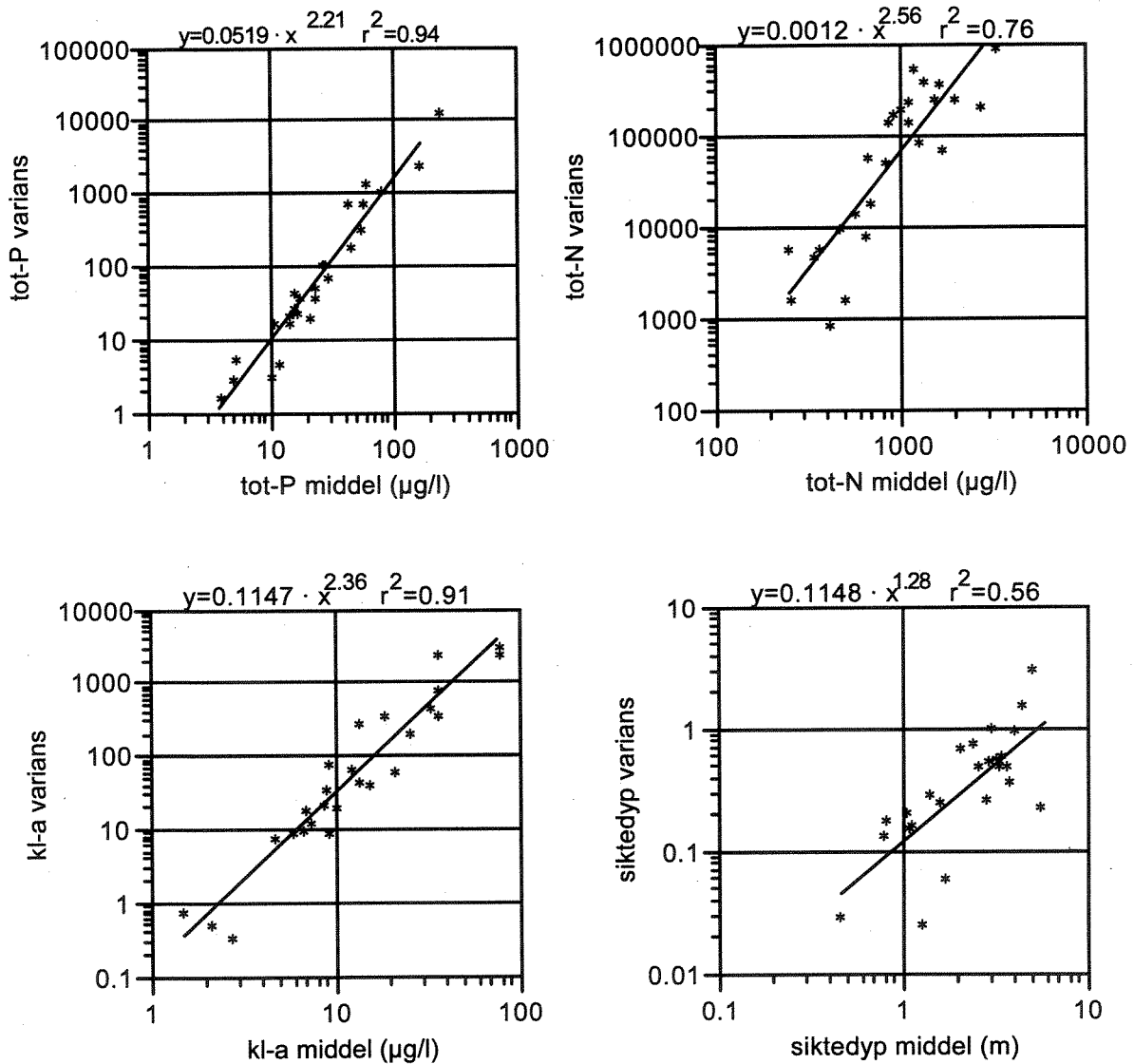
Hele denne prosessen kan gjøres for varierende prøvestørrelser, og en får da et innblikk i hvilken effekt prøvestørrelsen har for bredden av konfidensintervallet. I datasettet fra den danske undersøkelsen var det hensiktsmessig å gjøre dette for bootstrappede prøvestørrelser (n) lik 2, 3...20. For hver vannkvalitetsvariabel får en således en matrise av bootstrappede gjennomsnitt basert på prøvestørrelser fra 2 til 20 (19 kolonner med 1000 rekker). Bootstrappede gjennomsnitt og konfidensintervaller kan så beregnes for hver prøvestørrelse. Bootstrappings-prosedyrene er skrevet i MATLAB for Macintosh v. 4.2 (Mathworks Inc.)

De bootstrappede og konvensjonelle konfidensintervallene vil være omtrent identiske når n er stor, men vil kunne divergere når prøvestørrelsen er liten og utgangspopulasjonen avviker fra normalfordelingen. Ved små prøvestørrelser vil de bootstrappede konfidensintervallene kunne avvike fra en normalfordeling, og de vil være mer forventningsrette (unbiased) enn de konvensjonelle konfidensintervallene. Ved å sammenlikne konfidensintervallene fra modellene (likning 1, men hvor variansen er estimert ut fra likning 4) med konfidensintervallene fra bootstrappingsmetoden får man et inntrykk av hvor god modellen er til å predikere en relativ nøyaktighet ved forskjellige prøvestørrelser (n).

