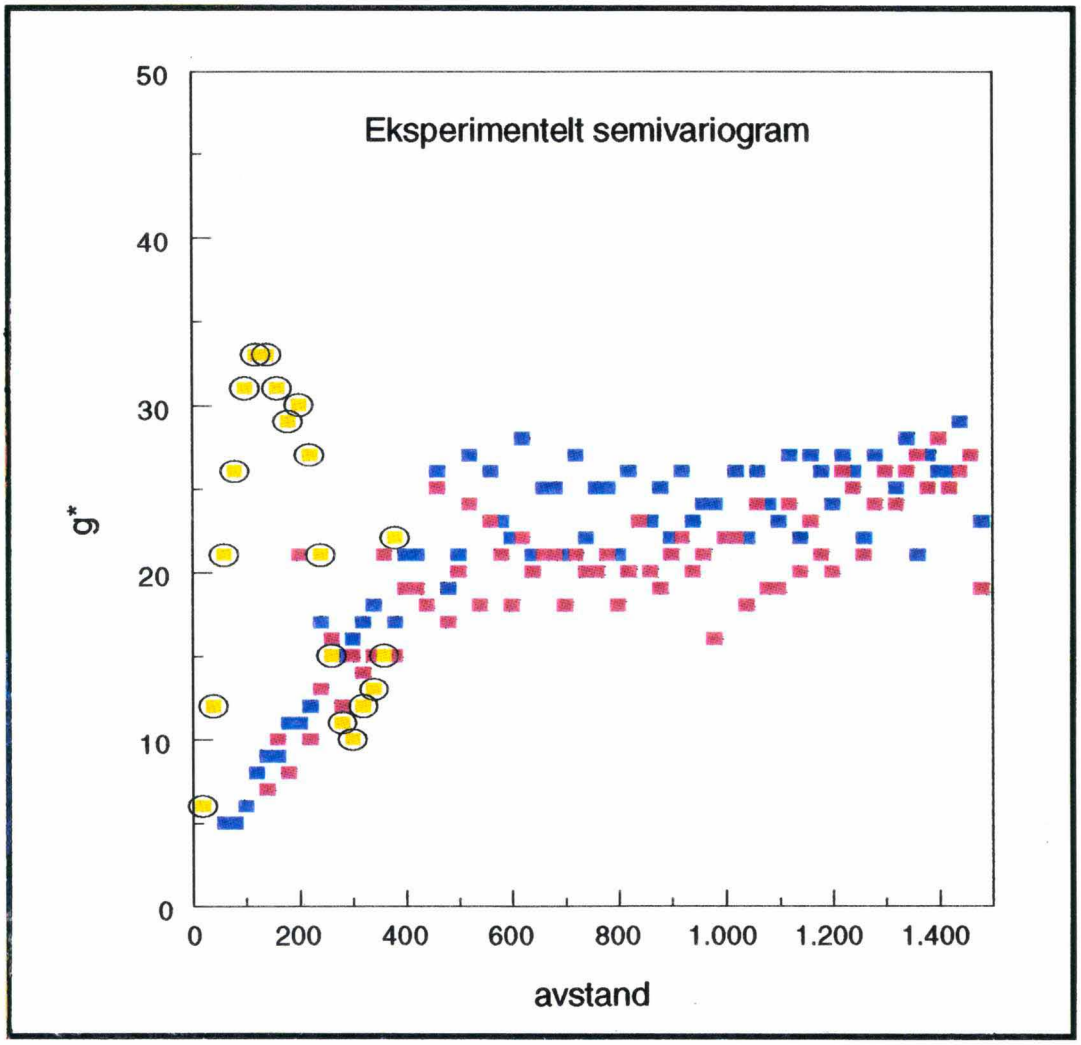
NOTAT

Nr 03
1995

NVE
NORGES VASSDRAGS-
OG ENERGIVERK

Nils-Otto Kitterød

GEOSTATISTIKK I GEOHYDROLOGI



HYDROLOGISK AVDELING

NVE
NORGES VASSDRAGS-
OG ENERGIVERK

TITTEL Geostatistikk i geohydrologi	**NOTAT** Nr. 03 1995 DATO 05.04.95
FORFATTER Nils-Otto Kitterød	

