

NGU-rapport 91.049

**Miljøkjemi og helse.  
Status pr. 01.02.91.**

|  |  |                          |                                    |                                     |            |
|--|--|--------------------------|------------------------------------|-------------------------------------|------------|
| Rapport nr. 91-049   |  | ISSN 0800-3416           |                                    | Åpen/ <del>Fortrolig</del>          |            |
| Tittel: Miljøkjemi og helse. Status pr. 01.02.91   |  |                          |                                    |                                     |            |
| Forfatter:<br>Øyvind Øyen  |  |                          | Oppdragsgiver:                     |                                     |            |
| Fylke:   |  |                          | Kommune:                           |                                     |            |
| Kartbladnavn (M. 1:250 000)  |  |                          | Kartbladnr. og -navn (M. 1:50 000) |                                     |            |
| Forekomstens navn og koordinater:  |  |                          | Sidetall: 14                       |                                     | Pris: 35,- |
|  |  |                          | Kartbilag:                         |                                     |            |
| Feltarbeid utført:   |  | Rapportdato:<br>29.01.91 |                                    | Prosjektnr.:<br>63.1856.26          |            |
|  |  |                          |                                    | Seksjonssjef: <i>Trine Bruilken</i> |            |
| <p>Sammendrag:</p> <p>Rapporten gir en oppsummering av 1,5 års arbeid med NAVF-prosjekt 363.88/012 "Miljøkjemi og helse", der kommunevise rater for sykkelighet av ulike typer kreft er sammenliknet med kommunevise verdier for geokjemisk sammensetning av løsmasser. Arbeidet er tidligere dokumentert i 9 fagrapporter ved NGU.</p> <p>Denne rapporten inneholder en oversikt over fagrapportene, hvilke metoder som er brukt og hvilke resultater arbeidet har gitt. Til slutt nevnes noen forslag til videre arbeid.</p> <p>Resultatene viser at vanligvis er det ingen samvariasjon mellom geokjemi og sykdom. I enkelte tilfeller er det imidlertid høye korrelasjoner mellom forekomst av kreft og visse elementer. Det er likevel ikke mulig på grunnlag av resultatene fra dette prosjektet å si om naturkjemi er en årsaksfaktor eller ikke.</p> |  |                          |                                    |                                     |            |
| Emneord Geokjemi   |  | Geomedisin               |                                    | Statistisk analyse                  |            |
| Fagrapport   |  |                          |                                    |                                     |            |
|  |  |                          |                                    |                                     |            |

## INNHOLD

|                                      | <u>Side</u> |
|--------------------------------------|-------------|
| 1. Summary in English                | 4           |
| 2. Innledning                        | 5           |
| 3. Oversikt over tidligere rapporter | 7           |
| 4. Diskusjon                         | 10          |
| 5. Videre arbeid                     | 12          |
| 6. Konklusjon                        | 13          |
| 7. Referanser                        | 14          |

## 1. SUMMARY IN ENGLISH

This report summarizes work documented in 9 reports at NGU dealing with geomedical associations between environmental geochemistry and municipal incidence rates of various types of cancer. This report gives an overview of the earlier reports, the methods being used and the results. Some recommendations to further work are also given.

The results show that some diseases and elements correlate. However, it is not possible to say whether or not the associations are causal or incidental.

## 2. INNLEDNING

NAVF-prosjekt 363.88/012 Miljøkjemi og helse har pågått siden 1/8 1989 med Øyvind Øyen ansatt som postdoktorstipendiat for en 3-årsperiode. Når jeg nå slutter (31/1 1991) har prosjektet pågått i 1,5 år. Denne rapporten er en sluttrapport over mitt arbeid, basert på 9 tidligere utgitte fagrapporter ved NGU.

Hensikten med prosjektet er å undersøke mulig samvariasjon mellom naturlig forekomst av grunnstoffer i berggrunnen og sykkelighet av ulike typer kreft. Prosjektet benytter allerede eksisterende data, som er samlet inn for andre formål. Denne rapporten gir en oversikt over de 9 NGU-rapportene som tidligere er utgitt. Til slutt gis en oppsummering og vurdering av de resultatene som har kommet frem, samt forslag til videre arbeid.

