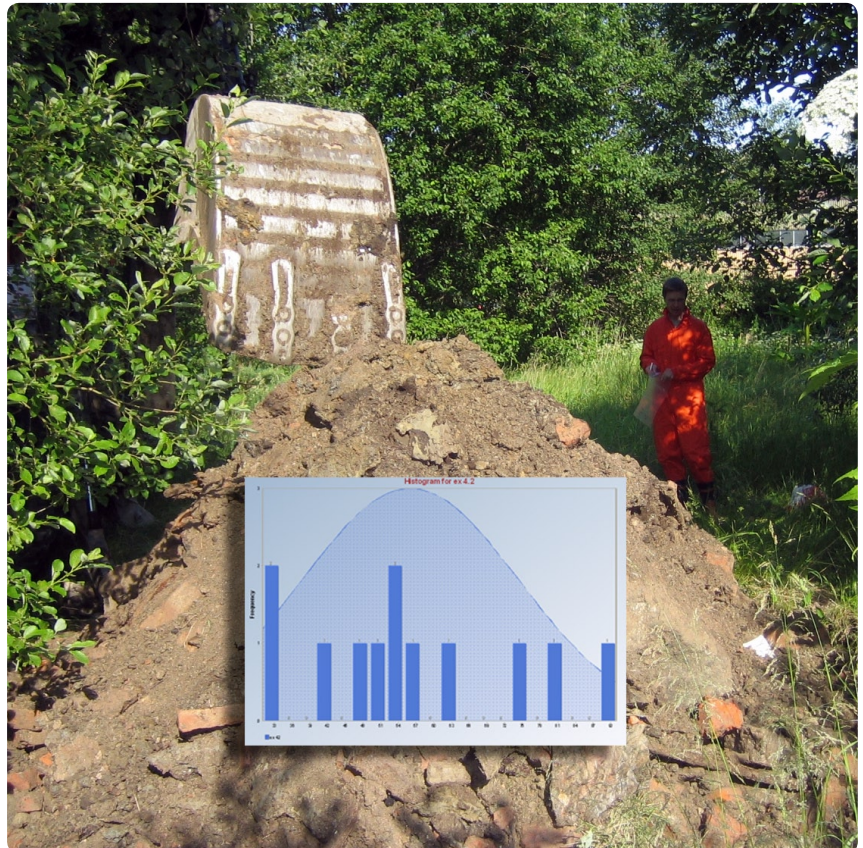


Metodik för statistisk utvärdering av miljötekniska undersökningar i jord

RAPPORT 5932 • JULI 2009



Kunskapsprogrammet



Metodik för statistisk utvärdering av miljötekniska undersökningar i jord

Jenny Norrman, Statens Geotekniska Institut (SGI)
Tom Purucker, US Environmental Protection Agency
Pär-Erik Back, SWECO
Fredric Engelke, SGI
Robert Stewart, University of Tennessee

Beställningar

Ordertel: 08-505 933 40

Orderfax: 08-505 933 99

E-post: natur@cm.se

Postadress: CM Gruppen AB, Box 110 93, 161 11 Bromma

Internet: www.naturvardsverket.se/bokhandeln

Naturvårdsverket

Tel: 08-698 10 00, fax: 08-20 29 25

E-post: registrator@naturvardsverket.se

Postadress: Naturvårdsverket, SE-106 48 Stockholm

Internet: www.naturvardsverket.se

ISBN 91-620-5932-3

ISSN 0282-7298

© Naturvårdsverket 2009

Tryck: CM Gruppen AB, Bromma 2009

Omslagsfoto: Stor bild: Ola Arvidslund, SGI, Liten bild: Göran Karlsson, SGI



Förord

Ett av riksdagens miljömål är Giftfri miljö, och i detta mål ingår att efterbehandla och sanera förorenade områden. Brist på kunskap om risker med förorenade områden och hur de bör hanteras har identifierats som hinder för ett effektivt saneringsarbete. Naturvårdsverket har därför initierat kunskapsprogrammet Hållbar Sanering.

Denna rapport redovisar projektet ”Metodik för statistisk utvärdering av miljötekniska undersökningar i jord” som har genomförts inom Hållbar Sanering. Arbetet i föreliggande rapport har utförts vid Statens Geotekniska Institut (SGI) i samarbete med US Environmental Protection Agency (US EPA), SWECO och University of Tennessee (UT). Rapporten har författats av Jenny Norrman (SGI), Tom Purucker (US EPA), Pär-Erik Back (SWECO), Fredric Engelke (SGI) och Robert Stewart (UT).

Under arbetets gång genomfördes ett seminarium med Anders Bank (Envipro Miljöteknik), Tommy Norberg (Matematiska vetenskaper, Chalmers och Göteborgs Universitet), Lars Rosén (kompetenscentrat FRIST, Chalmers tekniska högskola), Sonja Blom (FB Engineering), Bo Svensson (Linköpings Universitet) och Tomas von Kronhelm (SAKAB och kontaktperson för Hållbar Sanering), vilket gav projektgruppen viktiga kommentarer för det fortsatta arbetet.

Under tiden som arbetet pågick med rapporten arbetades nya riktvärden fram varför de riktvärden som används i exemplen inte alltid stämmer med de nu gällande. Detta påverkar dock inte principerna för beräkningarna på något sätt.

Rapporten har granskats och kommenterats av Lars Rosén, Mattias Bäckström (Örebro Universitet), Thomas von Kronhelm, Mikael Stark (SGI) samt Elisabeth Hansson (FB Engineering). Författarna tackar alla som bidragit med synpunkter och kommentarer på rapporten. Tomas von Kronhelm vid SAKAB var kontaktperson för Hållbar Sanering.

Naturvårdsverket har inte tagit ställning till innehållet i rapporten. Författarna svarar ensamma för innehåll, slutsatser och eventuella rekommendationer.

Naturvårdsverket juli 2009

Innehåll

FÖRORD		3
SAMMANFATTNING		9
SUMMARY		11
1	INLEDNING	13
1.1	Bakgrund	13
1.2	Syfte	13
1.3	Avgränsning och upplägg på rapporten	14
1.4	Utvärdering av mätdata i efterbehandlingsprojekt	14
2	RAMVERK FÖR STATISTISK UTVÄRDERING AV DATA	17
2.1	Steg 1 Bedömning av föroreningsgrad	19
2.2	Steg 2 Bedömning av andelen förorenade massor	20
2.3	Steg 3 Bedömning av rumslig korrelation	20
2.4	Steg 4 Interpolation	21
2.5	Åtgärds mål och val av efterbehandlingsvolym	22
3	STEG 1: BEDÖMNING AV FÖRORENINGSGRAD	24
3.1	Beskrivning av stickprovet	24
3.1.1	Beskrivande statistik	25
3.1.2	Parametriska och icke-parametriska metoder för att beskrivamålpopulationen	27
3.2	Representativ halt	28
3.3	Referenshalter att jämföra mot	30
3.3.1	Riktvärden	30
3.3.2	Jämförvärde för bakgrundshalt	31
3.3.3	Referenshalt för akuttoxicitet	32
3.4	Metoder för jämförelser mellan stickprov och referenshalt	33
3.4.1	Konfidensintervall för medelhalten	34
3.4.2	Hypotestester	37
3.5	Akuttoxiska föroreningar	43
3.6	Bedömning av sannolikheten att missa en hotspot	44
4	STEG 2: BEDÖMNING AV ANDEL FÖRORENADE MASSOR	45
4.1	Bedömning av andel med normalfördelningsplot	45
4.1.1	Metodik	46
4.2	Bedömning av andel med statistisk fördelning	47
4.3	Bedömning av andel med betafördelning	49
4.3.1	Betafördelningen	49
4.3.2	Användning av betafördelningen	51
4.3.3	Att inkludera förhandskunskap i en betafördelning	52

5	STEG 3: BEDÖMNING AV RUMSLIG KORRELATION	55
5.1	Hänsyn till förekomst av rumslig korrelation vid utvärdering av data	55
5.2	Metoder för bedömning av rumslig korrelation	56
5.2.1	Experimentella variogram	56
5.2.2	Korrelogram	57
5.2.3	Konfidensintervall och signifikanstester	57
5.2.4	Manteltest för rumslig korrelation	59
5.2.5	Stationaritetstester	59
6	STEG 4: INTERPOLATION	60
6.1	Interpolationsmetoder	60
6.1.1	Variogrammodellen	60
6.1.2	Kriging	62
6.1.3	Geostatistisk simulering	63
6.2	Modellvalidering	64
6.3	Strategi för kompletterande provtagning	64
7	PRAKTISKA ASPEKTER VID UTVÄRDERING AV DATA	66
7.1	Statistiska populationer	66
7.2	Hantering av datakluster	67
7.3	Hantering av olika typer av data	67
7.4	Data under detektionsgränsen	69
7.5	Duplikat	70
7.6	Outliers	70
7.7	Data med olika support	71
7.8	Data från olika provtagningsdjup	73
7.9	Projekt med få data	75
7.10	Data med osäker representativitet	75
7.11	Kvalitetskontroll och rimlighetsbedömningar	76
8	FALLSTUDIE MED BERÄKNINGSEXEMPEL	77
8.1	Områdesbeskrivning för fallstudien	77
8.2	Provtagning och analysresultat	78
8.2.1	Provtagningens syfte	78
8.2.2	Önskad säkerhet	78
8.2.3	Provtagning och analysresultat	79
8.3	Steg 1: Bedömning av föroreningsgrad	79
8.3.1	Beskrivande statistik	79
8.3.2	Goodness-of-fit test	80
8.3.3	Beräkning av UCLM95	81
8.3.4	Hypotestest: stickprov mot riktvärde	81
8.3.5	Bedömning av andel som överskrider referenshalt för akuttoxicitet	82
8.3.6	Sannolikheten för okänd hotspot	82

8.4	Steg 2. Bedömning av andelen förorenade massor	83
8.4.1	Bedömning av andel med normalfördelningsplot	83
8.4.2	Bedömning av andel med statistisk fördelning	84
8.4.3	Bedömning av andel med betafördelning	84
8.5	Steg 3. Bedömning av rumslig korrelation	85
8.6	Steg 4. Interpolation	86
9	DISKUSSION	88
10	REFERENSER	89
	BILAGA A. STATISTISK PROGRAMVARA	92
	ProUCL	92
	SADA – Spatial Analysis and Decision Assistance	92
	VSP – Visual Sample Plan	92
	BILAGA B. ORDLISTA	93
	BILAGA C. BETAFÖRDELNINGEN	99
	Referenser	99

Sammanfattning

Utvärdering och presentation av en miljöteknisk undersökning i ett förorenat område är ett underlag för de riskbedömningar och beslut om eventuella åtgärder som behöver göras. Utvärdering av insamlad mätdata sker ofta med ett stort inslag av subjektivitet. Subjektiviteten, eller snarare expertkunskapen och expertbedömningarna, är ett viktigt redskap i miljöundersökningar. För att kunna analysera och kvantifiera osäkerheterna i olika mätdata är det däremot nödvändigt att använda statistiska utvärderingsmetoder. Styrkan med statistik är att kunna skapa ett beslutsunderlag som har stöd i statistiska analyser och där beslutsfattaren blir informerad om vilka osäkerheter som föreligger. Vidare kan statistiska metoder ge information som annars inte framkommer, dvs. den statistiska analysen tillför ett mervärde.

Syftet med projektet har varit att ta fram ett ramverk för att strukturera statistisk utvärdering av tillgänglig information från förorenade områden, både förhandskunskap och insamlad data. Syftet med ramverket är att lyfta fram statistiska metoder som kan användas för vissa delar i en riskbedömning, men också som underlag för kostnadsuppskattningar i en åtgärdsutredning. Detta innebär att ramverket inte gör anspråk på att vara ett komplett arbetsschema för utredningar av förorenade områden, utan snarare en hjälp till att förstå var olika statistiska metoder kan komma till användning.

Projektet har inriktats på karaktärisering av föroreningskällan med avseende på föroreningsgrad och föroreningens rumsliga utbredning. Tillämpningarna och exemplen i rapporten hanterar endast föroreningshalter i jord, men metoderna kan användas för andra medier och andra typer av storheter och mätvärden. Ramverket omfattar en rad metoder som kan användas för att utvärdera ett enskilt egenskapsområde, dvs. ett område där föroreningen är genererad genom samma process och med relativt homogena egenskaper. De fyra huvudstegen i ramverket är:

- 1) Bedömning av föroreningsgrad,
- 2) Bedömning av andelen förorenade massor,
- 3) Bedömning av rumslig korrelation, samt
- 4) Interpolation.