4. RESULTATER

4.1 Sammenhengen mellom varians og middelverdi

Generelt var det en god lineær sammenheng mellom $\log(\text{varians})$ og $\log(\text{middelverdi})$ for de fire vannkvalitetsvariablene (figur 3.1). Best var den for tot-P, kl-a, og tot-N, med en determinasjonskoeffesient (R^2) på henholdsvis 0.94, 0.91 og 0.76. Determinasjonskoeffesienten, som kan variere mellom 0 og 1, gir et mål på hvor mye av variansen i y-variabelen som kan forklares av x-variabelen. Variansen i siktedypet var imidlertid mindre avhengig av middelverdien, her var $R^2 = 0.56$. Likningene, som er gitt i tabell 3.1, kan benyttes til å beregne variansen ut fra en gitt middelverdi.



Figur 4.1. Regresjonslinjer av varians på gjennomsnitt for de fire vannkvalitetsvariablene (log-transformerte data). Regresjonene er gitt i tabell 3.1.

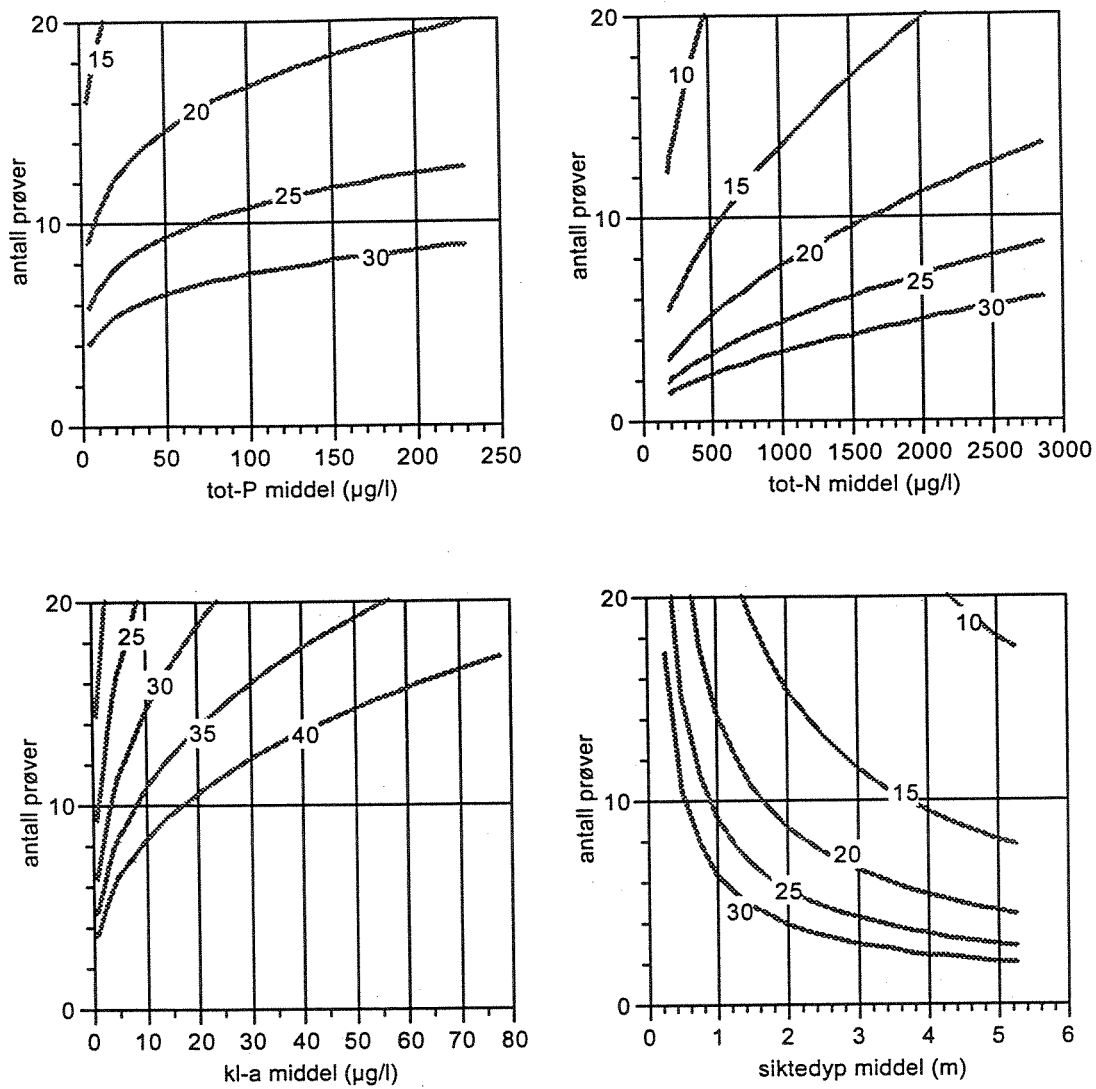
Tabell 4.1. Regresjoner av varians (s^2) på middelerverdi (\bar{x}). Regresjonene er basert på 26 observasjoner.

variabel	regresjonslikning	R ²	p
tot-N	$\log(s^2) = 2.5577 \times \log(\bar{x}) - 2.8905$	0.76	<0.001
tot-P	$\log(s^2) = 2.2072 \times \log(\bar{x}) - 1.2850$	0.94	<0.001
kl-a	$\log(s^2) = 2.3610 \times \log(\bar{x}) - 0.9404$	0.91	<0.001
siktedyp	$\log(s^2) = 1.2827 \times \log(\bar{x}) - 0.9401$	0.56	<0.001

Likningene for fosgfor og klorofyll er ganske like de som ble beregnet av Knowlton og medarb. (1984) for 188 innsjøer i USA og av France og Peters (1992) for 65 "nordlig, tempererte" innsjøer. Det viser at vi har med å gjøre med et generelt gyldig forhold; nemlig at variansen øker med økende sesongmiddelerverdi.