SAMMENDRAG

For alle som arbeider med hydrologi, er det en kjent sak at de fleste observasjoner og prognoser er beheftet med en viss usikkerhet. Til tross for disse usikkerhetene er det vår oppgave å utarbeide best mulige estimater av ulike hydrologiske parametre. I den senere tid er det utviklet statistiske metoder som gir optimale estimater samtidig som eventuelle usikkerheter i datagrunnlaget kan tas hensyn til. Det formelle begrepsapparatet som håndterer disse problemene har fått betegnelsen geostatistikk. Geostatistikk har etterhvert utviklet seg til en ny vitenskapsgren i krysningsområdet mellom geovitenskapene og tradisjonell statistikk. Det er vår oppfatning at disse metodene vil øke kvaliteten av våre hydrologiske analyser dersom det blir innført som en del av våre rutiner. I dette notatet gis en kort innføring i de viktigste begrepene innenfor geostatistikk samt en forklaring av de vanligste likningene.

ABSTRACT

A common problem in hydrology is spatially sparse observations often in an unregular grid. And there is always some uncertainties in the observations. In spite of this, we have to estimate critical attributes like hydraulic conductivities, soil moisture content or river discharge. In later years quantitative methods in spatial statistics have been developed that handle this kind of problems. These methods are grouped under the concepts of 'Geostatistics'. We think that these methods will increase the quality of our hydrological work. This document is a brief introduction to the theory of 'Geostatistics'.

EMNEORD /SUBJECT TERMS
geostatistikk, estimering, usikkerhet

ANSVARLIG UNDERSKRIFT

Arne Tollan
avd.direktør

Kontoradresse: Middelthunsgate 29
Postadresse: Postboks 5091. Maj.
0301 Oslo

Telefon: 22 95 95 95
Telefax: 22 95 90 00

Postgiro: 0803 5052055

Forord

Bakgrunnen for å utgi dette relativt 'teoretiske' notatet som en del av NVEs publikasjonsserier kan oppsummeres i følgende virkelighetsbeskrivelse av norsk vannforvaltning: Vannressursene - globalt og nasjonalt - utsettes for stadig større press ikke bare pga. økt forurensning, men også fordi bevisstheten om at vann i ulik form, skal dekke flere behov enn rent materielt forbruk. Dette skjerper kravene om en optimal forvaltning. Samtidig må forvaltningen forholde seg til en økonomisk realitet som innebærer at det må prioriteres skarpere. For Hydrologisk avdeling betyr dette at vi må øke kvaliteten av våre hydrologiske vurderinger samtidig som kostnadene knyttet til hydrologiske observasjoner må skjæres ned. Dette øker sårbarheten i de observasjonene som fortsatt blir gjort. En unøyaktig observasjon blant ti korrekte er langt mer alvorlig enn en feil blant hundre riktige. Disse forholdene gjør at vi må legge vekt på analyse- og prognoseverktøy som 'hedrer' høykvalitets observasjoner og samtidig legger mindre vekt på observasjoner av dårligere kvalitet. Disse aspektene kan bli håndtert ved hjelp av geostatistikk. Vi mener derfor det er bryderiet verdt å sette seg inn i de grunnleggende begrep innenfor dette temaet. Dette notatet vil forhåpentligvis kaste lys over problemet antydet ovenfor, pluss en del andre interessante muligheter.

Oslo 7. april 1995

Kjell Repp
seksjonssjef

Geostatistikk i geohydrologi

av Nils-Otto Kitterød

Innledning

Bedre datagrunnlag er et utbredt ønske innenfor hydrologi. Geohydrologi er intet unntak. Fordi markvann og grunnvann som regel befinner seg i et sterkt inhomogent og anisotropt medium, blir databehovet formidabelt. Markvanns- og grunnvannsdata er i tillegg ofte kostnadskrevende å samle inn. Dette gjør at vi ofte har behov for å estimere verdier i områder eller punkter hvor vi ikke har observasjoner. I løpet av de siste 10-årene er det utviklet en ny vitenskapsgren - geostatistikk - hvor et hovedsiktemål er å utvikle objektive metoder for å beregne optimale estimater.