Det första steget i ramverket syftar till att göra en bedömning om hälso- och miljöriskerna är acceptabla på basis av data från jordprover. Här föreslås ett antal moment med detta syfte: i) bedömning genom att jämföra en representativ halt för området med en referenshalt (långtidsrisker), ii) bedömning av hur stor andel av området som överskrider en referenshalt för akuttoxicitet (akuta risker), samt iii) bedömning av sannolikheten att provtagningen har missat en *hotspot* av en viss storlek.

Det andra steget är att göra en bedömning av andelen förorenade massor som har halter över riktvärdet. Andelen förorenade massor inom ett område är intressant både för riskbedömning, med avseende på bl.a. föroreningsmängder, och för kostnadsbedömningar i åtgärdsutredningen. Skattningen kan

användas för att avgöra om endast en del av området skall åtgärdas och om den delen behöver avgränsas, eller om hela området behöver åtgärdas. Om hela området behöver åtgärdas kan det vara mer kostnadseffektivt att direkt åtgärda utan detaljerad avgränsning. Om däremot endast en mindre del av området behöver åtgärdas är det viktigt att avgränsa dessa delområden/delvolymer.

Det tredje steget är att bedöma om det finns någon rumslig korrelation i området, dvs. om det finns något släktskap i halter mellan närliggande provpunkter. Syftet med detta är att kunna bedöma om geostatistisk interpolation bör/kan användas för att avgränsa områden med förhöjda halter. Om ingen rumslig korrelation förekommer i området, eller om korrelationslängden är alltför kort, är det olämpligt att gå vidare med geostatistiska metoder.

Det sista och fjärde steget utförs om det finns en relevant rumslig korrelation. Detta steg innebär att geostatistisk interpolation utförs med syfte att avgränsa områden med förhöjda föroreningshalter. För att interpolation skall kunna utföras och generera ett rimligt resultat måste interpolationsmetod väljas, modellen valideras och eventuell kompletterande provtagning utföras. Kartorna som genereras kan sedan användas som grund för att undersöka beslutsosäkerheter kopplade till utformning av åtgärder, volymer och kostnader.

I denna rapport ligger fokus på att kvantifiera typ I-fel, dvs. risken att felaktigt friskriva ett område som egentligen kräver efterbehandling. Det kan dock vara så att man istället vill kvantifiera risken att ett område efterbehandlas i onödan, dvs. typ II-fel. Valet har att göra med vem som är beslutsfattare, vilka kostnader och nyttor som är förknippade med en sanering och hur dessa värderas, samt vilka miljö- och hälsorisker som finns. Valet av representativ halt för det område man skall utreda är en viktig del i utredningen.

I rapporten beskrivs betafördelningen som ett verktyg där subjektiv kunskap formellt kan läggas in i en statistisk analys. Anledningen till att modeller som kan ta hänsyn till subjektiva data är ett intressant utvecklingsområde är att de personer som arbetar med utredningar av förorenade områden ofta besitter stor erfarenhet och expertkunskap som inte tillvaratas i klassiska statistiska analyser. Kombinationen av expertkunskap och klassisk statistik skulle kunna resultera i mycket effektivare verktyg än de som idag används i branschen.

För att statistiska metoder och verktyg skall börja användas i större utsträckning än idag krävs en allmän höjning av kunskapsnivån i Sverige inom branschen. En ökad kunskap kan bidra till att det blir lättare att kommunicera statistiska analyser mellan t.ex. konsulter, problemägare och tillsynsmyndigheter.

Summary

The evaluation and presentation of an environmental investigation at a contaminated site is a basis for the risk assessment and the decisions about possible remediation needed. The evaluation of collected data is often made with a large portion of subjectivity. This subjectivity, or rather this expert knowledge, is an important tool in environmental investigations. However, in order to analyse and quantify the uncertainties associated with data it is necessary to use statistical methods. The strength of using statistical methods is to be able to create a decision basis which is supported by statistical analyses and in which the decision-maker is informed about the existing uncertainties.

The aim of this project has been to develop a framework to structure the statistical evaluation of existing information from contaminated sites, both as soft and hard information. The aim of the framework is to lift forward statistical methods which can be used for some different parts of a risk assessment, but also for cost estimates and delimitation of highly contaminated areas. This means that the framework does not claim to be a complete working scheme for the assessment of contaminated sites, but rather a help to understand where different statistical methods can be useful.

The work has been limited to characterization of the contamination source regarding the degree of contamination and the spatial distribution of the contamination source. The application and the examples in the report do only handle contamination levels in soil, but the methods can be used for other media and for other types of variables. The framework suggests a number of methods for evaluating an area which is representing a single population, i.e. an area in which the contamination is generated by the same process and with relatively homogeneous geology. The four main steps of the framework are:

- 1) Estimation of the degree of contamination,
- 2) Estimation of the proportion of the site being contaminated,
- 3) Estimation of the spatial correlation,
- 4) Interpolation.

The first step in the framework aims at evaluating whether the health and environmental risks in the area are acceptable based only upon soil data. A number of analyses are suggested: (i) evaluation by comparing the site data with a reference concentration, (ii) estimation of the proportion of the site being contaminated above acute toxicity levels, and (iii) an evaluation of the probability to miss a hotspot of a certain size and shape given the sampling design used in the area.

The second step is to make an estimation of the proportion of the area being contaminated above a reference concentration. This proportion is interesting both for the risk assessment regarding estimations of the total amount of contaminants, and for an estimate of the remediation costs associated with this amount. The estimate of the proportion can be used to decide whether only a part of the area needs remediation and if this part needs delimitation,

or if the whole area needs to be remediated. If the whole area needs to be remediated, it can be more cost-effective to remediate without delimitating exactly where the highly contaminated areas are. On the other hand, if only a small proportion of the area needs remediation, it is of importance to delimitate those subareas.

The third step involves estimating if any spatial correlation is present at the site. The aim is to carefully consider whether geostatistical interpolation techniques should be used to delimitate areas in need of remediation. If no spatial correlation can be assumed to be present at the site, it is unsuitable to use geostatistical methods since they do assume that spatial correlation exists.

In the last and fourth step, which is carried out only if spatial correlation is assumed to be present, geostatistical interpolation is suggested for delimitating areas with high contamination levels before the remediation phase of the project. For an interpolation to be successful and generate reasonable results, the proper method should be chosen, the model should be validated and secondary sampling should be considered. The maps generated can then be used as a basis for evaluating uncertainties related to remedial design, volumes and costs.

In this report, focus has primarily been on quantifying the type I-error, i.e. the risk of wrongly leaving an area unremediated. However, one may wish to quantify the risk of unnecessary remediation at a site instead, i.e. the type II-error. This choice is depending on who the decision maker is, the costs and benefits are associated with a remediation and how these are valued, and the environmental and human health risks present at the site. The choice of a representative concentration at the site is an important part of the evaluation and a Swedish guidance for this choice would be useful for the evaluation of the degree of contamination.

In the report, the beta distribution is described as a tool where expert knowledge can be included in a formal statistical analysis. The reason that models which can take into account subjective information is an interesting development area, is that the persons usually working with evaluation of contaminated sites typically have a large portion of experience and expert knowledge which is not considered in classical statistics. The combination of expert knowledge and classical statistical tools could result in much more efficient tools than those used in the branch today.

For statistical methods and tools to be used to a larger extent than today, a general raise of the level of knowledge in the branch in Sweden is required. An increased level of knowledge can contribute to that statistical analyses can be communicated easier among e.g. consultants, problem owners and authorities.

1 Inledning

1.1 Bakgrund

Utvärdering och presentation av en miljöteknisk markundersökning i ett förorenat område ger underlag för de riskbedömningar som krävs och de beslut om eventuella åtgärder som behöver tas. Utvärdering av insamlad mätdata sker ofta med ett stort inslag av subjektivitet. Subjektiviteten, eller snarare expertkunskapen och expertbedömningarna, är ett viktigt redskap i miljöundersökningar. För att kunna analysera och kvantifiera osäkerheterna i olika mätdata är det däremot nödvändigt att använda statistiska utvärderingsmetoder. Styrkan med statistik är att kunna skapa ett beslutsunderlag som har stöd i statistiska analyser och där beslutsfattaren blir informerad om vilka osäkerheter som föreligger. Vidare kan statistiska metoder ge information som annars inte framkommer, dvs. den statistiska analysen tillför ett mervärde.

Utvärdering av förorenade områden bör baseras på all tillgänglig kunskap om området, både mät- och analysresultat (hårda data) och förhandsinformation som t.ex. inventeringsresultat, erfarenheter från andra områden och verksamhetshistorik (mjuka data). För att utvärderingen skall leda fram till en så realistisk bedömning som möjligt av föroreningshalter, rumslig utbredning och mängd förorening krävs att:

1. utvärderingen anpassas till syftet med utredningen,
2. osäkerheter redovisas och om möjligt kvantifieras,
3. redovisningen följer formella ramar, och
4. presentationen av analyser och slutsatser sker på ett transparent och spårbart sätt.

1.2 Syfte

Syftet med projektet har varit att ta fram ett ramverk för att strukturera statistisk utvärdering av tillgänglig information från förorenade områden, både för förhandskunskap och insamlad data. Ett sådant ramverk syftar till att lyfta fram statistiska metoder vid karaktärisering av föroreningssituationen i ett område. Vissa delar kan även vara till nytta i en riskbedömning och för kostnadsuppskattningar i en åtgärdsutredning. Detta innebär att ramverket inte gör anspråk på att vara ett komplett arbetsschema för utredningar av förorenade områden, utan snarare en hjälp att förstå var olika statistiska metoder kan komma till användning.

Ramverket tillämpas i första hand när ett antal mät- och analysresultat finns tillhanda och det kan användas både i översiktliga och mer omfattande undersökningar.

1.3 Avgränsning och upplägg på rapporten

Projektet har inriktat sig på karaktärisering av föroreningskällan med avseende på föroreningsgrad och föroreningens rumsliga utbredning. Tillämpningar och exempel i rapporten hanterar endast föroreningshalter i jord, men metoderna kan användas för andra medier och andra typer av storheter än koncentrationer.

Rapporten är uppdelad i följande kapitel:

- Kapitel 1 är en introduktion som beskriver syfte och avgränsning och har som mål att sätta in rapporten i rätt sammanhang.
- Kapitel 2 innehåller en översiktlig beskrivning av strukturen för de fyra stegen i ramverket.
- I efterföljande fyra kapitel 3, 4, 5 och 6 beskrivs respektive steg i ramverket med mer utförliga beskrivningar av de statistiska metoderna samt några exempel.
- Kapitel 7 går igenom vanliga frågeställningar som kan uppkomma när man utvärderar data, t.ex. hur hanteras datakluster, data från olika utredningar och mätningar som ligger under detektionsgränsen?
- Kapitel 8 är en redovisning av en fallstudie där flera av metoderna som beskrivits tidigare i rapporten tillämpas.
- Kapitel 9 är en kortfattad diskussion.

Ett antal bilagor finns också i rapporten:

- Bilaga A beskriver kortfattat några av de gratisprogramvaror som använts för beräkningsexemplen.
- Bilaga B är en ordlista där viktiga nyckelbegrepp förklaras.
- Bilaga C ger en mer detaljerad beskrivning av betafördelningen.

1.4 Utvärdering av mätdata i efterbehandlingsprojekt

Efterbehandlingsprojekt utförs vanligen i ett antal steg där man startar med ett obefintligt eller litet dataunderlag och successivt arbetar sig från till ett mer omfattande underlag. Syftet är att kunna ge säkrare bedömningar ju längre i processen projektet har kommit. Riskbedömningar av förorenade områden, både statligt finansierade utredningar och andra, utförs vanligen stegvis: från riskklassning i MIFO fas 1 till en fördjupad riskbedömning. Det är dock inte alltid man går vidare från en förenklad riskbedömning till en fördjupad och ibland utförs en fördjupad riskbedömning direkt; arbetsgången varierar mellan olika projekt. Riskbedömningen ligger till grund för bedömningen av åtgärdsbehov och behovet av kompletterande undersökningar om osäkerheterna är för stora.