Setter vi uttrykket for varians (fra tabell 4.1) inn i likningen for relativ usikkerhet (likning 2) kan vi framstille relativ usikkerhet som funksjon av n (prøvestørrelse) og gjennomsnitt. I figur 4.2 har vi vist konturplott av relativ usikkerhet (d-rel) ved varierende middelerverdier og n . Vi ser her at den relative usikkerheten øker med middelerverdien av tot-P, tot-N og kl-a, mens den minsker med siktedypet. For eutrofe innsjøer må man derfor ta flere prøver igjennom en sesong for å oppnå samme relativ usikkerhet i estimert middelerverdi som for mer oligotrofe innsjøer. Eksempelvis viser konturplottene at det i en oligotrof innsjø med midlere tot-P på 5 µg/l kreves omlag 6 prøver for å oppnå en relativ usikkerhet på 25% i estimatet, mens det i en eutof innsjø med midlere tot-P på 70 mg/m³ kreves 10 prøver for å oppnå samme usikkerhet. Etter SFTs klassifisering er dette innsjøer i hhv. tilstandssklasse I og V (SFT 1992). For nitrogen indikerer konturplottet at det i en innsjø med tot-N på 200 mg/m³ (klasse I) kreves kun 2 prøver for å oppnå en relativ usikkerhet på 25%, mens det i en innsjø med tot-N på 1000 mg/m³ (klasse V) kreves omlag 5 prøver for å oppnå samme usikkerhet.

Den statistiske metoden som ble brukt i denne analysen er ikke helt lik den innledende analysen (Faafeng og medarb. 1987), ved at det den gang ble beregnet varians i enkeltmålingene for hver innsjø. De beregnede antall nødvendige prøver er likevel ganske like for næringsfattige (oligotrofe) og middels næringsrike (mesotrofe) innsjøer, mens avviket var stort for eutrofe innsjøer. Avviket skyldes i stor grad at få eutrofe innsjøer inngikk i den tidligere analysen. Begge disse analysene gir samme resultat for hvilke av parametrene som gir størst (klorofyll) og minst (nitrogen) varians.



Figur 4.2. Konturplott av relativ nøyaktighet (d_{reb}) som funksjon av antall prøver og innsjøenes midlere tot-P, tot-N, chl-a og siktedyb. Ved å gå ut fra et gitt midlere nivå i de enkelte vannkvalitetsvariablene kan man avlese hvilken relativ nøyaktighet man får på estimert middelverdi ved ulike prøvestørrelser.

Som eksempler på resultatene viser vi under hvor stor relativ nøyaktighet en kan vente å få i sesong middelverdiene ved forskjellig antall prøver (tabell 4.2):

Tabell 4.2 Relativ nøyaktighet i sesongmiddelverdien for en innsjø i hver av de 5 tilstandsklassene

tilstandsklasse I:

ant.	4	6	8	10	12	15
prøver -->						
totalP	30	25	22	19	18	16
kl-a	39	34	29	25	23	20
siktedyp	20	17	14	13	12	11
totalN	17	14	11	10	9	8

tilstandsklasse II:

ant.	4	6	8	10	12	15
prøver -->						
totalP	32	26	23	20	19	16
kl-a	40	35	32	27	25	22
siktedyp	25	20	17	15	14	13
totalN	20	17	15	13	12	10

tilstandsklasse III:

ant.	4	6	8	10	12	15
prøver -->						
totalP	33	28	25	22	21	18
kl-a	50	43	36	32	30	25
siktedyp	27	22	19	17	15	14
totalN	22	18	16	14	13	12

tilstandsklasse IV:

ant.	4	6	8	10	12	15
prøver -->						
totalP	35	30	27	23	22	20
kl-a	-	50	45	40	37	33
siktedyp	35	27	23	20	18	16
totalN	25	20	18	16	15	13

>50

tilstandsklasse V:

ant.	4	6	8	10	12	15
prøver -->						
totalP	50	40	30	28	25	23
kl-a	-	-	50	47	45	40
siktedyp	50	40	35	30	27	25
totalN	30	27	23	22	19	17

>50

Tabell 4.3 Nødvendig antall prøver for å oppnå ønsket relativ nøyaktighet (d_{rel}) i middelverdien

total fosfor:

tilstandsklasse	I	II	III	IV	V
d_{rel} (%)					
15	16	18	-	-	-
20	9	10	12	15	20
25	6	7	8	9	12
30	4	5	5	6	9

klorofyll-a: NB! merk annen skala!

tilstandsklasse	I	II	III	IV	V
d_{rel} (%)					
25	10	14	20	-	-
30	7	8	13	19	-
35	5	6	9	14	-
40	4	5	8	11	20

