



# Hantering av geografiska data inom ett jordbruksfält

Kurskompendium

Kurs i precisionssodling, 5 poäng


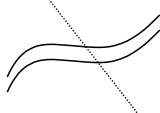
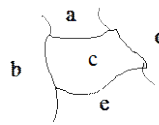
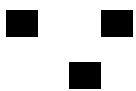
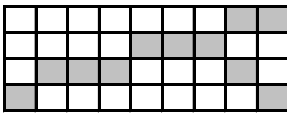
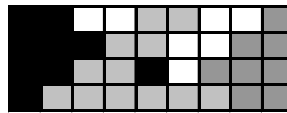
Sofia Delin

---

Institutionen för jordbruksvetenskap Skara  
Avdelningen för mark-växter

Precisionssodling i Väst  
Teknisk Rapport 4  
Skara 2000

provtagit. Av detta kan man endast göra en vektorkarta. Om man vill göra en rasterkarta måste man först interpolera fram värden för varje pixel på hela ytan. Av en digital flygbild kan man däremot göra en rasterkarta direkt från rådata.

	Punkter	Linjer	Ytor
<b>Attributdata i vektor-GIS</b>			
<b>Attributdata i raster-GIS</b>			

Figur 1. Olika sätt att geografiskt presentera attributdata med vektor- och rastermodeller

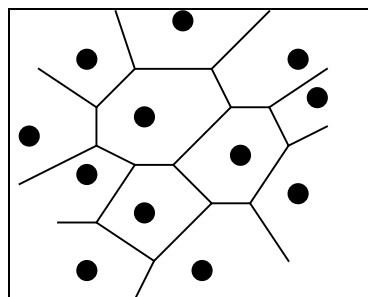
## 2.2 Interpolering

Interpolering är att mellan provpunkter skatta värden på platser som inte är provtagna och på så sätt omvandla data från punktoobservationer till kontinuerliga fält. Beräkning av värden på platser utanför det provtagna området benämns extrapolering. Det finns flera metoder att interpolera mellan mätpunkter.

### 2.2.1 Närmaste granne

En enkel interpoleringsmetod är att ge varje okänd punkt samma värde som närmaste belägna kända punkt. Om mätpunkterna är oregelbundet spridda innebär detta att fältet kommer delas upp i homogena polygoner av olika storlek och form. Dessa polygoner kallas

Thiessenpolygoner (Figur 2 samt Figur 5 a). Övergångarna blir skarpa, och det



Figur 2. Thiessenpolygoner

uppskattade värdet i en viss punkt avgörs endast av det värde som uppmäts i den punkt som råkar ligga närmast. Denna metod är enkel, men i de flesta fall inte den lämpligaste för variabler som förändras gradvis över ytan. Om det gäller kvalitativa data som jordartsklass är det svårt att interpolera, och då kan närmaste granne vara den bästa metoden. Men kvalitativa data används sällan i precisionsodlingssammanhang.

### 2.2.2 Invers distans

Interpolering med *invers distans* (Figur 5 b) ger gradvisa förändringar över ytan. Vid varje punkt beräknas ett medelvärde för omkringliggande punkter och där en viktning sker så att närmre belägna punkter får större betydelse än punkter som ligger längre bort. Varje provtagningspunkt får viktningfaktorn  $1/\text{avståndet}^r$ . Viktningsfaktorn anger hur stor vikt observationen får i det beräknade medelvärdet. Exponenten  $r$  kan ha olika värden. Ju mer man vill framhäva nära belägna punkter, desto högre exponent väljs. Vanligt är att  $r = 2$ . Oftast vill man inte inkludera alla observationer från hela fältet när man beräknar ut ett nytt värde i en punkt. Man anger därför en sökradie. Punkter utanför sökradien exkluderas.

Exempel: Ett värde ska räknas ut i punkten  $x$ . Inom sökradien från punkten  $x$  finns sju observationer.