Grovt sett kan de beslut som krävs i ett efterbehandlingsprojekt delas in i tre principiella kategorier:

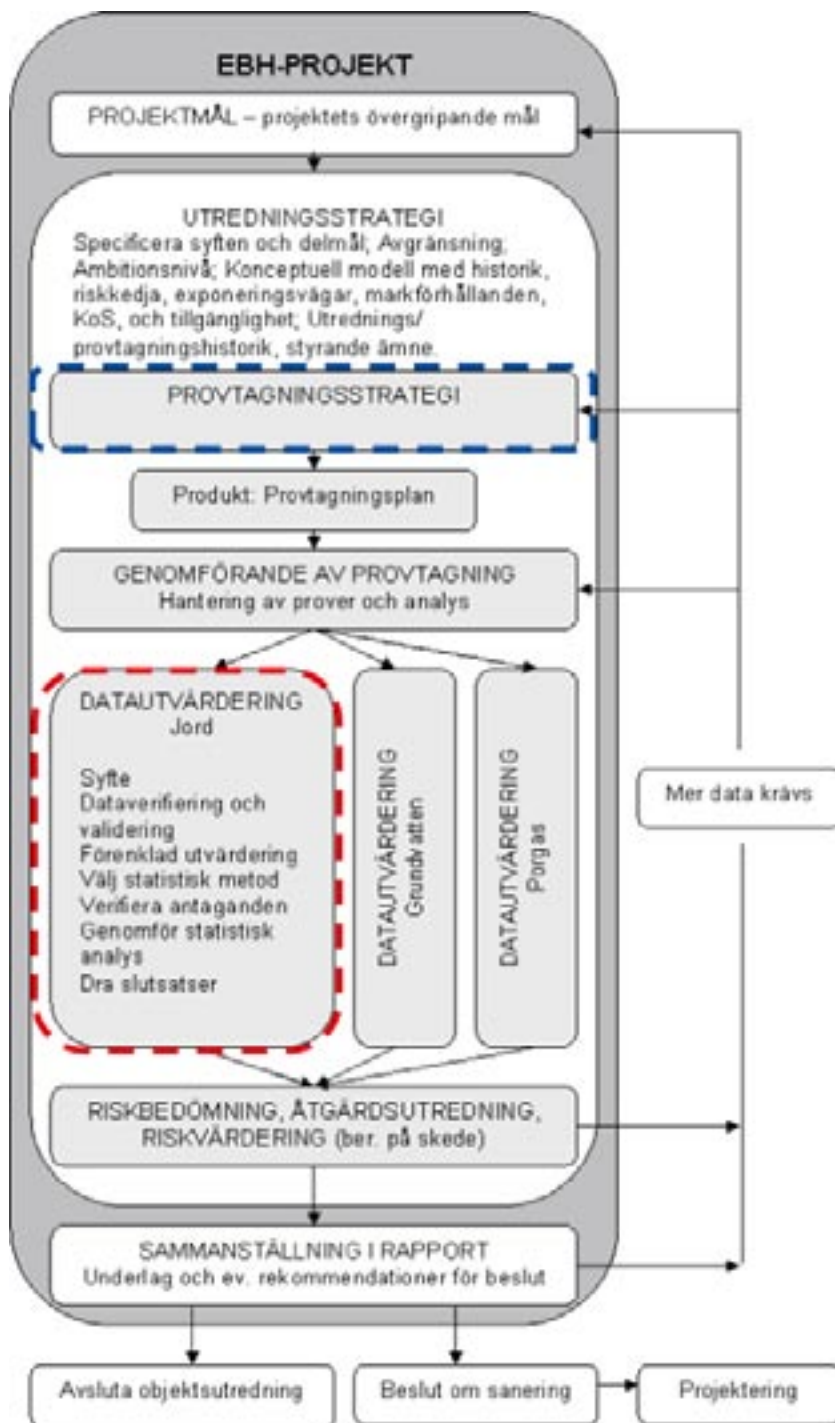
1. Oacceptabelt stor risk föreligger och ett beslut om åtgärder krävs.
2. Risken bedöms vara acceptabel och inga ytterligare åtgärder är nödvändiga.
3. Osäkerheten är för stor och mer data krävs för att beslut skall kunna tas.

Om beslutet faller inom kategori 1 eller 2 riskerar man att begå ett fel, dvs. (i) att utföra efterbehandling som inte är nödvändig och därmed ta onödiga kostnader, eller (ii) att felaktigt friskriva ett område som egentligen utgör en risk för hälsa och/eller miljö. I viss mån kan även kategori 3 vara ”fel beslut” – att samla in mer data kanske inte kommer att minska osäkerheten nämnvärt utan bara medföra kostnader.

Utvärdering av analys- och mätdata utgör en viktig del i det underlag som ligger till grund för att fatta beslut om hantering av ett förorenat område. Utvärderingen av insamlad data föregås av att definiera projektmål, framtagning av en konceptuell modell, framtagning av en provtagningsstrategi som omsätts i en provtagningsplan samt själva provtagnings utförande, se Figur 1-1. I denna rapport hanteras endast utvärdering av data från jordprovtagning, vilket är markerat med rött i figuren. En arbetsmetodik för hur en provtagningsstrategi för jord kan planeras beskrivs i rapporten *Provtagningsstrategier för förorenad jord* (Norrman et al. 2009) och är markerat med blått i figuren.

Mätdata används för olika syften i efterbehandlingsprojekt. De vanligaste är att bedöma om området är förorenat eller ej, utreda typ av förorening, avgränsa eventuella föroreningar och/eller genomföra riskbedömningar. Olika frågeställningar ställer olika krav på hur data skall utvärderas:

- Avgränsning av förorenade massor. Avgränsning av förorenade massor kan göras genom interpolation under förutsättning att det rumsliga sambandet mellan olika provtagningspunkter kan beskrivas på en lämplig skala. Förhandskunskap som t.ex. verksamhetshistorik och erfarenheter från liknande områden är viktigt för att bedöma om det finns stöd för att genomföra en interpolation.
- Riskbedömning. För riskbedömningen krävs ofta underlag som t.ex. områdets verkliga medelhalt, maximal halt, totalmängd osv. Parametrarna är delvis beroende av vilken förorening och exponeringsväg som är styrande för riskerna. För att göra en komplett riskbedömning ingår även andra analyser och bedömningar som t.ex. toxicitetstester, exponeringsanalyser och bedömning av spridningsförutsättningar.
- Bedömning av åtgärdsomfattning. För att bedöma omfattningen och volymen förorenade massor som kan behöva åtgärdas, behövs en rimlig skattning av hur stor del av området som är förorenat över ett riktvärde eller ett åtgärdsområde. Osäkerheterna i dataunderlaget bör kvantifieras och avspeglas i resultaten.



Figur 1-1 Hierarkisk struktur för ett efterbehandlingsprojekt. Denna rapport hanterar datautvärdering av jordprover (markerat med rött). En arbetsmetodik för att ta fram en lämplig provtagningsstrategi finns föreslaget i rapporten Provtagningsstrategier för förorenad jord (Norrman et al. 2009) och är markerat med blått i figuren.

2 Ramverk för statistisk utvärdering av data

I detta kapitel beskrivs ett ramverk för användning av statistiska metoder vid utvärdering av data från förorenad jord. Ett förfarande i fyra huvudsteg beskrivs, med ett antal olika möjliga metoder, men även vissa verktyg att tillgå i varje steg.¹ Ramverket kan ses som en något idealiserad bild av i vilket skede de olika metoderna kommer till användning; man kan också tänka sig att metoderna kan användas i andra skeden.

Motivet till att lyfta fram statistiska metoder är framförallt att osäkerheterna i de olika bedömningarna som görs kan kvantifieras och beskrivas. Resultatet från den statistiska utvärderingen ger både tyngd och transparens åt en undersökning som inte uppnås på samma sätt genom subjektiva bedömningar. Med statistikens hjälp blir beslutsunderlaget tydligare och mer objektivt. Detta är även relevant information att kommunicera med beställare och tillsynsmyndighet eftersom det ger en inblick i hur stor sannolikheten är för felaktiga bedömningar och beslut, dvs. det ger beslutsfattarna information om vilka felrisker som måste hanteras. Detta kan i sin tur kopplas till ekonomiska risker.

Ramverket omfattar en rad metoder att använda för att utvärdera ett enskilt *egenskapsområde*. Med egenskapsområde avses ett delområde inom vilket föroreningen är genererad genom samma typ av förorenande process och som uppvisar relativt homogena egenskaper med avseende på exempelvis geologi och föroreningssituation. Här antas att data som är insamlade från ett egenskapsområde tillhör samma statistiska målpopulation². Ett undersökningsområde kan i många fall delas upp i flera olika egenskapsområden. Indelningen i egenskapsområden bör göras redan vid planeringen av provtagningen, se vidare avsnitt 7.1.

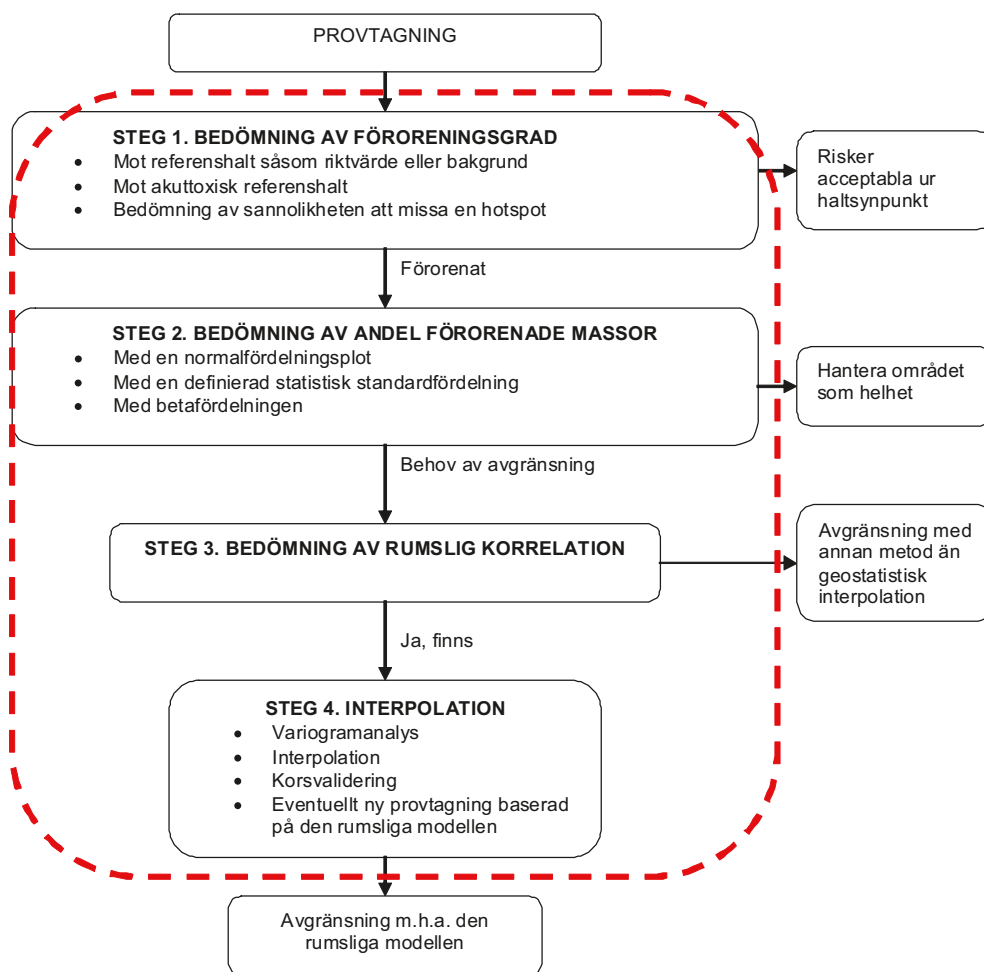
En förutsättning för de allra flesta statistiska metoder är att de mätdata som finns tillgängliga har samlats in med hjälp av slumpmässig provtagning (sannolikhetsbaserat angreppssätt), eller åtminstone så slumpmässig som möjligt. Om datamängden inte är helt slumpmässigt framtagen kan det ändå finnas verktyg som delvis kan kompensera för detta, vilket beskrivs kortfattat i kapitel 7. I rapporten ”Provtagningsstrategier för förorenad jord” (Norrman et al. 2009) finns en arbetsmetodik beskriven för att utforma provtagningsstrategier som kan användas för miljötekniska markundersökningar i jord. Delen som tar upp sannolikhetsbaserade angreppssätt ger viktig information för att skapa bra förutsättningar för att kunna tillämpa de metoder och verktyg som föreslås i denna rapport.

¹ Med statistisk metod avses själva beräkningsmetoden och principer för denna. Med verktyg avses t.ex. programvaror som utför de olika beräkningarna. Det kan vara programvaror som är utformade för statistiska analyser eller mer allmänna beräkningsprogram som t.ex. Excel.

² Målpopulationen är den totala mängd av koncentrationsvärden i egenskapsområdet som man önskar kunna säga något om (i princip en oändlig mängd). Eftersom målpopulationen inte är känd måste man ta prover för att karaktärisera den.