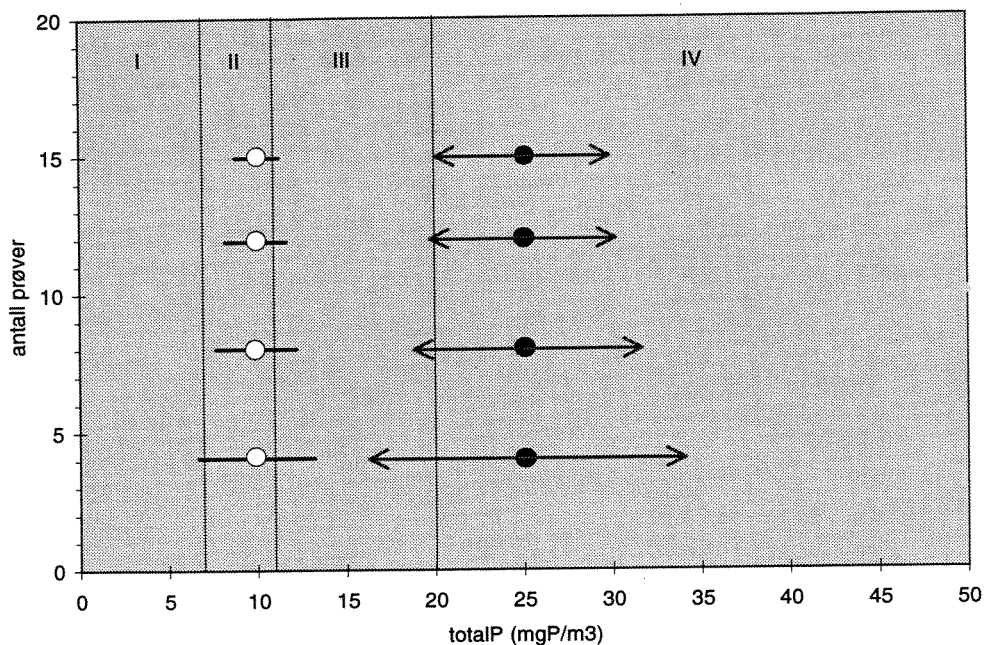
siktedyp:

tilstandsklasse	I	II	III	IV	V
d_{rel} (%)					
15	8	10	15	-	-
20	4	8	9	12	-
25	3	4	6	9	15
30	2	3	4	6	10

total nitrogen:

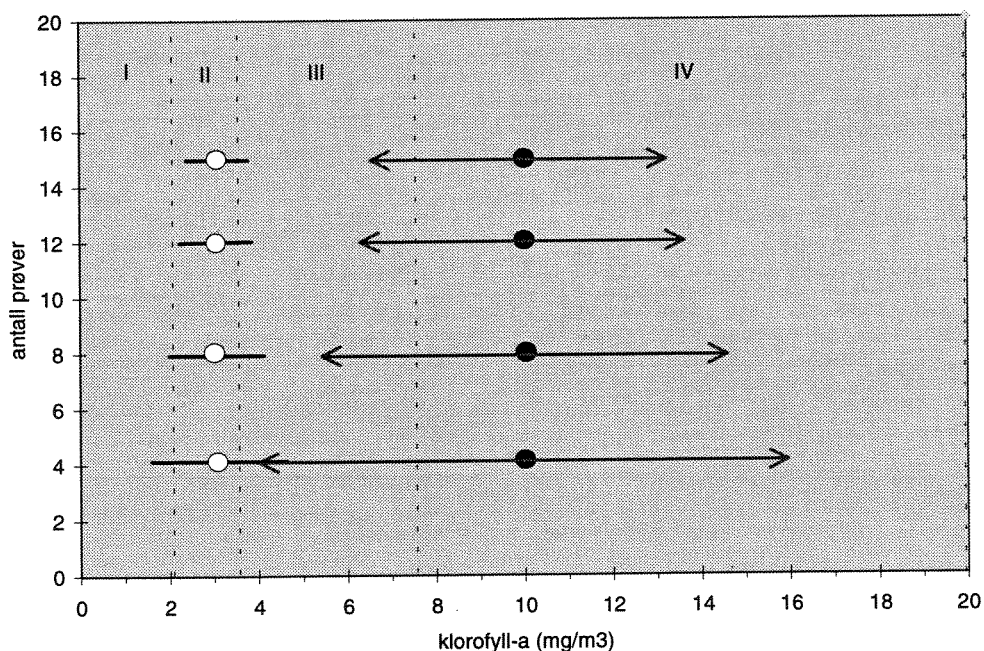
tilstandsklasse	I	II	III	IV	V
d_{rel} (%)					
15	8	9	10	13	20
20	4	5	6	8	12
25	3	3	4	5	7
30	2	2	3	4	5

Vi har også vist dette grafisk i figur 4.3. Her har vi valgt ut to tenkte innsjøer med hhv. tot-P på 10 og 25 mg/m³ og deres beregnede nøyaktighet ved 4, 8, 10 og 15 prøver pr. sesong.



Figur 4.3 Eksempel på nøyaktigheten ved klassifisering av en tenkt innsjø med sesongmiddelerdi av totalP på hhv. 10 mgP/m^3 (hvite sirkler) og 25 mgP/m^3 (mørke sirkler) og beregnet nøyaktighet (horisontale linjer) ved varierende antall prøver. Romertallene øverst angir SFTs tilstandsklasser.