De kommunevise sykdomsdataene er stilt til disposisjon av Kreftregisteret og publisert i Atlas over kreftinsidens i Norge (Glattre et al. [10]). Ettersom noen norske kommuner har lave innbyggertall, slås små nabokommuner sammen til kommuneaggregater med mer enn 10000 innbyggere. Dette gir mer statistisk pålitelige data.

De geokjemiske dataene er basert på 690 prøver av flomsedimenter fra hele Norge. Prøvene er analysert kjemisk på totalinnhold av 31 og den syreløselige delen av 30 grunnstoffer. Kjemisk sammensetning av flomsedimentene regnes å representere forholdene i løsmasser og berggrunn i dreneringsfeltet. Dreneringsfeltene er i størrelsesområdet 60-300 km<sup>2</sup>. For å knytte prøvene til kommuner, må det i mange tilfeller brukes skjønn. Det gjelder særlig når en prøve kan sies å representere flere kommuner.

For å sammenliknes med sykdomsdata, der de kommunevise tallene er små, slås også her kommuner med mindre enn 10000 innbyggere sammen til kommuneaggregater med innbyggertall større enn 10000.

I store datamatriser vil signifikante korrelasjoner oppstå med lovmessig hyppighet som resultat av statistisk tilfeldighet. Derfor er det gjort forsøk med å dele opp datamaterialet i undergrupper for å se om enkelte korrelasjoner går igjen i flere grupper. Oppdelingen er foretatt ved å dele inn kommunene etter geografi, klima og kommunetype (urban/rural). Det er også foretatt tilfeldige oppdelinger. Ettersom de rurale kommunene anses som mest interessante for undersøkelse av naturforhold, er de fleste grupperingene gjort blant disse. I det følgende betegnes denne metoden som "aggregatutvalgmetoden".

En annen angrepsmåte er "naboaggregatmetoden". Her beregnes parvise kvotienter og differanser mellom både elementinnhold og sykkelighet i naboaggregater. Fordelen med denne angrepsmåten er at naboaggregater i noen grad kan forutsettes like når det gjelder sosiale og en rekke naturlige parametre som ikke er registrert i denne undersøkelsen, eks. topografi, klima, næringsliv osv. Dermed burde virkningen av forskjellen i disse faktorene være forholdsvis liten, slik at geokjemiens betydning fremheves.

Det er også benyttet multivariable metoder som prinsipal komponentanalyse (PCA) prinsipal komponentregresjon (PCR) og Partial Least Squares (PLS) for å studere dataene. Disse metodene er ikke nærmere forklart i rapportene, for teoretisk bakgrunn henvises til Martens og Næs [11] og tilsvarende litteratur.

Prinsipal komponentregresjon forsøkes utfra teorien om at det ikke nødvendigvis er de første prinsipalkomponentene som modellerer sammenhengen best, slik forutsetningen er ved PLS-modellering. Ved å korrelere skårer fra et datasett mot variable fra et annet, kan en finne ut om enkelte faktorer gir bedre korrelasjon enn andre. De kan da brukes til modellering.

Videre er det gjort beregninger med ulike variasjoner av skalering og sentrering av datasettene, likeså med gruppering av variable.

Generelt gjelder at data- og programfiler, som er brukt i dette prosjektet, er permanent lagret på magnetband ved NGUs sentrale dataanlegg. Filnavnene er oppgitt i rapportene, sammen med kort beskrivelse av filene.

### 3. OVERSIKT OVER TIDLIGERE RAPPORTER

NGU-rapportene 90-015 [1] og 90-029 [2] inneholder dokumentasjon av datamaterialet. Rapportene 90-037 [3] og 90-061 [4] viser parvise korrelasjoner mellom elementer og krefttyper. Det samme gjelder rapport 90-100 [5], men her benyttes forhold og forskjell mellom nabokommuneaggregater som datagrunnlag. I rapport 90-115 [6] benyttes multivariable statistiske metoder. Rapport 90-119 [7] sammenlikner resultatene fra [3] og [4] med resultatene fra [5]. Rapport 90-153 [8] benytter multivariat statistikk på forskjellsdataene fra [5]. Rapport 91-003 [9] sammenlikner i likhet med [7] de to angrepsmåtene ([3] mot [5]), men her brukes skalerte data, som gir mer pålitelige resultater enn uskalerte.