Det bör också nämnas att betydelsen av att använda expertbedömningar och fler bevislinjer (*lines of evidence*) än bara halter i jorden är viktigt för en komplett utredning och riskbedömning. I denna rapport är utgångspunkten föroreningshalter i jord men samma typer av metoder kan lika väl användas för exempelvis toxicitetstester eller andra typer av data, eller då andra jämförvärden än riktvärden används. Figur 2-1 visar en schematisk bild av de fyra huvudstegen i ramverket:

1. Bedömning av föroreningsgrad.
2. Bedömning av andelen förorenade massor.
3. Bedömning av rumslig korrelation.
4. Interpolation.



Figur 2-1 Det föreslagna ramverket med Steg 1 till 4. Det som är inringat med röd streckad linje behandlas i föreliggande rapport.

2.1 Steg 1 Bedömning av föroreningsgrad

Det första steget i ramverket syftar till att göra en bedömning av om området är förorenat eller ej. Här föreslås ett antal statistiska analyser som kan göras med detta syfte:

1. Bedömning genom jämförelse mellan en representativ halt inom området och en referenshalt,
2. bedömning av hur stor andel av området som överskrider en akuttoxisk halt, samt
3. bedömning av sannolikheten att provtagningen har missat en *hotspot* (ett delområde med hög föroreningshalt) av en bestämd storlek.

Metoder och verktyg för dessa analyser beskrivs närmare i kapitel 3. Samtliga tre analyser relateras till egenskapsområdets storlek på olika sätt. Genom att ta fram en skattning av den verkliga medelhalten inom området kan detta, tillsammans med storleken på området och mäktigheten på det förorenade jordlagret användas för att skatta mängden förorening inom området. Mängden förorening är information som normalt inkluderas i en riskbedömning. I den andra analysen bedöms hur stor andel av området som har halter som överskrider akuttoxiska halter. Detta kan räknas om till en andel av områdets totala area även om det inte finns information om exakt *var* dessa massor finns inom området. Man kan också se det som ett mått på sannolikheten att ett barn som rör sig i området riskerar att exponeras för en jordvolym med akuttoxiska halter. Den tredje analysen, bedömning av sannolikheten att missa en *hotspot*, relaterar direkt till områdets storlek men också till provtagnings-tätheten inom området. Metoden som används är en form av geometrisk simulering som bygger på områdets area och form samt provpunkternas placering.

Genom att utföra dessa analyser och resonera omkring kvantifierade osäkerheter bör ett område inte kunna anses vara acceptabelt ur risksynpunkt, baserat på liten datamängd eller osäkra data. Om resultaten från den första analysen pekar på att den representativa halten för området underskrider referenshalten, att andelen massor inom området som kan påvisa akuttoxiska halter är acceptabel, samt att sannolikheten att ha missat en *hotspot* med relevant storlek och form är rimlig, så bör riskerna i området kunna bedömas vara acceptabla ur haltsynpunkt. Det kan däremot finnas helt andra aspekter i riskbedömningen som inte stödjer denna bedömning, såsom t ex ny kunskap om olika föroreningars egenskaper.

Slutsatsen från Steg 1 är att området antingen är rent, så till vida att det kan lämnas utan någon efterbehandlingsåtgärd, eller att området kräver någon form av efterbehandling, dvs. det bedöms utgöra en risk. Om området bedöms utgöra en risk och någon form av efterbehandling krävs är nästa steg i det föreslagna ramverket att bedöma hur stor andel av området som är förorenat.

2.2 Steg 2 Bedömning av andelen förorenade massor

Det andra steget är att göra en bedömning av andelen förorenade massor som har halter över riktvärdet. Andelen förorenade massor inom ett område är viktigt att bedöma för exempelvis riskbedömningen med avseende på mängd- och spridningsberäkningar, men även för att kunna göra kostnadsbedömningar.

Skattningen av hur stor andel av området som har massor över en viss referenshalt kan användas för att bedöma hur stor area eller volym förorenade massor som finns. Arealen, eller volymen, förorenade massor kan användas för att avgöra om endast de förorenade massorna med halter över referenshalten skall åtgärdas eller om hela området behöver åtgärdas. Om hela området behöver åtgärdas kan det vara mer kostnadseffektivt att direkt åtgärda utan att exakt avgränsa var de förhöjda halterna är lokaliserade. Däremot, om endast en mindre del av området behöver åtgärdas, är det ofta kostnadseffektivt att avgränsa dessa områden.

Andelen förorenade massor kan beräknas på olika sätt, här redovisas tre metoder som beskrivs mer ingående i kapitel 4:

1. med en normalfördelningsplot av data,
2. med en definierad statistisk standardfördelning, eller
3. med en betafördelning.

2.3 Steg 3 Bedömning av rumslig korrelation

Det tredje steget är att bedöma om det finns någon rumslig (spatial) korrelation mellan föroreningarna. Rumslig korrelation indikerar att uppmätta halter från datapar med ett givet avstånd från varandra är mer (eller mindre) lika än vad som kan förväntas om halterna vore helt slumpmässigt fördelade i rummet. Vanligtvis uppvisar datapar på små avstånd från varandra större likhet i uppmätta halter än datapar på stora avstånd från varandra. Vore halterna helt slumpmässigt fördelade i rummet skulle man inte kunna förvänta sig detta.

Om ingen rumslig korrelation kan antas finnas i området är det olämpligt att gå vidare med geostatistisk interpolation eftersom de bygger på antagande om att det faktiskt föreligger en rumslig korrelation. Om det fortfarande bedöms som att det är viktigt att försöka avgränsa områden med förhöjda halter bör andra metoder övervägas, t.ex. variansanalys eller geofysiska metoder. Om inga alternativa metoder finns att tillgå återstår avgränsning (eller klassning) antingen i själva saneringsskedet eller m.h.a. förtätad provtagning i efterbehandlingsenheter.

De ansträngningar man gör för att beskriva graden av rumslig korrelation i data bör till en början fokusera på att avgöra om rumslig korrelation över huvud taget finns eller inte. Om det inte existerar någon rumslig korrelation i data, då är antaganden om oberoende mellan data och de metoder som används för att beräkna exponering relevanta i förhållande till att avgöra om efterbehandling är nödvändig eller inte. Om rumslig korrelation däremot kan

observeras eller bestämmas med hjälp av statistiska tester, så är det möjligt att modellera fördelningen av föroreningar i rummet och använda den modellen för att avgränsa områden med förhöjda halter.

Statistiska metoder för att visualisera och bedöma rumslig korrelation finns tillgängliga, men många av dessa används sällan vid utvärdering av data från förorenade områden. Anledningen är att det inte finns lättillgängliga verktyg för att utföra flera av testerna. Några av de metoder som kan användas för att bedöma rumslig korrelation beskrivs mer i kapitel 5.

Innan statistiska metoder används är det väsentligt att ta fram en konceptuell modell av området och samtidigt göra en bedömning om det är rimligt att det kan finnas en rumslig korrelation. En rumslig korrelation bör alltid kunna förklaras med någon form av fysisk process som har genererat föroreningen på platsen och föroreningens fördelning i rummet. För en plym som sprids i grundvattnet är det ganska enkelt att koppla en fysisk förståelse av hur halter i plymen kan tänkas bero av varandra. I jord kan det vara svårare. Exempel på en fysisk process som kan ge upphov till rumslig korrelation i jord är atmosfärisk deposition av föroreningar, t.ex. bly från en väg eller partiklar från en skorsten där en viss vindriktning är mest vanlig. Andra exempel är om det funnits en lagringsplats där spill skett inom ett begränsat område eller om spill/utsläpp skett längs en ledningsgrav eller transportväg. Markföroreningar som genererats genom liknande processer har en rumslig riktning och korrelation.

Förorenade områden i urbana miljöer innehåller ofta fyllnadsmassor med varierande ursprung. På sådana områden kan det vara mycket svårt, eller omöjligt, att koppla till en fysisk process som har genererat en viss fördelning av föroreningshalter i rummet. En viss korrelation mellan punkter finns alltid, men på hur stora avstånd ett sådant samband finns är viktigt att fundera över. Om ett samband mellan punkter i rummet bara kan antas finnas på en skala av några decimetrar så är den informationen inte speciellt användbar för geostatistisk interpolation och avgränsning.

2.4 Steg 4 Interpolation

I det sista och fjärde steget som utförs i de fall en rumslig korrelation på tillräckligt stora avstånd kan påvisas och förklaras, är en geostatistisk interpolation. Syftet med interpolationen är att försöka avgränsa områden med förhöjda föroreningshalter från områden med lägre halter.

För att en interpolation skall kunna utföras och generera ett rimligt resultat väljs en lämplig interpolationsmetod, den geospatiala modellen skall valideras och eventuell ytterligare provtagning övervägas. Kartorna som genereras kan användas som grund för att undersöka beslutsosäkerheter i förhållande till utformning av åtgärder, volymer, kostnader eller för att få en bättre riskbedömning.

2.5 Åtgärds mål och val av efterbehandlingsvolym

Givet att en rumslig modell för halter över ett område finns tillgänglig³ finns det i princip två typer av beslutsskalor som åtgärds målet kan tillämpas på när efterbehandlingsinsatserna utformas. Den ena är en beslutsskala som tillämpas på varje enskilt block eller selektiva efterbehandlingsvolym (SEV). Den andra möjligheten är en beslutsskala som tillämpas på ett större område, ett område som kan motsvaras av en exponeringsenhet, dvs. som är kopplad till hur exponeringen för det styrande ämnet sker. Vardera av dessa beslutsskalor har olika mål för efterbehandling och kan ge mycket olika resultat beroende på hur föroreningen är distribuerad i området. De är kopplade till någon form av kriterium, t.ex. ett generellt eller platsspecifikt riktvärde eller ett humanrisk- eller ekorisk kriterium.

På en beslutsskala som tillämpas på varje enskild SEV, utgår man från att varje SEV skall uppnå beslutskriteriet dvs. att åtgärds målet skall tillämpas på varje enskild SEV. Till exempel kan detta innebära att medelhalten inom varje SEV skall underskrida referenshalten eller efterbehandlas. Detta kan givetvis kopplas till att en viss konfidensgräns för medelhalten inom blocket skall underskrida referenshalten.

På en beslutsskala som tillämpas på en exponeringsenhet, utgår man från att åtgärds målet skall tillämpas på varje exponeringsenhet. Till exempel kan detta innebära att de värst förorenade SEV-enheterna skall behandlas till dess att medelhalten inom *hela* exponeringsenheten underskrider referenshalten. Detta innebär att det kan finnas kvar SEV-enheter i området som överskrider åtgärds målet.

SADA (2008) har implementerat möjligheter att arbeta med dessa två olika beslutsskalor. För en beslutsskala som tillämpas på varje SEV, identifieras dessa baserat på den rumsliga modellen. För en beslutsskala som tillämpas på en exponeringsenhet ordnas de enskilda SEV inom exponeringsenheten från det mest förorenade till det minst förorenade och SEV-enheterna markeras för efterbehandling i den ordningen tills det att den representativa halten för hela exponeringsenheten underskrider åtgärds målet.

Det finns även en tredje möjlighet att arbeta med utformningen av efterbehandlingsinsatserna genom att kombinera dessa två beslutsskalor och att tillämpa *två olika åtgärds mål*: ett som tillämpas på en hel exponeringsenhet "Åtgärds mål Område" och ett som tillämpas på SEV-skala "Åtgärds mål SEV". Först tillämpas "Åtgärds mål Omr." för hela området och därefter tillämpas "Åtgärds mål SEV" för de enskilda SEV-enheter som återstår genom att undersöka om något av dessa överskrider "Åtgärds mål SEV". Om så är fallet identifieras återstående block som bör efterbehandlas. Typiskt tillämpas den kombinerade strategin så att "Åtgärds mål Omr." är striktare satt än "Åtgärds mål SEV". Ett exempel skulle kunna vara att inte tillåta att medel-

³ Den rumsliga modellen kan baseras på geostatistisk interpolation eller tät provtagning i fält.

halten i hela exponeringsenheten överskrider riktvärdet och på enskilda SEV-enheter får inte t.ex. en akuttoxisk referenshalt överskridas.⁴

Det bör uppmärksammas att skalan kan ha stor betydelse för hur mycket massor som planeras att efterbehandlas. Om ett åtgärds mål som tagits fram för ett större område, en exponeringsenhet, appliceras på varje enskild SEV-enhet riskeras området att översaneras.

⁴ Det är också på detta sätt man tillämpar t.ex. riktlinjer för buller: man har krav på dels en medelnivå, dels på topparna.