Med bare 4 prøver pr. sesong kan en risikere at en innsjø med en middelerdi på 10 mgP/m^3 kan falle innenfor tre forskjellige tilstandsklasser (I / II / III) pga. den beregnete relative nøyaktigheten. En tilsvarende innsjø med middelerdi på 25 mgP/m^3 vil kunne komme i klasse III og IV ut fra fire prøver. Figuren viser at usikkerheten ble betydelig redusert ved å øke antallet prøver.



Figur 4.4 Eksempel på nøyaktigheten ved klassifisering av en tenkt innsjø med sesongmiddelerdi av klorofyll på hhv. 3 mgP/m^3 (hvite sirkler) og 10 mgP/m^3 (mørke sirkler) med beregnet nøyaktighet (horisontale linjer) ved varierende antall prøver pr. sesong. Romertallene øverst angir SFTs tilstandsklasser.

Tilsvarende som for figur 4.3 for fosfor har vi vist resultater beregnet for klorofyll for to tenkte innsjøer med middelerdi på hhv. 3 og 10 mg/m³ (figur 4.4). Med bare fire prøver pr. sesong kan en innsjø med middelerdi på 3 mg klorofyll/m³ falle innenfor både klasse I, II og III pga. den relative nøyaktigheten i estimatet, mens en tilsvarende innsjø på 10 mg klorofyll/m³ kan falle innenfor klasse III og IV. Det er verdt å merke seg at 4 prøver kan gi samme tilstandsklassen for disse to innsjøene.

Ved 15 prøver pr. sesong er usikkerheten vesentlig mindre, men hver av de tenkte innsjøene kan falle innenfor to forskjellige klasser (hhv. II / III, og III / IV).

4.2 Validering

4.2.1 Kovariansanalyser

Vi kunne ikke påvise noen signifikante forskjeller mellom stigningskoeffesientene eller konstantleddene i regresjonene av varians (s^2) på gjennomsnitt (\bar{x}) i det norske og det danske datamaterialet. Dette gjalt for samtlige vannkvalitetsvariabler, skjønt for kl-a var forskjellene nær grensområdet for statistisk signifikans. De påviste relasjonene mellom varians og gjennomsnitt i det norske datamaterialet synes derfor å være av en mer generell karakter, og bør derfor kunne anvendes på liknende innsjøtyper fra et større geografisk område. Vi har vist resultatene fra kovariansanalysene i tabell 4.1. Kovariansanalysene testet modeller av følgende type:

likning 6:

$$\log s^2 = a \cdot \log \bar{x} + \begin{cases} -b & \text{om } N \\ +b & \text{om } DK \end{cases} + \begin{cases} -c \cdot \log \bar{x} & \text{om } N \\ +c \cdot \log \bar{x} & \text{om } DK \end{cases} + k$$

Regresjonsparametrene b og c tester om de danske og norske regresjonslinjene har hhv. samme skjæringspunkt med y-aksen og samme stigningskoeffisient. Når disse parametrene ikke er signifikant forskjellige fra null har vi intet grunnlag for å hevde at regresjonene fra det norske og danske datamaterialet er ulike.

I tabell 4.4 viser p-verdien at disse forutsetningene med rimelig sikkerhet holder for alle variablene ($p < 0.05$).

Tabell 4.4 Kovariansanalyse av regresjonsberegningene.**a. tot-N. $R^2=0.74$, $n=48$, $p<0.0001$**

	Parameter	Estimate	Std Error	t	p
Intercept	<i>k</i>	-1.882094	0.699948	-2.69	0.0101
log N-mean	<i>a</i>	2.2481702	0.230679	9.75	0.0000
Nasjon[DK-N]	<i>b</i>	1.0084058	0.699948	1.44	0.1568
Interaksjoner	<i>c</i>	-0.309541	0.230679	-1.34	0.1865

b. tot-P. $R^2=0.92$, $n=48$, $p<0.0001$

	Parameter	Estimate	Std Error	t	p
Intercept	<i>k</i>	-0.973331	0.197139	-4.94	0.0000
log P-mean	<i>a</i>	2.0634621	0.114569	18.01	0.0000
Nasjon[DK-N]	<i>b</i>	0.3116707	0.197139	1.58	0.1210
Interaksjoner	<i>c</i>	-0.143700	0.114569	-1.25	0.2164

c. klorofyll-a. $R^2=0.92$, $n=48$, $p<0.0001$

	Parameter	Estimate	Std Error	t	p
Intercept	<i>k</i>	-0.674063	0.139858	-4.82	0.0000
log Kl-a mean	<i>a</i>	2.1800219	0.106744	20.42	0.0000
Nasjon[DK-N]	<i>b</i>	0.2663273	0.139858	1.90	0.0634
Interaksjoner	<i>c</i>	-0.181013	0.106744	-1.70	0.0970