Som motivasjon til dette notatet vil jeg skissere et typisk problem i grunnvannsmodellering: Å estimere en parameter i ethvert punkt i akviferen på grunnlag av svært få observasjoner. Nøkkelparameteren er alltid hydraulisk ledningsevne eller transmissivitet. Denne parameteren er kostbar å måle, og forekommer derfor som regel i svært få punkter. Grunnvannstand derimot er enklere å måle og datadekningen er ofte brukbar i tid og rom. Kornfordeling eller kjerneprøver forekommer kun i et begrenset antall punkter. Dreietrykksonderinger eller annen geoteknisk informasjon kan forekomme hvis vi er heldige. Dersom det er observasjonsrør i området er det (relativt) enkelt å foreta pumpetester og registrere spesifikk pumperate, eventuelt infiltrasjonskapasitet ("slug" tester). I spesielle tilfeller kan vi også ha geofysiske data. (Elektriske sonderinger, georadar, seismisk-refraksjon eller -refleksjon). Alle disse dataene er mer eller mindre romelig korrelerte innbyrdes. I tillegg kan forskjellige dataserier være korrelerte seg i mellom. Hydraulisk ledningsevne er f.eks. korrelert med kornfordeling og spesifikk pumperate. I geostatistikken er det utviklet en metode hvor man kan estimere en datatype, f.eks. transmissivitet med andre korrelerte dataserier som f.eks kornfordeling. Disse dataene representerer imidlertid forskjellige skalaer eller volum. Langvarige pumpeforsøk gir hydraulisk ledningsevne over et større volum enn kortvarig pumping, mens kornfordeling kun gir verdier i et punkt i akviferen. For å løse slike problemer er det utviklet forskjellige estimeringsmetoder som går under fellesbetegnelsen **kriging**. Kriging gir alltid det best mulige estimatet ut i fra de observasjonene man har og de forutsetningene man arbeider innenfor.

I tillegg til optimale estimater gir kriging også den statistiske **usikkerheten** knyttet til estimatet. Slik "rommelig" usikkerhet er svært nyttig når man skal vurdere estimatene eller benytte resultatene videre. Siden usikkerheten ikke er knyttet til verdien av målingene, bare til den geografiske posisjon, kan kriging også benyttes for å optimalisere målenettet.

Geostatistikk

Hydrologiske variabler som f.eks. grunnvannstand, hydraulisk ledningsevne eller vannføring, har alle et element av tilfeldighet. I statistikken kalles ofte slike tilfeldige variabler stokastiske variabler Z. En stokastisk variabel kan variere som en funksjon av geografisk posisjon $Z=Z(x)$, eller som funksjon av tiden $Z(t)$. I geostatistikken er det $Z(x)$ som er gjenstand for analyse, mens $Z(t)$ er temaet i tidsserieanalyse. La oss anta at en hydrologisk observasjon kan sammenlignes med et eksperiment, og at vi kan gjenta eksperimentet men samme betingelser flere ganger. Vi har da flere **realisasjoner** ξ, av den stokastiske variablene $Z(x,\xi)$. Hvis oppgaven f.eks. er å måle bredden på samtlige kanaler i et forgreinet elveløpssystem, kan hver kanal sees som en realisasjon. Bredden kan variere fra punkt til punkt og fra kanal til kanal.

Grunnleggende begrep

For å gjøre statistiske slutninger om $Z(x,\xi)$ trengs følgende begrep:

1) Forventningen

$$E[Z(x_1,\xi)] = m(x_1)$$

2) Varians

$$E[(Z(x_1,\xi) - m(x_1))^2] = \sigma^2(x_1)$$

3) Covarians

$$E[(Z(x_1,\xi) - m(x_1))\,(Z(x_2,\xi) - m(x_2))] = Cov_Z(x_1,x_2) = C(x_1,x_2)$$

Korrelasjonskoeffisienten ρ, er gitt ved 3) og 2);

$$\rho = \frac{C(x_1,x_2)}{\sigma(x_1)\cdot\sigma(x_2)}$$

1) og 2) kalles 1. og 2. ordens moment.

For å beskrive $Z(x,\xi)$ fullstendig må man også kjenne sannsynlighetstetthetsfunksjonen (stf) til variabelen. De fleste statistiske metoder antar at stf er normalfordelt. I geostatistikken er ikke det noen forutsetning, men resultatene blir bedre desto mer normalfordelt stf er.

Forenklinger

<u>Sterk stasjonæritet:</u>

Sannsynlighetstetthetsfunksjonen til den stokastiske variablene er uavhengig av geografisk posisjon. Eller sagt på annen måte, alle momentene til Z er uavhengige av x.