3 Steg 1: Bedömning av föroreningsgrad

Följande kapitel innehåller tre huvudavsnitt. Det första avsnittet beskriver verktyg och metoder för att undersöka den data som finns tillgänglig, dvs. det insamlade stickprovet. Det andra avsnittet diskuterar begreppet representativ halt och referenshalt. Det tredje avsnittet beskriver olika typer av referenshalter som kan användas att jämföra med stickprovet för att bedöma föroreningsgraden.

I varje avsnitt ligger även exempel inbäddade för att läsaren lättare skall kunna få en känsla för den information analyserna ger. Fler exempel finns givna i fallstudien i Kapitel 8.

3.1 Beskrivning av stickprovet

Stickprovet utgörs av de enskilda prover/mätvärden som man har och som antas vara ett slumpmässigt urval från den målpopulation man önskar uttala sig om. Syftet med att undersöka och beskriva stickprovet är att dess egenskaper antas säga något om målpopulationens egenskaper, givet att proverna är insamlade på ett korrekt sätt. Målpopulationens egenskaper styr till stor del vilka metoder som är lämpliga att använda för att utföra olika statistiska beräkningar.

Några begrepp

Stickprov är med statistisk terminologi den mängd prover som tas för att karaktärisera en jordvolym, inte varje enskilt prov. Denna statistiska terminologi används genomgående i rapporten.

Med *representativ* halt avses en halt som tas fram för ett egenskapsområde utifrån insamlad data och som används för att jämförelse med ett riktvärde eller liknande. Den representativa halten är den halt som bäst representerar risksituationen på området utan att risken underskattas. Den kan uttryckas som en skattad medelhalt (med eller utan gardering för osäkerheter), 90-percentilen eller uppmätt maxhalt. Genom hela rapporten används dock UCLM (övre konfidensgräns för medelhalten) som representativ halt.

Med *referenshalt* avses det värde man vill jämföra mot, t.ex. generella eller plats-specifika riktvärden, en bakgrundshalt eller ett referensvärde för akuttoxicitet. För jämförelse mot bakgrundshalter har begreppet jämförelsevärde använts i MIFO (NV, 1999).

Konfidensgraden anger med vilken sannolikhet den verkliga medelhalten ligger inom det beräknade konfidensintervallet. Om ett tvåsidigt 95% konfidensintervall används så är sannolikheten 0,025 att det verkliga värdet är lägre än den undre gränsen och 0,025 att det verkliga värdet är högre än den övre gränsen. Om istället ett ensidigt övre konfidensintervall används är sannolikheten 0,05 att det verkliga värdet är högre än den beräknade nivån.

3.1.1 Beskrivande statistik

Det insamlade stickprovet bör undersökas genom att beräkna några grundläggande parametrar för stickprovet samt att visualisera stickprovets egenskaper genom olika typer av plottar (se t.ex. NV, 2008 och USEPA, 2006). Nedan listas olika typer av statistiska parametrar som kan beräknas för stickprovet:

- Centralmått (lägesmått), t.ex. medelvärde och median. Medelvärdet av stickprovet är en viktig parameter som används för att göra en väntevärdesriktig skattning av medelhalten.
- Spridningsmått, t.ex. bredden på variationen, variansen, standardavvikelsen och variationskoefficienten. Bredden på variationen anges genom min- och maxvärde. Variansen är standardavvikelsen i kvadrat och variationskoefficienten är ett mått som anger hur stor spridningen är i förhållande till medelvärdet.
- Mått på hur en eller flera värden relaterar till andra, t.ex. percentiler. Ofta är man intresserad av 25-, 50- och 75-percentilerna men även högre percentiler som 90-, 95- och 99-percentilerna kan vara intressanta.

Exempel 3-1 Beskrivande statistik.

Tolv samlingsprover valdes ut och analyserades för arsenik. Analysresultat och sammanfattande statistik redovisas i tabellerna nedan. Genom att ta fram beskrivande statistik får man en första bild av data, hur spridningen ser ut och datas fördelning.

Stickprov för uppmätta arsenikhalter (mg/kg).

Prov:	1	2	3	4	5	6	7	8	9	10	11	12
Halt:	52,7	62,6	80,9	33,5	50,0	91,1	74,8	40,8	47,2	53,7	31,0	55,1

Sammanfattande statistik för stickprovet beräknat med ProUCL⁵ (USEPA, 2007).

Variabel	Antal	Min	Max	Medel	Median	Varians	Std	Skevhets	Kurtosis	CV
As	12	31	91,09	56,1	53,19	339,8	18,4	0,572	-0,324	0,329
	Antal enskilda prov	Minsta värdet i stickprovet	Största värdet i stickprovet	Aritmetiskt medelvärde	Medianen (dvs. 50-percentilen)	Varians	Standardavvikelse	Skevhets <0=vänster >0=höger 0=symmetri	Toppighet >3= spetsig <3= flack	Variationskoefficient = standardavvikelse/medelvärde

Data i har inte någon tydlig skev fördelning och heller inte så stor relativ standardavvikelse (dvs. variationskoefficient), dvs. inte så stor spridning i relation till stickprovets medelvärde.

För att få en bra bild av data finns dessutom olika grafiska hjälpmedel:

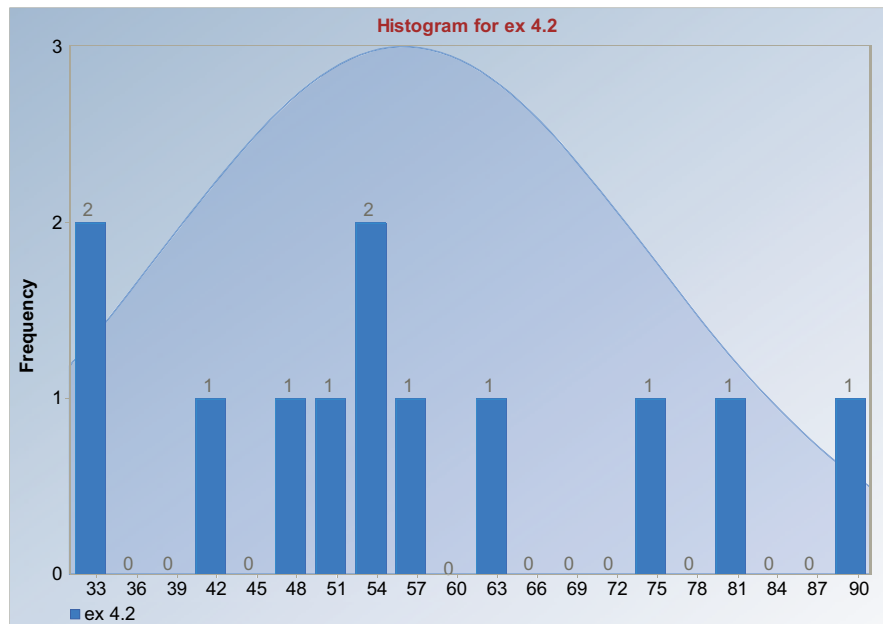
- Histogram, som är en typ av stapeldiagram som redovisar grupperad data. Histogram används för att få en överblick av hur data är fördelade, se Exempel 3-2 Histogram nedan.

⁵ ProUCL är en gratis programvara från USEPA. Se kort beskrivning i Bilaga A.

- Boxplottar, vilka ofta redovisar 25-, 50- och 75-percentilerna men även min- och max-värden. Boxplottar får med både centralmått, spridningsmått och en bild av hur data är fördelade. Exakt vilka parametrar som en boxplot redovisar kan dock variera mellan olika programvaror.
- Kvantilplottar och plottar med rankad data. Båda dessa är diagram som visar rankad data på y-axeln och antingen fraktionen av data eller jämna intervaller på x-axeln.
- Kvantil-kvantilplottar (Q-Q plottar), vilka visar stickprovets kvantiler plottade mot motsvarande kvantiler hos en viss statistisk fördelning, exempelvis en normalfördelning.
- Normalfördelningsplottar (sannolikhetsplottar), vilka plottar halterna hos ett stickprov mot percentilerna. I diagrammet har y-axeln en gradering som gör att en normalfördelning bildar en rät linje. Genom att logaritmera data får man istället en lognormalfördelningsplot.
- Scatterplottar, vilka beskriver hur en variabel varierar i förhållande till en annan.
- Plottar för tidsdata (trenddiagram).
- Plottar för rumslig data, se vidare i kapitel 5.

Exempel 3-2 Histogram

I ett histogram grupperas de uppmätta värden i grupper som visas på x-axeln. På y-axeln visas hur många datavärden det finns inom varje gruppering. I figuren visas ett histogram för stickprovet av arsenikhalter i föregående exempel.



Histogram för arsenikstickprovet i föregående exempel.

I enlighet med den beskrivande statistiken för stickprovet uppvisar inte heller histogrammet någon tydlig skevhet utan det är relativt symmetriskt.

3.1.2 Parametriska och icke-parametriska metoder för att beskriva målpopulationen

I statistiken skiljer man på parametriska metoder och icke-parametriska metoder för att beskriva sin målpopulation. Parametriska metoder innebär att man gör ett modellantagande och utgår ifrån att data (dvs. stickprovet) följer en viss statistisk standardfördelning, t.ex. normalfördelning, lognormalfördelning eller gammafördelning, som antas avspegla målpopulationen. För icke-parametriska metoder⁶ anpassas ingen modell till data, utan beräkningarna utförs direkt på stickprovet. Fördelen med att använda parametriska metoder är att man kan få ut mycket information av sitt stickprov. Nackdelen är att man gör ett antagande om fördelningstyp och detta antagande kan vara fel eller osäkert. Motsatsen gäller för icke-parametriska metoder: de är robusta och kan användas för all typ av data, men å andra sidan ger metoderna resultat som har större osäkerhet.

Det finns ett flertal olika tester för att undersöka vilken fördelning data har. De mest välkända testerna att testa för om data är normalfördelad eller ej är kanske Shapiro-Wilks test, Lilliefors test, Anderson-Darling eller Kolmogorov-Smirnovs test. Det är även väsentligt att bedöma rimligheten i att målpopulationen är fördelad enligt en viss statistisk standardfördelning.

Starzec et al. (2008) rekommenderar att Shapiro-Wilks test används (och även i många USEPA-dokument). Shapiro-Wilks test är svårt att beräkna för hand men finns implementerat i flera programvaror, bl.a. i ProUCL (2008). Tester för att undersöka huruvida data följer en viss fördelningstyp brukar kallas ”goodness-of-fit” (GOF) tester.

Goodness-of-Fit test i ProUCL

I ProUCL finns möjligheten att använda sig av Shapiro-Wilks (SW) test eller Lilliefors test för att testa normal- och lognormalfördelning, där SW kan användas för stickprover ≤ 50 och Lilliefors för stickprover > 50 (USEPA, 2007). Man kan dessutom testa för gammafördelning mha. Anderson-Darlings test eller Kolmogorov-Smirnovs test. Det är möjligt att välja signifikansnivå i testet: 90%, 95% eller 99%. Testen finns beskrivna i USEPA (2007) med rekommendationer. Rekommendationer ges även i USEPA (2006).

Exempel 3-3 Goodness-of-Fit test

Tolv slumpmässiga provpunkter valdes ut och analyserades för arsenik, se resultat nedan. Beskrivande statistik och en histogramplot tyder på att data är skevt fördelad, vilket kan vara en indikation på att datamängden är lognormalfördelad.

Stickprov med arsenikmätningar (mg/kg).

Nr:	1	2	3	4	5	6	7	8	9	10	11	12
Halt:	41,0	141,2	83,6	44,3	98,6	224,5	44,1	65,1	27,3	16,6	60,5	53,7

⁶ Kan också kallas fördelningsfria metoder, dvs. metoder som inte bygger på något antagande om att data kommer från en underliggande fördelning.