Observation	Värde ( $v$ )	Avstånd från $x$ ( $a$ )
1	3,2	10
2	3,5	12
3	3,0	14
4	3,7	15
5	2,9	17
6	3,3	20
7	3,6	21

$$x = \frac{\sum(v/a^2)}{\sum(1/a^2)} = (3,2/10^2 + 3,5/12^2 + 3,0/14^2 + 3,7/15^2 + 2,9/17^2 + 3,3/20^2 + 3,6/21^2) / (1/10^2 + 1/12^2 + 1/14^2 + 1/15^2 + 1/17^2 + 1/20^2 + 1/21^2) = 3,3$$

Invers distans är en enkel metod som är vanlig inom precisionsodling, både vid framställning av skördekartor och markkartor. Ofta är det fullt tillräckligt att använda denna metod, särskilt om man har gjort täta observationer. Men för att veta om och hur variationerna är rumsligt korrelerade, så att interpoleringen kan utföras på bästa sätt, bör man analysera sina data geostatistiskt.

## 2.2.3 Geostatistik

Intuitivt tror nog de flesta av oss att värdet av en variabel i en punkt sannolikt är mer lik värdet av samma variabel i en närliggande punkt, än i en punkt som ligger längre bort. Om detta är fallet finns det ett rumsligt samband. Detta rumsliga samband kan beräknas med hjälp av geostatistik och användas vid interpolering. Viktningen av olika observationer med skilda avstånd och riktning från punkten där ett värde ska beräknas grundas då på hur det ser ut statistiskt på det aktuella fältet. Geostatistiska metoder för interpolering delar in den rumsliga variationen i tre komponenter: *deterministisk variation* som kan förklaras av någon annan variabel, *rumsligt korrelerad men fysiskt oförklarlig variation* och *okorrelerat brus*.

Deterministisk variation kan t.ex. bero på höjdskillnader. Den rumsligt korrelerade men fysiskt oförklarliga variationen vet man inte den direkta orsaken till, men punkter som ligger närmare varandra tenderar till att vara mer lika än punkter som ligger längre ifrån varandra. Okorrelerat brus är variation som uppträder helt slumpvis i rummet och som därför inte går att förutsäga. Ibland kan den senare variationen vara rumsligt korrelerad, men i en mycket mindre skala än den man fokuserar på. Den uppfattas därför som brus. Bruset kan också bero på t.ex. mätosäkerhet vid kemisk analys.

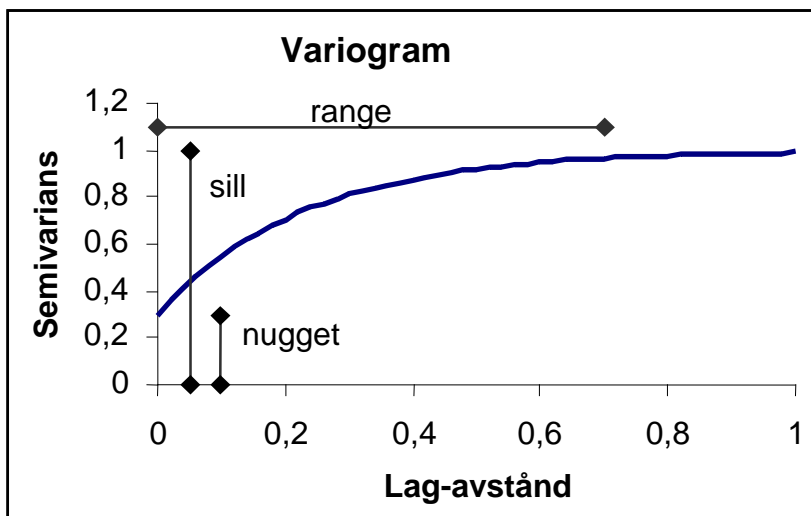
### 2.2.3.1 Grundläggande statistik och trendanalys

Innan man startar att analysera sina data geostatistiskt bör man kontrollera att data är normalfördelade, att det inte finns några observationer som avviker extremt samt om det finns någon geografisk trend. Detta är saker som kan störa och försvåra den geostatistiska analysen. Om data inte är normalfördelade kan man behöva göra en transformation, t.ex. logaritmera värdena. Om de är binomialfördelade, d.v.s. har två toppar som många observationer hamnar runt i ett histogram, kan det vara aktuellt att dela in fältet i två delar (se vidare 3.1.1 Klassificering före interpolering). Om det förekommer observationer som avviker extremt från resten, kan det vara lämpligt att plocka bort dem under den inledande geostatistiska analysen. Sedan kan man ta tillbaka dem inför interpoleringen. Med trend menas att det föreligger en genomgående förändring längs med hela fältet i någon riktning. Om lerhalten exempelvis ökar linjärt med avståndet från ena fältkanten i  $x$ -led finns det en trend. Denna trend beskrivs bättre med en vanlig regressionsanalys mellan lerhalt och läge i  $x$ -led. Den bör tas bort före geostatistisk analys och kan tas tillbaka efter interpolering. Den geostatistiska