#### 3.1. 90-015.

Rapporten gir tabeller over aritmetisk og geometrisk gjennomsnitt for kjemisk sammensetning av flomsedimenter i kommuner og kommuneaggregater. Både aggregering og assosiering av prøver til kommuner er dokumentert i tabeller. Alle kommunene og kommuneaggregatene har minst 10000 innbyggere.

#### 3.2. 90-029.

Rapporten gir tabeller over sykdomsrater i kommuner/kommuneaggregater med mer enn 10000 innbyggere.

#### 3.3. 90-037.

Rapporten gir korrelasjoner mellom kreftsykelighet og elementinnhold i kommuner/kommuneaggregater. Beregningene er basert på dataene som er vist 90-015 [1] og 90-029 [2].

I beregningene er det brukt syreløselig del av 25 elementer og totalinnhold av 30 elementer. De øvrige dataene ble regnet som mindre pålitelige.

Oppdelingen i aggregatutvalg er vist i tabeller.

Resultatene er vist i korrelasjonsmatriser. For å gi bedre oversikt er bare koeffisienter signifikante på 10%-nivå eller bedre skrevet ut.

#### 3.4. 90-061.

Rapporten gir en oppsummering av resultatene fra 90-037 [3]. Det gis tabeller over hvilke parkombinasjoner av sykdom og element som gir flest signifikante korrelasjoner på ulike nivåer. Resultatene viser at de fleste kombinasjonene gir få signifikante korrelasjoner, mens andre går igjen ofte. Det mest karakteristiske er at "Annen hudkreft" gir svært mange høye korrelasjoner med flere elementer. Dette vil bli nærmere diskutert i kap. 4.

Totalt er antallet signifikante korrelasjoner høyere enn forventet ut ifra statistisk tilfeldighet. Et forhold som kan spille inn her, er at både sykdommer og elementer er innbyrdes

korrelerte. Dette medfører større sannsynlighet for å få tilfeldige signifikante korrelasjoner enn dersom variablene er uavhengige. Problemet blir diskutert nærmere i kap. 4.

### 3.5. 90-100.

Rapporten viser resultater fra "naboaggregatmetoden", en annen angrepsmåte enn "aggregatutvalgsmetoden", som ble brukt i 90-037 [3].

Resultatene viser at det er likegyldig om korrelasjonene beregnes for differansen mellom naboverdier eller for kvotienter av naboverdier. Derimot avviker resultatene fra "aggregatutvalgsmetoden", 90-061 [4]. Resultatene i rapport 90-100 [5] er imidlertid ikke helt pålitelige, ettersom det ble brukt uskalerte data. "Naboaggregatmetoden" blir vurdert på nytt i rapport 91-003 [9], se punkt 3.9.

### 3.6. 90-115.

Rapporten benytter multivariable metoder (prinsippal komponentanalyse (PCA) og Partial Least Squares (PLS), Martens og Næs [11]) for å studere dataene. Det er brukt data fra syreløselig del av 22 og totalinnhold av 28 elementer, de øvrige analyseverdiene er vurdert som upålitelige.

Det er gjort beregninger med ulike variasjoner av skalering og sentrering av datasettene, likeså med gruppering av variable, der sykdommer som er innbyrdes korrelerte, ble samlet i 18 grupper. Resultatene viser at dataene bør sentreres og skaleres, her til varians=1 for hver variabel. Forsøkene med gruppering ga ingen nye opplysninger. Resultatene gir ikke holdepunkter for å vurdere om det er forskjell mellom syreløselig og totalinnhold m.h.t. hvor godt dataene egner seg til geomedisinsk analyse.