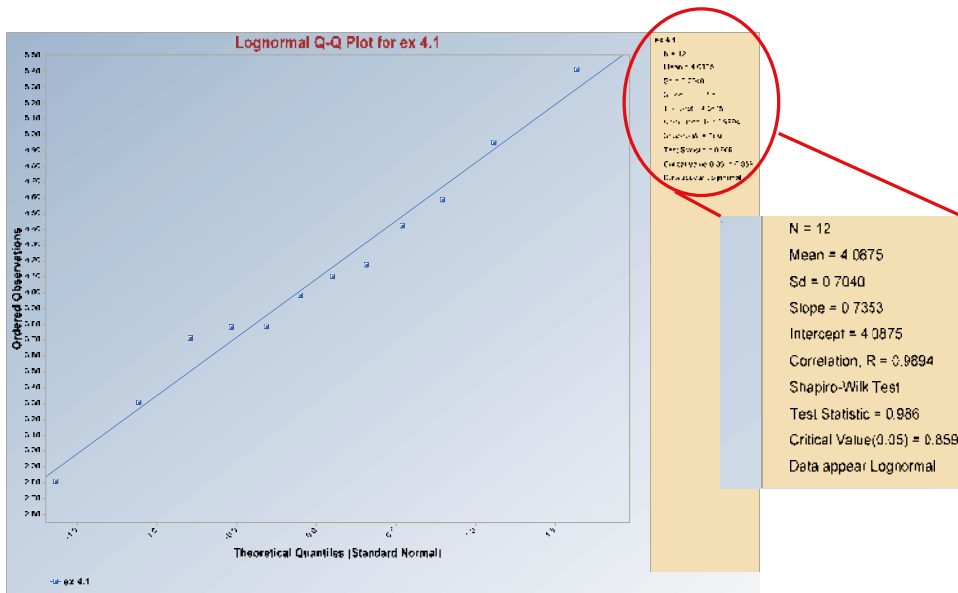
Sammanfattande statistik för stickprovet.

Variabel	Antal	Min	Max	Medel	Median	Varians	Std	Skevhets	Kurtosis	CV
As	12	16,6	224,5	75,04	57,1	3351	57,89	1,811	3,525	0,771

Antal enskilda prov	Minsta värdet i stickprovet	Största värdet i stickprovet	Aritmetiskt medelvärde	Medianen (dvs. 50-percentilen)	Varians	Standardavvikelse	Skevhets <0=vänster >0=höger 0=symmetri	Toppighet >3= spetsig <3= flack	Variationskoefficient = standardavvikelse/medelvärde
---------------------	-----------------------------	------------------------------	------------------------	--------------------------------	---------	-------------------	--	---------------------------------------	--

Datamängden plottas därför i en lognormal Q-Q-plot för att undersöka om data följer en lognormalfördelning. Här är detta gjort mha. ProUCL och ett Shapiro-Wilks test för att testa om data är lognormalfördelad. Nollhypotesen formuleras som att data är lognormalfördelade och testet utförs på signifikansnivån 95%.⁷

I grafen ges även teststatistikan och det kritiska värdet för testet (*Critical Value*) på signifikansnivån 95%. På den nivån accepteras en felrisk på 5%, dvs. ett p-värde på max 0,05. Här är teststatistikan 0,986 och det kritiska värdet 0,859. Eftersom värdet från testet överstiger det kritiska värdet kan man anta att data är lognormalfördelade.



Lognormal Q-Q-plot i ProUCL samt resultat av Shapiro-Wilkstestet.

3.2 Representativ halt

Det har hittills inte funnits några tydliga riktlinjer för vilken halt hos stickprovet som skall jämföras med ett riktvärde men i det vägledningsmaterial om förorenade områden som Naturvårdsverket avser att publicera under 2009 kommer detta att diskuteras. Den halt som representerar stickprovet och som används för jämförelsen benämns ”representativ halt” (Arnér, 2009).

⁷ För mer information om hypotestest se avsnitt 3.4.2

Den representativa halten definieras som den halt som bäst representerar risk-situationen på området utan att risken underskattas.

Som representativ halt bör man välja ett statistiskt mått som exempelvis medelvärdet av uppmätta värden, den övre konfidensgränsen för medelhalten (UCLM, *Upper Confidence Limit of the Mean*), det maximalt uppmätta värdet, en viss percentil av uppmätta värden, eller något annat värde som grundas på bearbetade data. I den här rapporten används en representativ halt baserad på en skattad medelhalt, men med en adderad säkerhetsmarginal för att gardera mot osäkerheter i skattningen, se nedan.

En representativ halt kan bara tas fram för områden som är någorlunda homogena ur föroreningssynpunkt. Förorenade områden med stor variation i föroreningsgrad måste därför först delas in i delområden, eller egenskapsområden, och därefter kan metodiken tillämpas på respektive delområde. Exemplet nedan i Tabell 3.1 visar att man kan komma till vitt skilda slutsatser beroende på vilket statistiskt mått som används som representativ halt. Det är därför viktigt att noga tänka igenom hur den representativa halten ska bestämmas. Valet av representativ halt bör baseras på bl.a:

- Vilken typ av risk som avses; långtidsrisker eller akuta risker.
- Hur säker man vill vara att inte göra fel⁸ vid jämförelsen mot en referenshalt.
- Hur enkel eller avancerad metod man vill använda sig av.
- Hur stort dataunderlaget är.
- Hur pass representativa mätdata är. Riktad provtagning kan ge icke-representativa data och ur denna synvinkel är sannolikhetsbaserade metoder (slumpmässig, systematisk provtagning) att föredra som underlag till den statistiska utvärderingen.
- Hur god förhandskunskap och annan information man har om området.

Tabell 3-1 nedan visar några olika statistiska mått som beräknats på samma datamängd. De tre mest relevanta måtten (fet stil) är i detta fall de som inte bygger på någon antagen statistisk fördelning. Notera att beräknade värden som bygger på antagandet om lognormalfördelade data är oralistiska, vilket beror på att antagandet i detta fall är felaktigt. Dessutom kan man konstatera att medianhalten förmodligen är alldeles för låg för att kunna användas som representativ halt (betydligt lägre än medelhalten).

I den här rapporten har vi utgått från att en skattning av medelhalten med 95 % konfidensnivå används för bedömning av långtidsrisker och att en percentil (dock ej föreslaget vilken) används för bedömning av akuta risker.

⁸ Man kan göra två typer av fel: (1) Den beräknade representativa halten är lägre än riktvärdet trots att den "verkliga" representativa halten är högre, eller (2) den beräknade representativa halten är högre än riktvärdet trots att den "verkliga" representativa halten är lägre. Det första felet innebär att området felaktigt klassas som rent, vilket normalt är det fel som är viktigast att undvika.

Tabell 3-1 Exempel på olika statistiska mått beräknade från en verklig datamängd från en dioxinförorenad upplagsyta. Datamängden består av 25 mätvärden och jämförvärdet är 50 ng/kg. De värden som är lämpligast att använda som representativ halt anges med fet stil.

Typ av representativ halt	Värde [ng/kg]
Medelvärde:	
Aritmetiskt medelvärde	41
Medelvärde baserat på lognormalantagande, MVU-skattning enligt Gilbert (1987)	440
Medelvärde baserat på lognormalantagande, förenklad skattning enligt Gilbert (1987)	13
95% UCLM:	
Baserat på normalantagande (Students t-fördelning)	66
Baserat på lognormalantagande (Land, se Gilbert, 1987)	240 000
Baserat på lognormalantagande (Chebyshev, se USEPA, 2007)	2100
Utan antagen fördelning (standard bootstrap)	65
Utan antagen fördelning (Halls bootstrap, se USEPA, 2007)	68
Utan antagen fördelning (Chebyshev, se USEPA, 2007)	106
Percentil:	
Medianvärde	9,1
90-percentilen	165
99,9-percentilen (extrapolering från data)	Ca 500
Maximalt uppmätt värde	280

För att bedöma om riskerna på ett förorenat område är acceptabla eller inte, ska den representativa halten jämföras med någon typ av referenshalt (eller jämförvärde) i riskbedömningen. Sådana referenshalter beskrivs i följande avsnitt.

3.3 Referenshalter att jämföra mot

Med referenshalt avses här den halt som stickprovet (den representativa halten) jämförs med för att bedöma föroreningsgrad på området. Den vanligaste typen av referenshalt som används idag är ett riktvärde, generellt eller platsspecifikt. En annan typ av referenshalt är kopplad till bakgrundshalterna av ett visst ämne, men det finns även referenshalter som kan kopplas till bedömningsgrunder för miljö kvalitet, TDI-värden, akuttoxiska dosnivåer, fastlagda acceptabla risknivåer etc. I följande avsnitt beskrivs hur ett antal olika typer av referenshalter kan tas fram.

3.3.1 Riktvärden

Många som arbetar med förorenade områden är väl orienterade i principerna för beräkning av riktvärden. Riktvärden baseras på en acceptabel risknivå som med hjälp av ”bakåträkning” räknas om till en referenshalt i jorden med hjälp av toxikologiska och fysikaliska parametrar, transport- och exponeringsmodeller samt exponeringsscenarioer. Generella riktvärden baseras på generella transportmodeller, fysikaliska parametrar och exponeringsscenarioer, medan platsspecifika riktvärden grundas på platsspecifika förhållanden. Generella

riktvärden används vanligen vid förenklade riskbedömningar medan platspecifika riktvärden normal används vid fördjupade sådana. NV (1997a) beskriver principerna för framtagning av riktvärden i Sverige.

Det börjar även bli vanligare att använda sig av framräknade risknivåer vid en riskbedömning, istället för jämförelser mellan stickprov och riktvärden, på samma sätt som normalt sker i en del andra länder. Det mest dominerande förfaringssättet i Sverige är dock fortfarande att använda sig av riktvärden vid en riskbedömning.

3.3.2 Jämförvärde för bakgrundshalt

I MIFO (NV, 1999) beskrivs något som kallas för jämförvärde vilket kan användas för indelning i tillstånd, dvs. i vilken grad objektet är påverkat av punktkällor. Ett jämförvärde skall avspegla den naturliga förekomsten av ett ämne inklusive eventuellt diffust antropogent tillskott. Urbana miljöer har oftast högre jämförvärden än andra områden och även ämnen som inte förekommer naturligt kan ha jämförvärden (NV, 1999). De bästa jämförvärdena är baserade på data från närområdet (utan punktkällor). I tabell 3-2 ges principerna för vilket jämförvärde som bör användas i de fall data från närområdet finns tillgänglig.

Tabell 3-2 Principerna för vilket jämförvärde som ska användas enligt NV (1999).

Stickprovets storlek	Rekommenderat jämförvärde
<5	Kan ej användas
5 - 20	Högsta eller näst högsta värdet
>20	90:e eller 95:e percentilen

Ofta finns inte undersökningar av bakgrundshalter i närheten av undersökningsområdet. Då används istället regionala eller nationella undersökningar. En rapport från Naturvårdsverket (NV, 1997b) redovisar bakgrundshalter i Sverige i tätorter och i naturliga miljöer. Redovisningen är baserad på SGUs geokemiska kartering samt provtagning i 19 tätorter i landet på uppdrag av NV. Resultaten i NV (1997b) används för att ange jämförvärdet i MIFO, Bilaga 5, Tabeller 1 till 4. Här är jämförvärdet 90-percentilen av antingen SGUs geokemiska kartering eller NVs tätortsprovtagning.

I ProUCLs tekniska guidedokument (USEPA, 2007) anges fyra olika parametrar som kan användas som jämförvärde och metoder för att beräkna dessa:

- Övre percentiler: t.ex. 90-, 95- eller 99-percentilen.
- Övre prediktionsgränser⁹, vanligtvis gränsen för det övre 95%iga prediktionsintervallet.
- Övre toleransgränser för medelvärdet¹⁰.
- IQR övre gräns¹¹, för ickeparametrisk data.

⁹ UPL – Upper prediction limit

¹⁰ UTL – Upper tolerance limit

¹¹ IQR – Interquartile range

För närmare förklaring av dessa statistiska parametrar hänvisas till USEPA (2007) och andra statistiska texter. I ProUCL har man möjlighet att beräkna dessa olika jämförvärden med både parametriska och icke-parametriska metoder. ProUCL rekommenderar att stickprovet med bakgrundsdata skall bestå av minst 8-10 prover, helst fler.

Exempel 3-4 Beräkning av jämförvärde för bakgrundshalt.

Ett stickprov med 12 enskilda prover samlades in för att beräkna ett jämförvärde i ProUCL, se analysresultaten nedan.

Stickprovet av bakgrundshalter, arsenik (mg/kg).

Nr	As (mg/kg)	Nr	As (mg/kg)
1	48,3	7	10,6
2	28,2	8	7,9
3	12,4	9	18,2
4	35,2	10	5,8
5	8,5	11	4,6
6	23,3	12	15,8

Ett Shapiro-Wilks test utfördes i ProUCL för att testa om data följer en lognormalfördelning, vilket den befanns göra. Bakgrundstatistik beräknades med ProUCL med antagandet att data och målpopulationen är lognormalfördelad och ProUCL rekommenderar att den övre prediktionsgränsen med konfidensen 95% används. Man kan välja konfidensnivå för den övre prediktionsgränsen (UPL, *Upper Prediction Limit*) samt för den övre toleransgränsen (UTL, *Upper Tolerance Limit*). För den senare skall man även välja för vilken percentil man beräknar den övre toleransgränsen (Coverage).