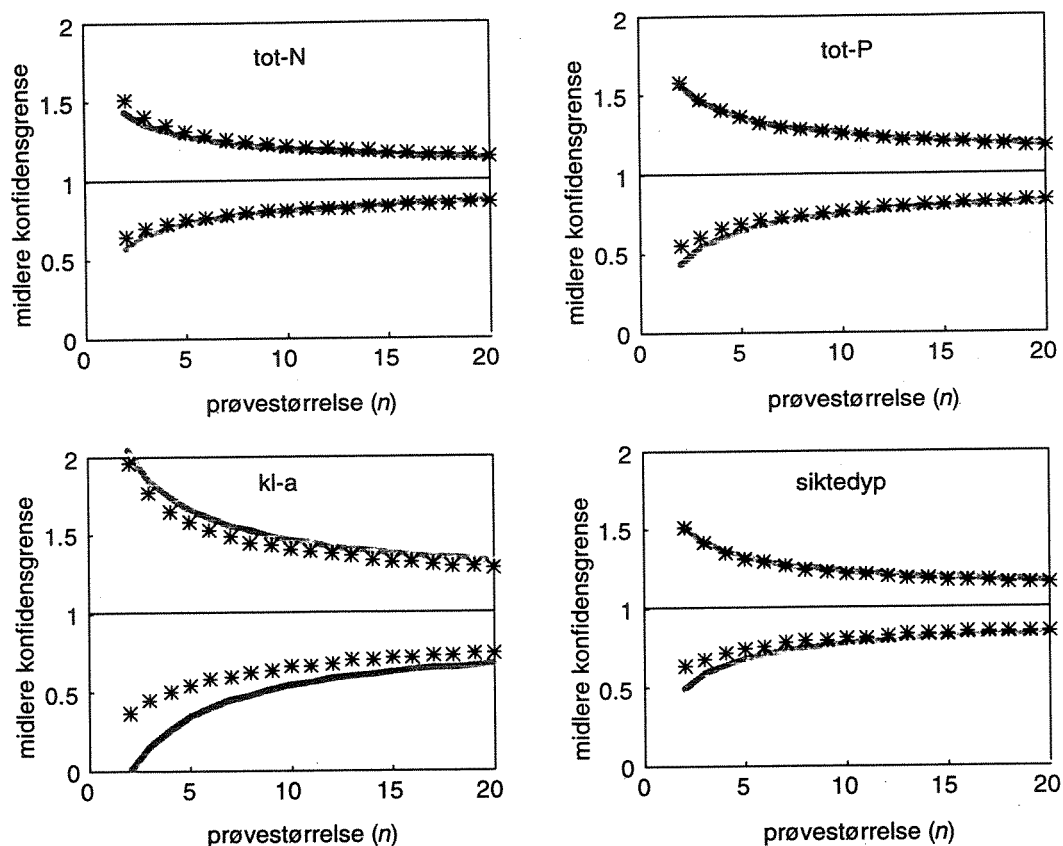
d. siktedyp. $R^2=0.53$, $n=48$, $p<0.0001$

	Parameter	Estimate	Std Error	t	p
Intercept	<i>k</i>	-0.938962	0.080517	-11.66	0.0000
log sikt mean	<i>a</i>	1.4656814	0.207900	7.05	0.0000
Nasjon[DK-N]	<i>b</i>	0.0011841	0.080517	0.01	0.9883
Interaksjoner	<i>c</i>	0.1829667	0.207900	0.88	0.3836