<u>2. ordens stasjonæritet:</u>

Stasjonæritet m.h.t. forventning

$$E[Z(x,\xi)] = m \qquad \forall \quad x$$

Stasjonæritet m.h.t. varians

$$E[(Z(x,\xi - m(x))^2] = \sigma^2 \quad \forall \quad x$$

Stasjonæritet m.h.t. kovarians

$$E[(Z(x_1,\xi) - m)(Z(x_2,\xi) - m)] = Cov_Z(h) = C(h)$$

hvor $h = x_1 - x_2$ h er altså relativ avstand mellom målepunktene. I geologisk sammenheng blir ofte dette kalt for **homogenitet**.

Ergodisitet:

Ergodisitet betyr at en unik realisasjon av den stokastiske variabelen $Z(x)$ oppfører seg med samme sannsynlighetstetthetsfunksjon som alle tenkelige realisasjoner av samme variabel. Er prosessen ergodisk kan man beregne de statistiske momentene i rommet (eller tiden hvis det er snakk om tidsserier) i stedet for over alle mulige realisasjoner. ("Alle mulige realisasjoner" kalles ofte "ensembelet"). I praksis har vi kun en realisasjon av prosessen å forholde oss til, og dermed må vi ofte anta at prorosessen er ergodisk uten å kunne bevise det formelt. Matematisk kan dette uttrykkes på følgende måte:

$$E[Z(x,\xi) = E[Z(x)] = m(x)$$

$$E[(Z(x,\xi) - m)^2] = E[(Z(x) - m)^2] = \sigma^2(x)$$

$$E[(Z(x_1,\xi) - m)(Z(x_2,\xi) - m)] =$$
$$E[(Z(x_1) - m)(Z(x_2) - m)] = Cov_Z(x_1 - x_2)$$

Enkel kriging

Anta 2. ordens stasjonæritet, ergodisitet og at forventningen m, varians σ^2, og kovarians $C(h)$ er kjent. Estimatet $Z^*(x_0)$ er gitt som en vektet sum av observasjonene $Z(x_i)$;

$$Z^*(x_0) = Z_0^* = \sum_{i=1}^{n} \lambda_i Z(x_i)$$

Vi er interessert i et optimalt estimat, og det finner vi ved å minimalisere feilvariansen.

La:

$$Y = Z(x) - m$$

og

$$Y_0^* = \sum_{i=1}^{n} \lambda_i Z(x_i)$$

Det optimale estimatet får vi når

$$A = Var(Y_0^* - Y_0) = E[(Y_0^* - Y_0)^2] = minimum$$

$$A = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(x_i - x_j) + \sigma_y^2 - 2\sum_{i=1}^{n} \lambda_i C(x_i - x_0)$$

hvor vi antar at $\sigma_y{}^2$ og $C(x_i - x_j)$ er kjent.

Det beste estimatet får vi når

$$\frac{\partial A}{\partial \lambda_i} = 2\sum_{j=1}^{n} \lambda_j C(x_i - x_j) - 2C(x_i - x_0) = 0$$

eller

$$\sum_{j=1}^{n} \lambda_j C(x_i - x_j) = C(x_i - x_0) \qquad i=1,2,\ldots,n$$

som er den **enkle kriging likningen**.

Hvis $n = 3$ har vi matrisen

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} = \boldsymbol{C}$$

$$\boldsymbol{C} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ C_{30} \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \boldsymbol{C}^{-1} \begin{bmatrix} C_{10} \\ C_{20} \\ C_{30} \end{bmatrix}$$

Skal λ_i beregnes må $\boldsymbol{C}$ være invertibel, eller definit positiv. Dersom observasjonspunktene har fast posisjon (og det har de som regel), kan $\boldsymbol{C}$ matrisen beregnes en gang for alle.

Variansen i feilestimatet får vi ved å sette kriginglikningen inn i $\boldsymbol{A}$;

$$\sigma_{err}^2 = \sigma_y^2 - \sum_{i=1}^{n} \lambda_i C(x_i - x_0)$$

I selve målepunktet er estimatet eksakt, $\sigma_{err}{}^2 = 0$. Desto større avstand der er mellom estimatet og observasjonspunktet desto større blir $\sigma_{err}{}^2$.

Ordinær kriging

I praksis er antagelsen om 2.ordens stasjonæritet sjelden eller aldri oppfylt og forventning og varians er ukjent. Vi trenger derfor en mer generell antagelse. Dette behovet blir tilgodesett i den såkalte **intrinsic hypotesen** (eller "indre antagelsen").