Rekommenderat jämförvärde i detta fall är enligt ProUCL 56 mg/kg, vilket är den beräknade 95%iga övre prediktionsgränsen (UPL). Här redovisas inte metoden hur detta värde beräknas, se istället USEPA (2007). Som jämförelse kan nämnas att om man hade följt NVs rekommendationer skulle jämförvärdet vara högsta eller näst högsta värdet, dvs 48 eller 35 mg/kg.

3.3.3 Referenshalt för akuttoxicitet

För vissa ämnen är det aktuellt att även göra en jämförelse av uppmätta halter mot ett referensvärde för akuttoxicitet. Arsenik och cyanid är ämnen som i Naturvårdsverkets modeller betraktas som akuttoxiska (t.ex. NV, 1997a). För dessa ämnen beaktas akuttoxicitet vid beräkning av de svenska generella riktvärdena. Modellen som används bygger på att små barn i åldern 0-2 år har större benägenhet att få i sig jord, speciellt vid s.k. pica-beteende, samtidigt som dessa barn har en låg kroppsvikt. NV (2007) ger ekvationen för att beräkna referenshalten i jord för akuttoxicitet:

$$C_{AE} = \frac{ARV \cdot m_{barn}}{m_{intag}}$$

I ekvationen står referenshalten C_{AE} för den föroreningskoncentration [mg/kg] vid vilken akuttoxiska effekter inte förväntas uppkomma, ARV är referensdosen för akuttoxiska effekter [mg/kg kroppsvikt], m_{barn} är vikten hos ett litet barn som exponeras [kg] och m_{intag} är jordintagets storlek vid ett enskilda tillfälle [kg]. Enligt NV (2007) används kroppsvikten 10 kg och jordintagets storlek 5 gram. Referensdosen ARV beror på vilken typ av akuttoxisk effekt som avses. Enligt Rosén et al. (2008a) används en referensdos där ingen akuttoxisk effekt förväntas i Naturvårdsverkets riktvärdesmodell, dvs. TDAE (Tolerabel Dos Akuta Effekter). Rosén et al. (2008a) flaggar för att eventuellt använda en referensdos vid vilken dödsfall förväntas i samband med kostnadsnyttoanalyser.

Exempel 3-5. Beräkning av referenshalt för akuttoxicitet för arsenik

För Arsenik anges TDAE av Naturvårdsverket (remissversion 2007-10-19) till 0,05 mg/kg kroppsvikt. Detta ger en referenskoncentration för jord på 100 mg/kg med de antaganden som görs i Naturvårdsverkets modell.

Enligt White (1999) kan doser om ca 1 mg/kg kroppsvikt leda till dödsfall. Referenskoncentrationen i jord för dödlig dos skulle då motsvaras av 2000 mg/kg med motsvarande antaganden.

3.4 Metoder för jämförelser mellan stickprov och referenshalt

I följande avsnitt beskrivs tre olika principer för att jämföra ett stickprov (den representativa halten) med en referenshalt eller ett annat stickprov.

De tre metoderna är:

- Konfidensintervall för medelhalten.
- Hypotestest: stickprov mot referenshalt.
- Hypotestest: stickprov mot stickprov.

För alla tre tillvägagångssätten gäller att stickprovet först måste undersökas för att man ska kunna välja lämplig metod.

Att använda konfidensintervall vid jämförelse med riktvärde

Med hjälp av stickprovet kan en skattning göras av den verkliga men okända medelhalten. Denna skattning har en viss osäkerhet och genom att beräkna ett konfidensintervall kan man ta hänsyn till denna osäkerhet. Konfidensintervallets övre gräns benämns UCLM och med konfidensnivån 95% blir beteckningen UCLM95. Om UCLM95 beräknas till exempelvis 900 mg/kg kan man göra ett uttalande som: "Sannolikheten att den verkliga medelhalten i området är lägre än 900 mg/kg är 95%", vilket är detsamma som att säga att "sannolikheten för att den verkliga medelhalten i området är högre än 900 mg/kg är 5%". Om UCLM95 är *lägre* än riktvärdet kan man göra ett uttalande som: "Sannolikheten att den verkliga medelhalten i området överskrider riktvärdet är mindre än 0,05 (5%)". Observera att om däremot UCLM95 är *högre* än riktvärdet så kan man inte enkelt kvantifiera sannolikheten att den verkliga medelhalten överskrider riktvärdet. Slutsatsen blir emellertid i det senare fallet att den verkliga medelhalten i området bedöms vara högre än riktvärdet.

3.4.1 Konfidensintervall för medelhalten

Konfidensintervall är en statistisk term som anger graden av säkerhet för en skattad parameter. Konfidensintervallet anges ofta i form av en punktskattning med felmarginal, till exempel 30 ± 3 samt den konfidensgrad som gäller, t.ex. 95%. Både ensidiga och tvåsidiga konfidensintervall förekommer. Med UCLM95 menas den ensidiga övre konfidensgränsen för medelhalten vid konfidensgraden 95% (*Upper Confidence Limit of the Mean*). Att jämföra UCLM95 med referenshalten (riktvärdet) innebär att man jämför medelhalten mot referenshalten, men att man som en extra säkerhet lägger på ett definierat säkerhetsintervall för medelhalten. Eftersom den verkliga medelhalten aldrig är känd utan skattas med hjälp av stickprovet så är detta ett sätt att gardera sig mot denna osäkerhet så att inte risken underskattas. I praktiken innebär det att man säger att man maximalt accepterar en sannolikhet på 5% att den verkliga medelhalten överskrider UCLM95. Att beräkna UCLM95 innebär alltså i praktiken att beräkna storleken på säkerhetsintervallet.

Det finns flera olika metoder att beräkna UCLM och vilken metod som bör användas beror på vilka antaganden som kan göras om populationens statistiska fördelning. För normalfördelade data brukar t-fördelningen användas (t-kvantiler), se exempelvis USEPA (2006; 2007).

Hur bör man tänka angående en förorenings fördelning?

När man undersöker en datamängd är det rimligt att ha en idé om hur data är fördelad som bygger på kunskap om hur data har genererats. Föroreningshalter brukar exempelvis ofta vara skevt fördelade, t.ex. lognormalfördelade. Andra fenomen kan uppvisa en annan fördelning. När man testar data med avseende på fördelning är det viktigt att ha med sig ett modelltänkande gällande fördelningen och inte bara blint fokusera på huruvida konfidensnivån 95% är uppnådd eller ej.

I praktiken följer många datamängder både en lognormalfördelning och en gammafördelning och de två fördelningstyperna kan vara svåra att skilja åt, speciellt vid små stickprov ($n < 50$ till 70), enligt USEPA (2007). Dessutom noteras att UCLM-beräkningar baserade på gammafördelningar resulterar i mer pålitliga och stabila resultat, vilket är av praktisk nytta (USEPA, 2007). Därför rekommenderar ofta ProUCL att ett gamma UCLM används istället för ett lognormalt UCLM. I många fall följer dock inte data någon specifik fördelning och då kan det vara lämpligast att beräkna UCLM med någon fördelningsfri beräkningsmetod, dvs. en metod som inte kräver något antagande om en specifik statistik fördelning.

För lognormalfördelade datamängder är beräkningarna mer komplicerade än för normalfördelade data och det finns det flera olika metoder utvecklade. Land's metod är en exakt metod, men den är känslig för om data verkligen är lognormalfördelad eller ej, särskilt om lognormalmodellen avviker från de allra högsta mätvärdena. I så fall kan metoden generera mycket höga värden på UCLM95, värden som i praktiken kan vara orimliga. I ProUCL finns förutom Land's en metod som bygger på Chebyshev's olikhet (vilken egentligen är icke-parametrisk metod) och som är anpassad för lognormalfördelad data.

Denna metod rekommenderas ofta i de fall Land's metod genererar orimligt höga värden på UCLM95. Metoderna beskrivs i USEPA (2007). Tabellerade värden på h-kvantiler, som används i Land's metod, redovisas i Gilbert (1987) och i USEPA (2006).

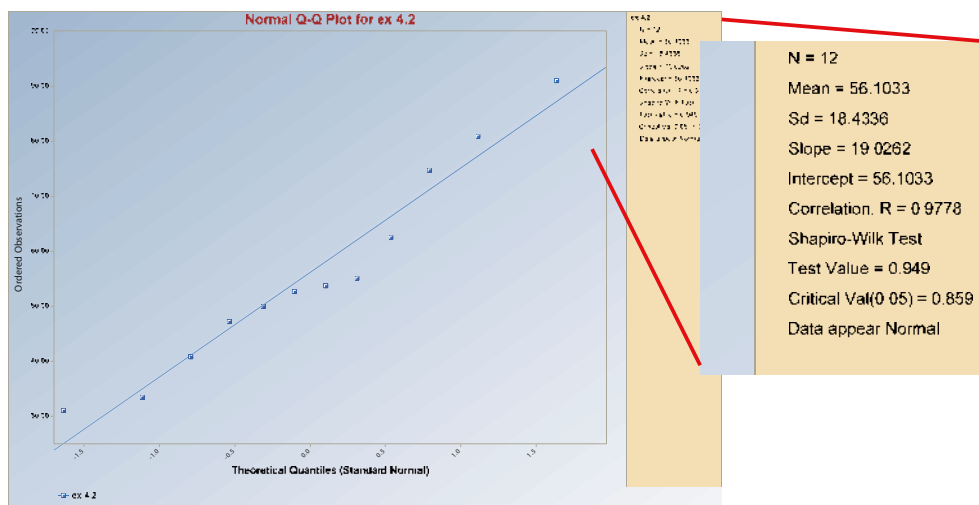
För gammafördelad data finns två olika metoder implementerade i ProUCL, approximativ gamma UCLM och anpassad gamma UCLM, vilka båda finns beskrivna i USEPA (2007). Dessa metoder är inte lika känsliga för extremvärden som Land's metod ovan.

Det finns även icke-parametriska metoder som inte kräver något antagande om en viss underliggande statistisk fördelning. USEPA (2007) beskriver tio metoder som finns implementerade i ProUCL. En bygger på Jackknifing och fem på Bootstrap-metoder, vilka är tekniker som använder sig av upprepade teoretiska provtagningar ur det insamlade stickprovet. Dessutom beskrivs en metod som bygger på t-kvantiler modifierad för skeva fördelningar samt ytterligare några metoder, bl.a den icke-parametriska Chebyshev's olikhet.

Ibland kan den beräknade övre konfidensgränsen för medelhalten vara högre än det maximalt uppmätta värdet, speciellt vid små stickprov (få prover) och skeva datamängder. I USA har därför en praxis utvecklats där man väljer det minsta av UCLM95 och den maximalt uppmätta halten (USEPA, 2007). I USEPA (2004) har dock simuleringar gjorts som visar att för skeva datamängder och små stickprov (<10-20) så underskattar maxvärdet UCLM95 och för större stickprov överskattar maxvärdet UCLM95. Därför rekommenderas i USEPA (2007) att man undviker att använda maxvärdet som representativ halt och istället använder UCLM-värdet. Tidigare har det varit problematiskt att beräkna rimliga värden på UCLM men idag finns så många metoder tillgängliga att det sällan är något stort problem, se exempelvis ProUCL version 4.0 (ProUCL, 2008). Rekommendationer om vilken beräkningsmetod som bör användas finns i USEPA (2007) för normal, lognormal-, gammafördelningar samt datamängder som inte följer någon tydlig fördelning.

Exempel 3-6. Beräkning av UCLM95 för medelhalten

UCLM95 skall beräknas för stickprovet som redovisas i exempel 3-1. Baserat på beskrivande statistik görs ett Shapiro-Wilks test för att undersöka om data kan antas vara normalfördelad, se Q-Q plot från ProUCL nedan.



Q-Q plot för stickprovet.

Baserat på Shapiro-Wilks test antas data vara normalfördelad. UCLM95 beräknas m.h.a. Student's t-fördelning till 65,7 mg/kg och medelhalten skattas till 56,1 mg/kg. Båda dessa värden överskrider tydligt riktvärdena för arsenik, både för känslig markanvändning och för mindre känslig markanvändning, på 15 respektive 40 mg/kg¹².

Om UCLM95 istället skall jämföras med det jämförvärde på 56,3 mg/kg som tagits fram utifrån bakgrundshalter i närområdet (se Exempel 3-4), blir slutsatsen densamma: området är förorenat över referenshalten med en maximal felrisk på 5 % (i det här fallet en referenshalt baserat på bakgrundshalter). Detsamma gäller för de jämförvärden som kan användas enligt NV: 48,3 resp. 35,2 mg/kg.