analysen och interpoleringen görs i det fallet på residualerna (observationernas avvikelser från regressionslinjen) istället för på rådata.

### 2.2.3.2 Variogram

Den rumsliga korrelationen inom ett fält kan beskrivas i ett s.k. semivariogram (ofta används denna term synonymt med termen ”variogram”) (Figur 3). Semivarians är variansen mellan två punkter separerade av ett visst avstånd. Detta avstånd kallas *lag*. I semivariogrammet plottas semivariansen mot detta avstånd. Medelvärden av semivariansen hos alla par som ligger på ungefär samma avstånd mellan varandra bildar en punkt i variogrammet. Om ett rumsligt samband föreligger är variansen större mellan provpunkter som ligger längre bort ifrån varandra. Variansen mellan provpunkter som ligger oändligt nära varandra beror bara på brus i form av provtagnings- och mätfel. Denna varians kallas *nugget* och kan avläsas där



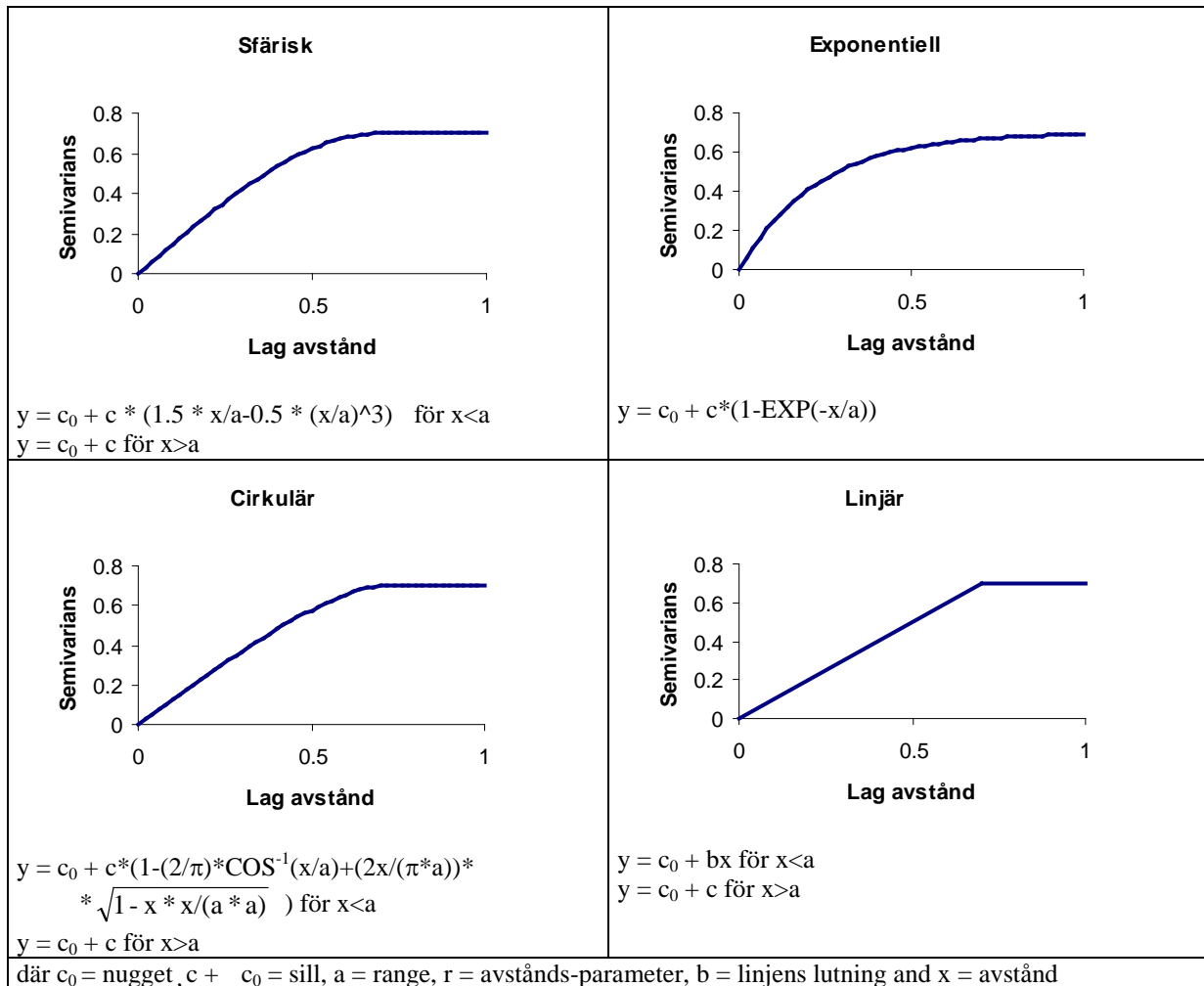
linjen i variogrammet korsar y-axeln. Vid ett visst avstånd slutar variansen att öka. Variansen vid detta avstånd kallas *sill*. Avståndet kallas *range* eller *räckvidd*. Inom räckvidden är punkterna sannolikt mer lika ju närmare varandra de ligger.