4.2.2 Bootstrapping

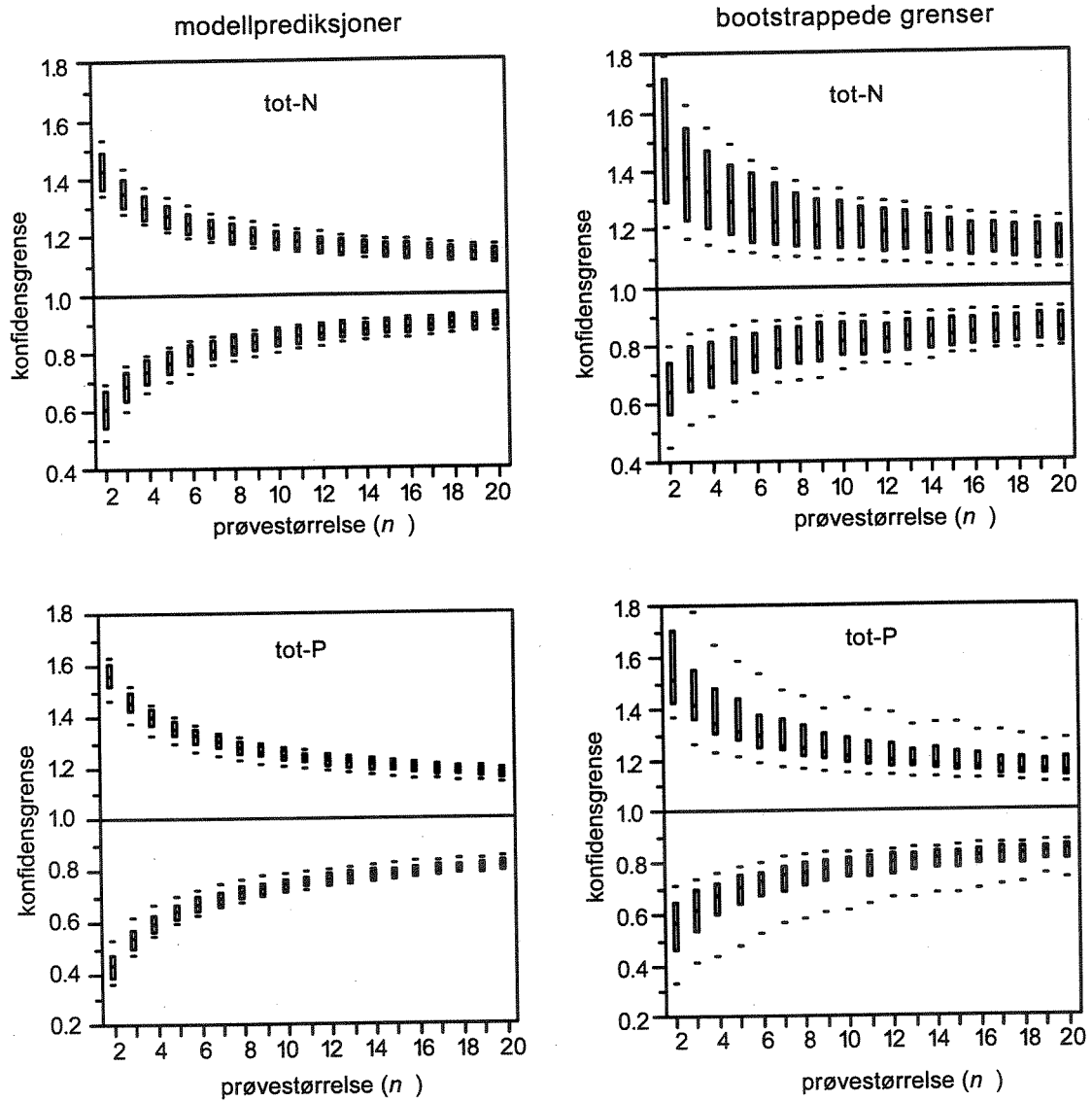
Bootstrappings-analysene genererte store mengder data som vi har vist i figur 4.5. Figuren presenterer gjennomsnittlige konfidensintervaller ved ulike prøvestørrelser. Før middelveiden ble beregnet har vi imidlertid normalisert intervallene, dvs. at vi har delt dem på middelveiden de tilhører. Dette gjør at øvre og nedre konfidensgrense ligger på hver side av 1, noe vi har indikert med en horisontal strek i diagrammene. For å kunne sammenlikne konfidensintervallene beregnet på konvensjonell måte ut fra modellene (etter likning 1 men hvor variansen er estimert etter likningene i tabell 3.1), og bootstrappede konfidensintervaller har vi plottet dem inn på de samme diagrammene. Figurene viser at de midlere modellerte konfidensgrensene stemmer godt overens med de midlere bootstrappede konfidensgrensene. En av de viktigste forutsetningene for modellene som beregner nødvendig prøvestørrelse (likning 5, figur 3.2) er derfor oppfylt. Figurene viser også hvorledes konfidensgrensene nærmer seg middelveiden etterhvert som prøvestørrelsen øker, og at forskjellene mellom modellerte og bootstrappede konfidensgrenser minsker ved økende prøvestørrelse.

Tar vi verdiene som er angitt på y-aksene i figur 4.1 og trekker fra 1, for så å multipliserer differansen med 100 får vi den relative nøyaktigheten oppgitt i %. Vi ser da at den relative nøyaktigheten ved en gitt prøvestørrelse i gjennomsnitt er best for tot-N og dårligst for kl-a.