Hypotesen går i all enkelhet ut på at man ikke krever 2. ordens stasjonæritet av selve observasjonene, men i differansen mellom observasjonene.

Dette innebærer at:

$$E[Z(x)] = m(h)$$

$$E[(Z(x+h) - Z(x))^2] = 2\gamma(h)$$

Dvs. forventningen og varians er funksjon av relativ avstand h, ikke av absolutt posisjon x. Intrinsic hypotesen innebærer også at variansen **ikke** behøver å være endelig.

$\gamma(h)$ kalles **semi-variogrammet**. I definisjonen av semi-variogrammet - i motsetning til kovariansen - inngår ikke forventningen. Semi-variogrammet er dermed en mer generell definisjon av kovarians.

$\gamma(h)$ erstatter $C(h)$ i den enkle kriging likningen, og samme krav gjelder $\gamma(h)$ som $C(h)$. Dvs. at $\gamma(h)$ må være **definit positiv**. For at dette kravet skal være oppfylt er det utarbeidet en rekke teoretiske semi-variogram modeller som kan tilpasses det semi-variogrammet som er beregnet for vårt datasett. (Semi-variogram beregnet fra observerte data kalles ofte **eksperimentelle semi-variogram** og betegnes $\gamma^*(h)$.)

Vi stiller nå to krav til estimatet $Z_0^* = \Sigma\lambda_i Z_i$. Det skal være:

1) Forventningsriktig (unbiased)
2) Optimalt, dvs. feilvariansen skal minimaliseres

Første krav medfører

$$E[Z_0^* - Z_0] = 0$$

$$E[\sum_{i=1}^{n} \lambda_i Z_i] = E[Z_0]$$

$$\sum_{i=1}^{n} \lambda_i E[Z_i] = E[Z_0]$$

$$\sum_{i=1}^{n} \lambda_i m = m$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

I tillegg skal variansen i estimasjonsfeilen A, minimaliseres. Vi skal altså finne et minimum med tilleggskrav at $\Sigma\lambda_i = 1$.

$$B = A + \mu(\sum_{i=1}^{n} \lambda_i - 1)$$

μ kalles Lagranges multiplikator. Med intrinsic hypotesen er $A=\frac{1}{2}E[(Z_0^*-Z_0)^2]$.

$\partial B/\partial\lambda_i=0$ gir de **ordinære kriging likningene**

$$\sum_{j=1}^{n} \lambda_j \, \gamma(x_i-x_j) + \mu = \gamma(x_i-x_0)$$

$$\sum_{j=1}^{n} \lambda_i = 1 \qquad i = 1,2,\ldots,n$$

Feilvariansen i estimatet er

$$\sigma_{err}^2 = Var(Z_0^* - Z_0) = E[(Z_0^* - Z_0)^2] = \sum_{i=1}^{n} \lambda_i \, \gamma(x_i - x_0) + \mu$$

Kriging matrisen har fått en rad og en kolonne i tillegg pga. Lagranges multiplikator

$$\begin{bmatrix} 0 & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & 0 & \gamma_{23} & \cdots & \gamma_{1n} & 1 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \gamma_{n1} & \gamma_{n2} & \cdot & \cdots & 0 & 1 \\ 1 & 1 & \cdot & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{n0} \\ 1 \end{bmatrix}$$

hvor $\gamma_{ij} = \gamma(x_i - x_j)$ og $\gamma_{ii} = \gamma(x_i - x_i) = 0$

Første steg i krigingen er å finne en god modell for $\gamma^*(h)$. Det er 3 begreper i variogram analysen det er viktig å kjenne til;

1) C_0, nugget (gullklump)
2) $C(0)$, sill (terskel)
3) a, range (rekkevidden)

Nugget effekten tilsvarer en bakgrunnsvariasjon i dataene. Tar man to observasjoner i samme punkt (til samme tid) er $\gamma^*(h) = 0$. Hvis man derimot tar to observasjoner svært nær hverandre kan differansen i målingene ha gjort et tilfeldig hopp. Dette tilfeldige hoppet kalles altså nugget effekten.

Sill er den konstante verdi variogrammet får når avstanden mellom målepunktene er så stor at det ikke lenger er korrelasjon mellom observasjonene. Sill verdien tilsvarer variansen i datamaterialet. Dersom korrelasjonen mellom punktene er uendelig, har ikke variogrammet noen sill eller terskelverdi.