Generelt tyder resultatene på at prinsippal-komponentmetoder ikke er egnet som verktøy til å analysere de datasettene som er brukt her. Egenverdiene viser ingen sterkt fallende tendens med avtakende rangering av prinsippalkomponenter. Dette tyder på at dataene ikke lar seg beskrive med noen få prinsippalkomponenter.

PLS-beregningene viser at flomsedimentdataene ikke gir noen god prediksjon av sykdomsdataene. Det er med andre ord ingen påvist matematisk sammenheng mellom de to datasettene. Prediksjonen blir heller ikke særlig bedre ved bruk av mange faktorer.

### 3.7. 90-119.

Rapporten sammenlikner resultatene fra "aggregatutvalgsmetoden", 90-037 [3] og 90-061 [4] med resultatene fra "naboaggregatmetoden", 90-100 [5].

Rapporten summerer resultatene fra de to metodene og viser hvilke parkombinasjoner som går igjen i begge metodene. Bl.a. viser resultatene at klimafaktorene som er undersøkt her, antall soldager og antall dager overskyet (1931-1960), ikke ser ut til



å ha noen innvirkning på resultatene. Dette ble undersøkt fordi sykdommen "Annen hudkreft" viste høye korrelasjoner med flere elementer, og fordi forekomst av hudkreft antas å ha sammenheng med hudens eksponering for solstråler.

Sammenlikningen mellom metodene viser i store trekk ulikheter, men også visse likheter. Som nevnt under 3.5. er imidlertid resultatene fra rapport 90-100 [5] upålitelige, slik at sammenlikningen ikke er særlig interessant. En ny vurdering blir foretatt i rapport 91-003, se pkt. 3.9.

### 3.8. 90-153

I likhet med i rapport 90-115 [6] benyttes her prinsipal komponentmetodene PCA og PLS. I tillegg brukes prinsipal komponentregresjon (PCR). (Martens og Næs [11].) Den siste forsøkes utfra teorien om at det ikke nødvendigvis er de første prinsipalkomponentene som modellerer sammenhengen best, slik forutsetningen er ved PLS-modellering. Ved å korrelere skårer fra et datasett mot variable fra et annet, kan en finne ut om enkelte faktorer gir bedre korrelasjon enn andre. De kan da brukes til modellering.

Resultatene viser at PCR gir dårligere prediksjon enn PLS. Forøvrig bekreftes resultatene fra rapport 90-115 [6]. Prinsipal komponentmetoder, slik de er anvendt i rapportene 90-115 [6] og 90-153 [8] ser ikke ut til å være egnete hjelpemidler for å finne ut om det er samvariasjoner mellom kreftdata og flomsedimentdata i Norge.

Samtidig som det ble beregnet korrelasjoner mellom skårer og variable, ble det også beregnet parvise korrelasjonskoeffisienter mellom element- og sykdomsvariable. Her ble det brukt de samme dataene som i rapport 90-100 [5]. Disse dataene er differansen mellom naboaggregater, den såkalte "naboaggregatmetoden". Forskjellen fra beregningene i rapport 90-100 [5] er at da ble ikke dataene skalert etter at differansene hadde blitt beregnet. Resultatene med uskalerte data anses som upålitelige, ettersom kommunepar med stor differanse veier uforholdsmessig tungt i beregningene når dataene ikke skaleres.

Resultatene viser store ulikheter mellom uskalerte og skalerte data. De skalerte dataene viser at "naboaggregatmetoden" gir resultater som samsvarer noe mere med "aggregatutvalgmetoden" enn det kunne se ut som ved bruk av uskalerte data.

Resultatene fra rapportene 90-100 [5] ("naboaggregatmetoden") og 90-119 [7] (sammenlikning mellom metodene) må derfor vurderes på nytt, se pkt. 3.9.

### 3.9. 91-003

Rapporten gir en ny vurdering av "naboaggregatmetoden" i forhold til 90-100 [5] og 90-119 [7], i og med at skalerte data er brukt. Resultatene viser til en viss grad samsvar mellom resultatene fra "naboaggregatmetoden" og "aggregatutvalgmetoden", 90-061 [4].