Figur 3. Exempel på ett variogram

Vid upprättandet av ett variogram utifrån observationer i ett dataset, krävs det ett stort antal observationer för att få ett stabilt variogram. Olika forskare har olika tumregler för hur många observationer som krävs, men omkring 50 observationer kan räcka i bästa fall. Ett medelvärde för semivariansen räknas ut för par separerade av ungefär samma avstånd, och det krävs tillräckligt med par för varje lag för att få ut ett vettigt medelvärde. För få par per lag ger en ”hoppig” kurva.

Punkterna i variogrammet benämns det experimentellt variogrammet. Till det anpassas en modell (i form av en kurva), vilket benämns modellvariogrammet. Det finns flera modeller med olika form (Figur 4). Vilken modell man ska använda grundar man i första hand på hur modellen passar in på punkterna rent visuellt. I andra hand ser man på vilken modell som ger minsta kvadratsumma på residualerna. Man föredrar oftast en enklare modell framför en mer avancerad.

Ibland ser det geostatistiska sambandet olika ut i olika riktningar. En punkt kan t.ex. tendera att vara mer lik sina grannar i nordsydlig riktning än i östvästlig. Detta kallas *anisotropi*. Genom att modellera fram variogram för olika riktningar kan man upptäcka om det finns anisotropi.



**Figur 4. Några modeller för variogram**

### 2.2.3.3 Kriging

För att uppskatta värden av en variabel på punkter, där inga observationer gjorts, kan den geostatistiska interpoleringsmetoden *kriging* användas (figur 5 c). Kriging är liksom invers distans en metod med vilken man viktar omkringliggande värden till ett medelvärde. Vikterna beror av variogrammet och configurationen av mätpunkter. Det finns olika typer av kriging. I punktkriging uppskattar man värdet i en punkt, medan man i blockkriging uppskattar medelvärdet i ett område av en viss storlek runt punkten. Det uppskattade värdet,  $\hat{Z}(x_0)$ , i en punkt uttrycks i följande ekvation:

$$\hat{Z}(x_0) = \sum_{i=1}^N \lambda_i z(x_i)$$

där  $\lambda_i$  är vikterna och  $z(x_i)$  är rådata. Summan av vikterna är lika med ett. Mer vikt ges åt en mätpunkt nära den punkt där uppskattning ska beräknas. Mindre vikt ges till varje mätpunkt om flera punkter är grupperade tätt tillsammans. På så vis blir det de punkter som ligger närmast i varje riktning som påverkar uppskattningen. Antalet mätpunkter som inkluderas när uppskattningen beräknas bör inte vara för få och inte heller onödigt många. För få punkter inom sökradien ger ett dåligt medelvärde, medan många punkter inkluderar även de med en vikt nära noll och därför mycket liten betydelse för resultatet. Sökradien bör vara ungefär lika med räckvidden (dvs det avstånd inom vilket det föreligger ett statistiskt samband).