Rang er den avstanden hvor variogrammet når sill-verdien. Dvs. den rekkevidden hvor det er korrelasjon mellom observasjonspunktene.

Kryssjekking

Etter at man har tilpasset en variogram-modell til det eksperimentelle variogrammet må modellen kryssjekkes. Kryssjekkingen består i at man regner ut et estimat for hvert enkelt målepunkt på grunnlag av de andre dataene uten at målepunktet selv benyttes i estimatet. Kryssjekking blir også kalt for 'cross validation', eller "Thomas-prosedyre" etter apostelen Thomas - han som tvilte på oppstandelsen.

$$Z_1^* = \sum_{i=2}^{n} \lambda_i Z(x_i)\,, \quad Z_2^* = \sum_{\substack{i=1 \\ i\neq 2}}^{n} \lambda_i Z_i\,, \quad \ldots\,, \quad Z_n^* = \sum_{i=1}^{n-1} \lambda_i Z_i$$

Hvor god modellen er bestemmes av følgende kriterier:

1) Estimatet må være forventningsriktig, $E[Z^*-Z] \approx 0$,eller $1/n\sum Z_i^*-Z_i \approx 0$
2) "Standardisert feilvarians" $E_i[\{(Z^*-Z)/\sigma_i\}^2] \approx 1$,eller $1/n\sum_i\{(Z^*-Z)/\sigma_i\}^2 \approx 1$
3) Variansen $E[(Z^*-Z)^2]$, skal være så liten som mulig

Anisotropi

Hittil har vi latt semi-variogrammet vært en funksjon av avstanden mellom målepunktene,$\gamma=\gamma(h)$. Dersom dataene er anisotrope kan man i tillegg til avstanden også la semi-variogrammet være en funksjon av retningen, $\gamma=\gamma(h,\theta)$. Dette krever selvsagt en stor data tetthet, og i praksis må man søke innenfor en sektor.

For å undersøke om det er anisotropi i dataene bør man alltid undersøke de 4 hovedretningene. På bakgrunn av dette kan man lage en anisotropi ellipse. I kryssjekkingen og i selve estimeringen (kriging-rutinen) transformeres dataene i henhold til anisotropi ellipsen slik at datasettet blir isotropt, og de vanlige kriging likningene kan benyttes.

Blokk-kriging

Til nå har vi kun betraktet observasjonsdataene som punkt målinger og estimatene kun som punktestimater. I blokk-kriging estimeres volum verdier eller areal verdier. Dette gjør blokk-kriging til et aktuelt verktøy i hydrologi fordi hydrologiske data ofte representerer arealverdier. Vannføringsmålinger er et godt eksempel på slike data.

I blokk-kriging må vi arbeide med integraler i stedet for enkelt verdier som i vanlig kriging. Hvis Z_0 representerer arealet S_0 får vi gjennomsnittsverdien av Z_0 ved;

$$\bar{Z}_0 = \frac{1}{S_0} \int_{S_0} Z(x)\, dx$$

og estimatet blir;

$$\bar{Z}_0^* = \frac{1}{S_0} \int_{S_0} Z^* ds = \frac{1}{S_0} \int_{S_0} \sum_{i=1}^{n} \lambda_i Z_i\, ds$$

Mens kriging likningene blir modifisert til;

$$\sum_{j=1}^{n} \lambda_j\ \gamma(x_i - x_j) + \mu = \bar{\gamma}(x_i - x_0)$$

$$\sum_{j=1}^{n} \lambda_i = 1 \qquad i = 1,2,\ldots,n$$

hvor

$$\bar{\gamma}(x_i - x_0) = \frac{1}{S_0} \int_{S_0} \gamma(x_i - x_0)\, ds$$

Og variansen i feilestimatet

$$\sigma_{err}^2 = Var(\bar{Z}_0^* - Z_0) = \sum_{i=1}^{n} \lambda_i\, \bar{\gamma}(x_i - x_0) + \mu - \bar{\bar{\gamma}}(x_0)$$

hvor

$$\bar{\bar{\gamma}}(x_i - x_0) = \frac{1}{S_0^2} \int_{S_0}\int_{S_0} \gamma(x - y)\, dx\, dy$$

Variansen blir m.a.o. mindre når vi ser på volum eller areal estimater fordi variansen i en gjennomsnittsverdi er mindre enn variansen i punktmålinger. Eller sagt på en annen måte, det er lettere (eller mindre usikkert) å estimere gjennomsnittsverdien på en variabel i et område enn å estimere punktverdiene i samme område.