#### 4. DISKUSJON

I og med at det totalt er beregnet så mange korrelasjoner, må det av statistiske grunner ventes at enkelte kombinasjoner element/sykdom viser høye korrelasjoner. Det er ikke mulig på grunnlag av resultatene her å fastslå signifikansen av disse korrelasjonene. Likevel er det grunn til å merke seg de korrelasjonene som opptrer flest ganger og ved bruk av ulike metoder.

De enkelte elementene samvarierer innbyrdes. Når et element er korrelert med en sykdom, vil også andre samvarierende elementer være korrelerte med den samme sykdommen. Dermed kan visse tilfeldige korrelasjoner komme til å opptre oftere enn ventet. Dette går frem av korrelasjonene med "Annen hudkreft", der flere innbyrdes korrelerte elementer viser positive eller negative korrelasjoner med sykdommen. Dermed kan inntrykket av mulig signifikans bli for sterkt.

I rapport 90-061 [4] ble antallet beregnete signifikante korrelasjoner sammenliknet med forventet antall. På bakgrunn av det som er nevnt ovenfor, er det ikke mulig å trekke noen konklusjon ut fra at det beregnete antall er større enn det forventete. Prinsipalkomponentene forutsettes etter teorien å være uavhengige. Problemet med interkorrelasjon mellom forklaringsvariable er mindre for prinsipalkomponenter enn for de enkelte variablene. I rapport 90-153 [8] er prinsipalkomponenter for de geokjemiske dataene korrelert med prinsipalkomponenter for sykdomsdataene (vedlegg 3 side 8). Forutsatt at en kan regne signifikans som for enkeltvariabler, vil en på basis av 81 naboaggregatpar få resultater som i tabell 1.

Tabell 1.

Antall beregnete signifikante korrelasjoner mot antall forventet. 15 prinsipalkomponenter, geokjemi, mot 15 prinsipalkomponenter, sykdom.

|                                |       |      |
|--------------------------------|-------|------|
| Signifikansnivå                | 5%    | 1%   |
| Grense for signifikans         | 0,22  | 0,28 |
| Antall beregnete korrelasjoner | 29    | 13   |
| Antall forventete              | 11-12 | 2-3  |

Siden det ikke er interkorrelasjon mellom prinsipalkomponentene, skulle tabell 1 tilsi at det forekommer flere høye korrelasjoner enn forventet.

Påliteligheten av sykdomsdataene er ikke vurdert her, siden det ligger utenfor NGUs kompetanseområde. For "Nyrer, M", som gir mange høye korrelasjoner med Sr, er sykdomsratene forholdsvis lave. Dette indikerer at dataene er usikre statistisk sett, fordi et befolkningsgrunnlag på 10000 i dette tilfelle er lavt. "Annen

hudkreft", som gir høy korrelasjon med flere elementer, er ikke en enkelt sykdom, men en "samlepost". Dataene er publisert i Atlas over kreftinsidens i Norge (Glattre et al. [10]), men sikkerheten av insidensratene bør vurderes av medisinere.

Sannsynligvis gir naturkjemien høyest et beskjedent bidrag til utbredelsen av kreft. Det kreves derfor solide statistiske metoder for å finne mulige assosiasjoner mellom geokjemi og kreft. I dette prosjektet er det brukt kjente og velprøvde metoder. Det har ikke ligget innenfor arbeidsområdet å utvikle ny metodikk. I geomedisin vil problemstillingen ofte være å undersøke påvirkningsfaktorer som er i støyområdet. Dette er forholdsvis lite behandlet i litteraturen.

I tillegg til at påvirkningen sannsynligvis er liten, er spranget fra berggrunn til kreft meget langt. Det synes rimelig å se på sammenhengen mellom sammensetning av berggrunn og løsmasser på den ene siden og elementinnhold i planter, dyr og mennesker på den andre for å se i hvor stor grad naturkjemien påvirker levende organismer mere generelt. Resultatene fra dette prosjektet gir få holdepunkter for å si noe sikkert om hvorvidt naturkjemi er en medvirkende faktor når det gjelder forekomst av kreft i Norge.