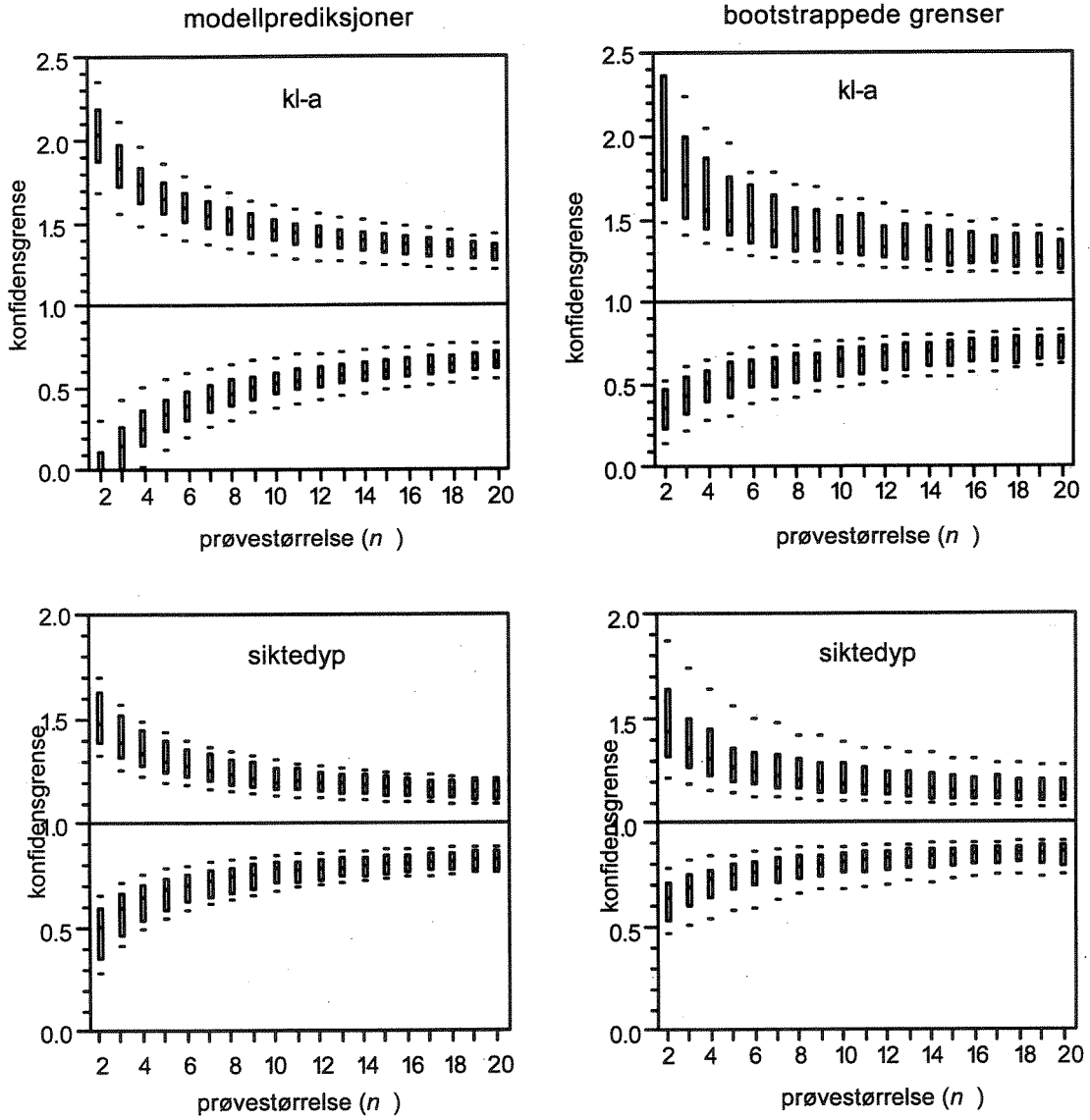


Figur 4.5 Midlere konfidensgrense som funksjon av prøvestørrelsen (n). Konfidensgrensene er normalisert ved å dele dem på deres tilhørende middelvei. Konfidensgrensene beregnet ved modellene (etter likning 1, med varians beregnet etter tabell 3.1) er angitt som en kurve, mens bootstrappede konfidensgrenser er angitt med stjerner.

Ved å framstille *midlere* normaliserte konfidensgrenser mister vi imidlertid viktig informasjon om variasjonen i estimatet. Vi har derfor i tillegg laget figurer som viser den *reelle* fordelingen av de normaliserte konfidensgrensene ved ulike prøvestørrelser (figur 4.6 og 4.7). Modellprediksjonene er framstilt på de venstre diagrammene, mens de bootstrappede er framstilt på de høyre diagrammene. Spredningen i modellestimatene reflekterer kun den effekten økende middelvei har på variansen i modellene. Resultatet av dette er jo som før nevnt at det er de mest eutofe innsjøene har de største konfidensintervallene. Sammenlikner vi konfidensgrensene fra modellprediksjonene med de bootstrappede konfidensgrensene ser vi at modellprediksjonene gir mindre spredning. En viktig årsak til dette er at vi ikke har bygd inn noen usikkerhet i modellene våre, og man bør derfor være noe konservativ når man anvender modellene til å anslå nødvendig prøvestørrelse eller oppnådd relativ nøyaktighet.



Figur 4.6 Normaliserte konfidensgrenser for tot-N og tot-P som funksjon av prøvestørrelsen (n). Venstre figurer viser modellprediksjonene, mens høyre figurer viser bootstrappede grenser. Søylene (eller boksene) ved en gitt n viser 25 og 75 % prosentilene. Den horisontale streken inne i søylen viser medianen, mens de horisontale strekene utenfor søylen viser 10 og 90% persentilene.



Figur 4.7 Normaliserte konfidensgrenser for *kl-a* og *siktedyb* som funksjon av prøve-
størrelsen (n). De to venstre figurer viser modellprediksjonene, mens de to til høyre viser
bootstrappede grenser.

REFERANSER

- Diaconis, P. og Bradley, E. 1983. Computer-intensive methods in statistics. *Sci. Amr.* 248(5): 96-129.
- Efron, B. og Tibshirani, R. J. 1993. *An introduction to the bootstrap*. Chapman and Hall. NY. 436 s.
- France, R.L. og R.H. Peters 1992. Temporal variance function for total phosphorus concentration. *Can. J. Fish. Aquat. Sci.* 49: 975-977.
- Faafeng, B., T. Tjomsland og B. Bjernkes 1987. Regional undersøkelse av eutrofierte innsjøer. Statistisk analyse av prøvetakingsfrekvens. NIVA-notat (O-87127), 16. okt. 1987. 25s.
- Jensen, J.P., E. Jeppesen, M. Søndergaard, J. Windolf, T.L. Lauridsen og L. Sortkjær 1995. Ferske vandområder - søer. Vandmiljøplanens Overvågningsprogram 1994. Danmarks Miljøundersøgelser. Faglig rapp. DMU 139: 116s.
- Knowlton, M.F., M.V. Hoyer og J.R. Jones 1984. Sources of variability in phosphorus and chlorophyll and their effects on use of lake survey data. *Wat. Res. Bull.* 20(3): 397-407.
- SFT. 1992 Klassifisering av miljøkvalitet i ferskvann. Kortversjon. SFT Rapport TA-905/1992. 31 s.

Norsk institutt for vannforskning

Postboks 173 Kjelsås
0411 Oslo

Telefon: 22 18 51 00
Telefax: 22 18 52 00

Ved bestilling av rapporten,
oppgi løpenummer 3427-96

ISBN 82-577-2961-2