Cokriging

Dette er en nyttig estimeringsteknikk dersom man har 2 eller flere dataserier som er korrelerte. La $Z_1(x)$ og $Z_2(x)$ være 2 korrelerte datasett. Estimatet av $Z_1(x)$, $Z_1^*(x)$, kan da beregnes som en vektet sum av $Z_1(x)$ og en vektet sum av $Z_2(x)$.

$$Z_1^*(x_0) = \sum_{k=1}^{n} {}^1\lambda_k Z(x_k) + \sum_{k=1}^{m} {}^2\lambda_k Z_2(x_k)$$

Som tidligere kreves det at estimatet er
1) Forventningsriktig, $E[Z_1^*]=0$.
2) Variansen i feilen er minimum, $Var(Z_1^*-Z_1)$ ***minimum***.

Siden $Z_1(x)$ **og** $Z_2(x)$ er korrelerte må kryss-korrelogram også beregnes. Disse kryss-korrelogrammene inngår selvsagt i kriginglikningene. (For nærmere detaljer se "Quantitative Hydrology" av de Marsily, 1986).

Ikke stasjonær kriging

Til nå har vi antatt 2.ordens stasjonæritet enten i selve datamaterialet eller i differansen (intrinsic hypotesen). Ofte er dette en for streng antagelse. Forventning og variansen vil da variere med geografisk posisjon. I slike tilfeller må kriging likningene modifiseres.

Dersom det er en trend i observasjonene i en bestemt retning kan man unngå ikke stasjonær kriging ved å se på dataene som **anisotrope**. I grunnvannssammenheng kan man tenke seg en homogen akvifer med en dominerende strømningsretning. Statistisk sett er ikke grunnvannsspeilet homogent (eller stasjonært) fordi gjennomsnittlig grunnvannstand endres med geografisk posisjon. Man kan imidlertid unngå ikke stasjonær kriging ved å beregne $\gamma(h)$ vinkelrett på ekvipotensial linjene.

Stasjonæritet er ofte et spørsmål om skala. I regional skala er ofte observasjonene ikke stasjonære, men på lokal skala kan man til en viss grad si at dataene er stasjonære. I slike tilfelle kan man unngå ikke stasjonær kriging ved å benytte såkalt **"bevegelig naboskap"** eller **moving neighbourhood**. Man tillater da bare observasjoner innenfor en begrenset radius å være med i beregningene av kriginglikningene. Dette begrensede området flyttes fra plass til plass gjennom hele feltet. Lokale data danner hele tiden grunnlaget for de ordinære kriginglikningene. Bevegelig naboskap forutsetter selvsagt god datadeknig.

Hvis dataene ikke er stasjonære mht. første moment (forventningen) kan man unngå ikke-stasjonær kriging ved å tilpasse et polygon som beskriver hvordan forventningen endres med posisjonen. Polygonet kan i prinsippet være av en hvilken som helst orden. Er polygonet av 1. orden sier man ofte at polygonet beskriver driften eller trenden i dataene. Differansen mellom de observerte dataene og polygonet dannner "residualet". Residualet kan betraktes som stasjonært og kan dermed danne utgangspunkt for ordinær kriging.

Dersom polygonet tilpasses med minstekvadraters metode har man imidlertid et problem. I minstekvadraters metode forutsetter man at dataene er ukorrelerte. Hvis residualet er korrelert kan estimatet (muligens) forbedres ved kriging. Dette kan fungere i praksis, men man behandler dataene på en inkonsistent måte.

I visse tilfeller kan det tenkes at driften kan modelleres deterministisk (f.eks. ved en grunnvanns-modell). Prosedyren blir den samme som ovenfor. Ordinær kriging anvendes på residualet. Dette er en konsistent men tungvint metode.

Universal kriging

Denne teknikken er som navnet indikerer, en mer generell ikke stasjonær kriging teknikk, men er i den senere tid blitt mer og mer avløst av en annen ikke stasjonær metode som kalles 'indicator kriging'.