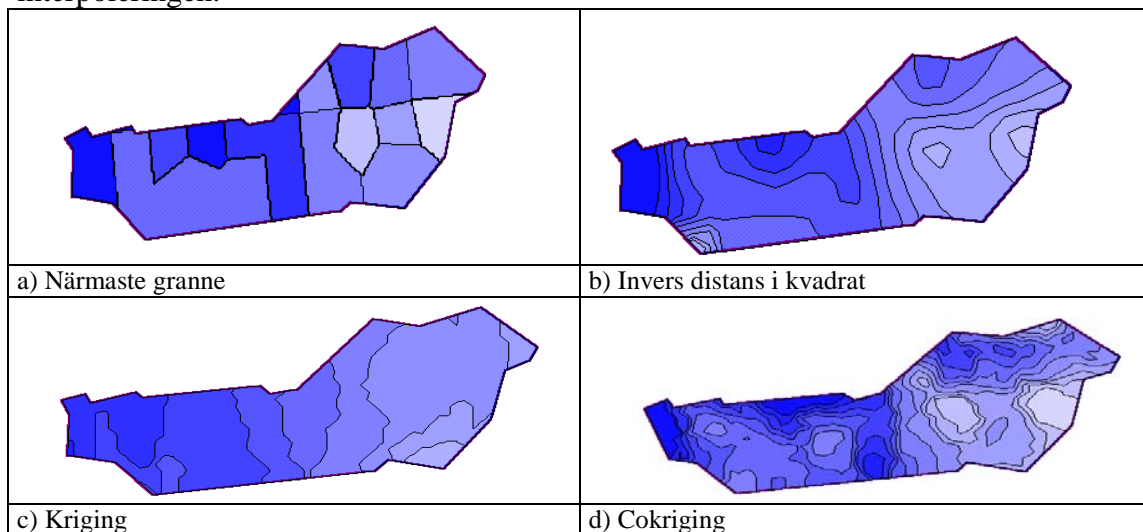
Om anisotropi föreligger, d.v.s. om sambandet ser olika ut i olika riktningar, kan man ta hänsyn till detta vid kriging. Större vikt ges då åt de observationer som ligger i de riktningar där semivariansen ökar minst med avståndet.

Det finns flera datorprogram för variogram, modellering och kriging. Några exempel är GeoEas, GS+ och Genstat. Även i en del GIS-program finns kriging, men ofta har man små möjligheter att påverka modelleringen av variogram i dessa program. Vanligt är att man gör geostatistiken i ett specialiserat geostatistikprogram och sedan för över interpolerade data till ett GIS-program för presentation och eventuella vidare geografiska analyser.

En fördel med kriging framför invers distans är att hänsyn tas till mätpunkternas konfiguration, så att inte tätt provtagna områden får större vikt. En annan fördel är att man har möjlighet att ta hänsyn till om det geografiska sambandet varierar i olika riktningar. Dessutom tar man hänsyn till en uppmätt räckvidd istället för en tillhöftad och får en mer välgrundad viktning med avseende på avstånd. Nackdelen med kriging är att den kräver mer arbete, kunskap och relativt avancerade datorprogram. I fall där man har tätt mellan provtagningspunkterna och jämn spridning på dem finns inte så mycket att tjäna på kriging jämfört med invers distans.

### 2.2.3.3 Cokriging

Om man har gles data av en variabel, men tätare med data av en annan variabel (kallad *covariabel*) som varierar enligt samma mönster, kan man använda sig av en metod som kallas *cokriging* (Figur 5 d). Då styr till viss del observationerna och variogrammet för covariabeln interpoleringen.



**Figur 5.** Lerkartor framställda med olika interpoleringsmetoder från ett dataset med 20 provpunkter analyserade med avseende på lerhalt, samt vid cokriging, ca 480 punkter med bestämning av elektriskt ledningstal. Anledningen till att kartan som framställts genom kriging här är ganska utjämnad är att nugget är ganska hög. Med en nugget nära noll hade höga och låga värden framträtt på samma sätt som med de andra interpoleringsmetoderna.