## 5. VIDERE ARBEID

Det er liten grunn til å fortsette arbeidet med bare de datasettene som her er brukt. Velkjente metoder ser ikke ut til å føre frem, og dataene egner seg dårlig til metodeutvikling, ettersom sammenhengene i beste fall er svake.

Det foreligger datasett over elementinnhold i jord og moser. Disse er i større eller mindre grad påvirket av forurensing. Det kan være interessant å sammenlikne disse både med flomdataene og kreftdataene. I tillegg finnes det også data over innhold av enkelte grunnstoffer i melketenner som kan brukes i denne sammenhengen.

Dersom disse datasettene brukes, bør en starte med vanlig parvis korrelasjonsanalyse. De små forsøkene som er gjort med multi-variabel statistikk gir ingen holdepunkter for å påstå at slike metoder er velegnede hjelpemidler i dette arbeidet.

Ettersom naturforhold er en liten parameter i sykdomssammenheng, bør en søke å trekke inn dominerende faktorer som f.eks. røyking og forurensing, for deretter å se på eventuell restvarians i datasettet når disse store faktorene er tatt hensyn til. En kan enten trekke ut andre parametre enn geokjemi, eller en kan se om noen av disse går sammen med geokjemien. Aktuelle parametre kan være avstand fra kysten eller befolkningstetthet, ettersom de kan kvantifiseres.

Det finnes også andre sykdomsdata som kan være av interesse. Hjertekarsykdommer har høyere rater enn kreft, og de kan derfor brytes ned til mindre geografiske enheter. Dermed vil en få mindre variasjoner innen hvert enkelt område.

Endelig bør det vurderes om det skal utarbeides et nytt kreftatlas. Det forrige gjaldt perioden 1970-79, og en har nå data for en ny 10-årsperiode. Det gir muligheter for å oppdatere, å slå sammen til en 20-årsperiode, eller å se på utviklingen fra et tiår til et annet.

## **6. KONKLUSJON**

Det er ikke mulig på grunnlag av resultatene så langt i dette prosjektet å avgjøre om naturkjemi er en medvirkende faktor eller ikke når det gjelder utbredelse av kreft i Norge.

## 7. REFERANSER

1. Ø. Øyen, B. Bølviken, R. Nilsen:  
Geokjemisk karakterisering av norske kommuner ved hjelp av flomsedimentdata,  
NGU-rapport 90-015.
2. Ø. Øyen, B. Bølviken, R. Nilsen:  
Kreftsykelighet i norske kommuner og kommuneaggregater,  
NGU-rapport 90-029.
3. Ø. Øyen, B. Bølviken, R. Nilsen:  
Samvariasjon mellom sykkelighet av kreft og geokjemisk sammensetning av flomsedimenter,  
NGU-rapport 90-037.
4. Ø. Øyen, B. Bølviken, R. Nilsen:  
Oversikt over korrelasjoner mellom kreftsykelighet og kjemisk sammensetning av flomsedimenter,  
NGU-rapport 90-061.
5. Ø. Øyen:  
Samvariasjon mellom kreftsykelighet og kjemisk sammensetning av flomsedimenter i nabokommuner,  
NGU-rapport 90-100.
6. Ø. Øyen:  
Prinsippal komponentanalyse av flomsediment- og kreftdata,  
NGU-rapport 90-115.
7. Ø. Øyen:  
Sammenlikning av metoder for beregning av korrelasjoner mellom geokjemi og kreft,  
NGU-rapport 90-119.
8. Ø. Øyen:  
Prinsippal komponentregresjon flomsediment- og kreftdata,  
NGU-rapport 90-153.
9. Ø. Øyen:  
Geokjemi og kreft. Sammenlikning av to metoder med skalerte data,  
NGU-rapport 91-003.
10. E. Glattre, T.E. Finne, O. Olesen og F. Langmark:  
Atlas over kreftinsidens i Norge 1970-1979,  
Landsforeningen mot kreft/Kreftregisteret (1985).
11. H. Martens og T. Næs:  
Multivariate Calibration,  
John Wiley & Sons, 1989.