La $E[Z]$ variere regelmessig og mulig å representere lokalt med et sett av kjente basis funksjoner.

$$E[Z] = m(x) = a_0 + a_1 X + a_2 Y + a_3 X^2 + a_4 XY + a_5 Y^2 + \ldots = \sum_{k=0}^{n} a_k p^k(x)$$

$\gamma(h)$ er kjent og stasjonær. For ikke å få for ubekvemme uttrykk, la $m(x)$ være beskrevet ved et 1.ordens polynom.

$$m(x) = a_0 + a_1 X + a_2 Y$$

Som i ordinære kriging krever vi at;

$$E[Z_0^* - Z_0] = 0$$

$$E[\sum_{i=1}^{n} \lambda_i Z_i] = E[Z_0]$$

$$\sum_{i=1}^{n} \lambda_i E[Z_i] = E[Z_0]$$

$$\sum_{i=1}^{n} \lambda_i m(x_i) = m(x_0)$$

Innsatt for $m(x)$ gir dette

$$\sum_{i=1}^{n} \lambda_i (a_0 + a_1 X_i + a_2 Y_i) = a_0 + a_1 X_0 + a_2 Y_0$$

For et 1.ordens polygon må altså følgende 3 krav være oppfylt;

$$\Sigma\lambda_i a_0 = a_0$$
$$\Sigma\lambda_i a_1 X_i = X_0$$
$$\Sigma\lambda_i a_2 Y_i = Y_0$$

La $A = \frac{1}{2}E[(Z_0^* - Z_0)^2]$ som før. $B = A$ + tilleggskravene ovenfor multiplisert med hver sin Lagrange multiplikator μ_1, μ_2 og μ_3.

$$B = A + \mu_1(\sum_{i=1}^{n}\lambda_i - 1) + \mu_2(\sum_{i=1}^{n}\lambda_i X_i - X_0) + \mu_3(\sum_{i=1}^{n}\lambda_i Y_i - Y_0)$$

B må nå minimaliseres både m.h.t. λ og μ. $\partial B/\partial\lambda_i = 0 \quad \partial B/\partial\mu_i = 0$.

Dette gir en 6x6 matrise i stedet for en 3x3 matrise som ved ordinær kriging.

$$\begin{bmatrix} 0 & \gamma_{12} & \gamma_{13} & 1 & X_1 & Y_1 \\ \gamma_{21} & 0 & \gamma_{23} & 1 & X_2 & Y_2 \\ \gamma_{31} & \gamma_{32} & 0 & 1 & X_3 & Y_3 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ X_1 & X_2 & X_3 & 0 & 0 & 0 \\ Y_1 & Y_2 & Y_3 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ 1 \\ X_0 \\ Y_0 \end{bmatrix}$$

Variansen i feilestimatet er:

$$\sigma^2_{err} = Var(Z_0^* - Z_0) = \sum_{i=1}^{n}\lambda_i\,\gamma(x_i - x_0) + \mu_1 a_0 + \mu_2 a_1 X_0 + \mu_3 a_2 Y_0$$

Dette er "nice but dangerous". Hvis polygonet er av høyere orden må man ha gode observasjoner og en klar formening om hvilket intervall estimatene bør ligge. Det skal lite til før estimatene "skeier ut".

Oppsummering

Geostatistiskk kan deles inn i 3 separate deler:

Statistiske slutninger består av 3 faser. Første fase består i å se på dataene, d.v.s. undersøke frekvensfordelingen. Desto mer normalfordelt dataene er desto bedre blir estimatene. Vanligvis analyserer man på logaritmen av dataene dersom fordelingen har tyngdepunktet i lave verdier. Andre fase består i å tilpasse en semi-variogram modell til det eksperimentelle semi-variogrammet. Tredje fase består i kryssjekking av semi-variogram modellen.

Estimering består av enkel kriging, ordinær kriging, blokk-kriging, ko-kriging og universal kriging.

Simulering blir ikke berørt nærmere her, men flg. stikkord kan nevnes: i) Simulering av rent stokastisk felt uten romelig korrelasjon mellom dataene. ii) Simulering av korrelert stokastisk felt. iii) Simulering av betinget korrelert stokastisk felt. Punkt i) og iii) er relativt enkelt, mens punkt ii) derimot er langt mer komplisert.

Denne serien utgis av Norges vassdrags- og energiverk (NVE)
Adresse: Postboks 5091 Majorstua, 0301 Oslo

I 1995 ER FØLGENDE NOTATER UTGITT:

Nr 1 IKKE OFFENTLIG

Nr 2 Østen A.Tilrem: Tensiometeret. (21 s.)

Nr 3 Nils-Otto Kitterød: Geostatistikk i geohydrologi. (13 s.)