Exempel 3-7. Beräkning av UCLM95

UCLM95 skall beräknas med hjälp av stickprovet nedan. Genom ett Shapiro-Wilks test har stickprovet kunnat antas vara lognormalfördelat, varför Land's metod kan användas för beräkning av UCLM95. ProUCL används för beräkningarna.

Stickprov av arsenik

Prov nr	As mg/kg	Prov nr	As mg/kg
1	21,67	9	1,46
2	5,70	10	4,93
3	7,71	11	17,24
4	27,71	12	56,76
5	0,79	13	2,70
6	6,49	14	1,12
7	19,34	15	16,48
8	13,40		

Sammanfattande statistik för stickprovet

Variabel	Antal	Min	Max	Medel	Median	Varians	Std	Skevhet	Kurtosis	CV
As	15	0.79	56.76	13.57	7.71	213.6	14.62	1.992	4.963	1.077
	Antal enskilda prov	Minsta värdet i stickprovet	Största värdet i stickprovet	Aritmetiskt medelvärde	Medianen (dvs. 50-percentilen)	Varians	Standardavvikelse	Skevhet <0=vänster >0=höger 0=symmetri	Toppighet >3= spetsig <3= flack	Variationskoefficient = standardavvikelse/medelvärde

Beräkningar med Land's metod ger ett UCLM95 som är ca 48 mg/kg. Medelhalten (utan säkerhetsintervall) skattas till ca 15 mg/kg. Här har vi alltså en situation där den skattade medelhalten ligger under riktvärdet för mindre känslig markanvändning på 40 mg/kg, men där UCLM95 beräknat med Land's metod ligger över detta riktvärde. I en sådan situation bör man behandla området som att det har en medelhalt som överskrider riktvärdet.

ProUCL ger rekommendationen att testa om data kan antas följa en gammalfördelning, och i så fall kan ett gamma UCLM användas istället. Om samma stickprov testas mot en gammalfördelning så kan man i detta fall även anta att data är gammalfördelade. Ett approximativt gamma UCLM95 beräknas då till 23 mg/kg och ett anpassat gamma

¹² Generella riktvärden från 1997, dessa ersattes i oktober 2008 med nya generella riktvärden.

UCLM95 beräknas till 25 mg/kg. Om de framräknade gamma UCLM95 (se avsnitt 3.4.1 ovan) används istället för UCLM95 beräknat med Land's kan alltså området behandlas som att det har en medelhalt som ligger under riktvärdet på 40 mg/kg med (minst) 95%-ig säkerhet.

3.4.2 Hypotestester

Principer

Hypotesprövning innebär att man med en statistisk test försöker avgöra om ett påstående verkar stämma eller inte. Vid hypotesprövning kan man, med hjälp av ett hypotestest, kvantifiera exakt hur stor risken är att man har fel när man gör ett visst uttalande, utifrån en nollhypotes som man formulerat. Till skillnad från att jämföra ett beräknat UCLM95 med referenshalten (se ovan), där man vet att risken är *maximalt* 5%, kan man genom hypotesprövning ta reda på den faktiska risken.

Formuleringen av nollhypotesen är viktig och för de allra flesta fall är det lämpligt att formulera den som:

H_0 : medelhalten inom området är större än eller lika med referenshalten.

Mothypotesen, eller alternativhypotesen, blir då:

H_A : medelhalten inom området är mindre än referenshalten.

Genom att utföra ett hypotestest kan man göra uttalanden i form av "sannolikheten att ha fel om nollhypotesen förkastas är p ", där p ges av det i testet framräknade *p-värdet*. Felrisken, eller egentligen risken för typ I-fel tas alltså fram genom att ta beräkna detta *p-värde*. Ekvationen för beräkning av *p-värdet* beror av vilket statistiskt test man använder.

Tolkningen av *p-värdet* är inte helt självklar. En bra regel är att för ett givet *p-värde* (t.ex. 0,03) tänka på följande sätt: "Sannolikheten att ha fel om nollhypotesen förkastas är 0,03". Detta innebär att felrisken är 3% om man på basis av hypotestestet säger att medelhalten inom området är lägre än referenshalten. Med ett lågt *p-värde* kan man alltså vara ganska säker på att nollhypotesen kan förkastas utan att man har fel. Om *p-värdet* däremot är 0,25 så är risken 25% att felaktigt förkasta nollhypotesen, dvs. i ett fall av fyra har man fel.

Hypotestester kräver en del av användaren men de är mycket användbara om man önskar att kvantifiera felrisken. För vissa hypotestester kan man dessutom relativt enkelt beräkna *styrkan* i testet. Testets styrka är ett mått på hur stor sannolikheten är att man har rätt när nollhypotesen förkastas (vilket alltså inte är detsamma som $1 - p$). Man kan också genom att ange acceptabla värden på *p-värdet* och *styrkan* beräkna hur många prover man bör ta. Detta beskrivs dock inte i den här rapporten men behandlas i Norrman et al. (2009).

Hypotestest: stickprov mot referenshalt

Olika hypotestester bör användas beroende på vilken typ av fördelning data följer. USEPA (2006) rekommenderar t-test för normalfördelad data, Chen-test för lognormalfördelad data och tecken-test då inget antagande om en viss fördelning kan göras (parameterfri metod), se även Starzec et al. (2008). Man ska dock vara medveten om att icke-parametriska tester inte testar medelhalten mot ett riktvärde utan *medianen* mot ett riktvärde. För skeva datamängder kan medianen vara en grov underskattning av medelhalten, vilket kan leda till att en hälso- eller miljörisk kan underskattas.

Statistiska hypoteser följer ett strikt mönster, där man först formulerar en nollhypotes och därpå en eller flera mot- eller alternativhypoteser. Den felrisk man är villig att ta (att förkasta en *sann* nollhypotes) definieras av *signifikansnivån*, eller α (typ I-fel). Storleken på typ I-felet är den risk man tar att felaktigt förkasta nollhypotesen. Vid all hypotesprövning kan man hamna i fyra olika situationer, se Tabell 3-2. Om nollhypotesen förkastas medan den *de facto* är sann begår man ett α -fel (typ I-fel). Alternativt, om nollhypotesen inte förkastas trots att den är falsk begås ett β -fel (typ II-fel).

Tabell 3-2 De fyra möjliga utfallen vid hypotesprövning.

		Verkligt förhållande (okänt)	
		H ₀ sann (området är förorenat)	H ₀ falsk (området är rent)
Resultat av hypotestestet	H ₀ förkastas (området klassas som rent)	Typ I-fel, α	Korrekt beslut, 1- β (styrka)
	H ₀ behålls (området klassas som förorenat)	Korrekt beslut, 1- α	Typ II-fel, β

Som Grandin (2003) påpekar har det blivit vanligare att man arbetar med den faktiska signifikansnivån i testet som ges av *p-värdet* i ett test. Det beräknade p-värdet i ett test ger sannolikheten för att begå ett α -fel. Om $p < 0,05$ så är sannolikheten att man har fel om man förkastar nollhypotesen mindre än 5%. Detta är i allmänhet (men inte alltid) en acceptabel felrisk, och slutsatsen blir då att mothypotesen blir gällande. Om däremot $p = 0,4$ så är sannolikheten att man har fel om man förkastar nollhypotesen så stor som 40%. Så stor felrisk är i de flesta sammanhang inte acceptabelt och slutsatsen blir därför att man inte kan förkasta nollhypotesen.

Vid utvärdering av data från förorenade områden är typ I-felet (α) förknippat med konsekvenserna av att lämna ett förorenat område utan åtgärd (felaktig friklassning). Typ II-felet (β) är däremot förknippat med konsekvenserna av att klassa ett område som förorenat fast det inte är det. I förlängningen kan konsekvensen då bli att området saneras i onödan. Denna konsekvens är tydligast kopplad till den ekonomiska efterbehandlingskostnaden, men det kan även uppstå andra oönskade konsekvenser av efterbehandlingen i form av t.ex. buller, ökad olycksrisk vid transporter, ökad damning, onödigt utnyttjande av naturresurser samt utsläpp av växthusgaser.

Styrkan i ett hypotestest är beroende av fyra faktorer:

1. Signifikansnivå (α),
2. Provstorlek, dvs. stickprovets storlek (n),
3. Effektstorlek (Δ), eller den minsta förändring eller skillnad man anser vara av vikt att upptäcka (se faktaruta nedan),
4. Variationen hos den studerade variabeln.

De tre första kan man i viss mån styra över, medan variationen beror på egenskaper hos det man studerar (föroreningens variabilitet i jorden).

Även vilken typ av statistiskt test som utförs spelar viss roll. Hög signifikansnivå, liten provstorlek, liten effektstorlek och stor variation i data leder till låg styrka. Detta innebär att om man har få prover och skillnaden mellan medelhalten i området och referenshalten är liten, och man dessutom har en stor variation i data (typiskt för lognormalfördelade data) så kan det vara svårt att dra säkra slutsatser.

För normalfördelade data rekommenderar USEPA (2006) att kontrollera att testet har tillräcklig styrka (eller egentligen: att stickprovet är tillräckligt stort), vilket beskrivs i Norrman et al. (2009). När en icke-parametrisk metod används (t.ex. teckentest) används samma ekvationer för att kontrollera att stickprovet är tillräckligt stort med avseende på krav på testets styrka, men med ett tillägg: stickprovets storlek skall vara +20 % mot det beräknade värdet. För lognormalfördelad data rekommenderas här att Chen-testet används. Att beräkna styrkan i Chen-testet är dock inte lika lätt, det krävs avancerade simuleringar för detta. I Norrman et al. (2009) har dock en förenklad metod tagits fram för att beräkna storleken på stickprovet genom att anta att variationskoefficienten är konstant och sedan logaritmera data och utgå från normalteorin. Den metoden kan användas för att kontrollera stickprovets storlek.

Effektstorlek, minsta detekterbara skillnad eller *gray region*

Effektstorleken kan beskrivas som den förändring eller skillnad man vill kunna upptäcka vid hypotesprövning. I detta sammanhang är det vanligen skillnaden mellan det förorenade områdets verkliga medelhalt och det uppmätta medelvärdet. Effektstorleken kan även beskrivas som den minsta skillnad man anser vara av vikt att upptäcka. Ju mindre skillnad man vill kunna upptäcka, desto svårare är det, och desto fler prover behövs.

US EPA (2002) kallar effektstorleken för *gray region*, vilket i riskbedömningshang kan beskrivas som det område i närheten av riktvärdet där det är acceptabelt att göra en felaktig klassning, dvs. där ett potentiellt beslutsfel inte bedöms få så stora negativa konsekvenser. Ett hypotestest bör formuleras så att en felklassning sker på den säkra sidan om riktvärdet, vilket innebär att inom *gray region* så accepteras att ett område kan klassas som förorenat trots att det är rent.

Exempel 3-8. Hypotesprövning för ett stickprov mot en referenshalt

Ett stickprov med arsenikdata skall jämföras med ett riktvärde för arsenik på 15 mg/kg. Stickprovet från det misstänkt förorenade området innehåller 25 enskilda prover enligt nedan.

Stickprov från misstänkt förorenat område.

Prov nr	As mg/kg	Prov nr	As mg/kg
1	7,1	14	17,0
2	10,9	15	9,3
3	1,8	16	88,6
4	26,4	17	14,0
5	9,9	18	8,7
6	7,5	19	42,8
7	10,5	20	40,0
8	29,9	21	7,5
9	8,9	22	1,3
10	36,0	23	38,1
11	23,7	24	17,2
12	7,6	25	11,6
13	19,2		

Sammanfattande statistik för stickprovet i ProUCL.

Variabel	Antal	Min	Max	Medel	Median	Varians	Std	Skevhet	Kurtosis	CV
As	25	1,3	88,6	19,82	11,6	351,6	18,75	2,27	6,716	0,946
	Antal enskilda prov	Minsta värdet i stickprovet	Största värdet i stickprovet	Aritmetiskt medelvärde	Medianen (dvs. 50-percentilen)	Varians	Standardavvikelse	Skevhet <0=vänster >0=höger 0=symmetri	Toppighet >3= spetsig <3= flack	Variationskoefficient = standardavvikelse/medelvärde

Ett Shapiro-Wilks test utfördes i ProUCL för att testa om fördelningen är lognormalfördelad, vilket den befanns vara. Ett Chen-test utfördes (i Excel) för att testa om medelhalten i det misstänkt förorenade området överskrider riktvärdet på 15 mg/kg. De två hypoteserna formulerades som:

H0: Medelhalten av arsenik i området ≥ 15 mg/kg

HA: Medelhalten av arsenik i området < 15 mg/kg

Resultatet av hypotestestet blir att nollhypotesen inte kan förkastas och p-värdet för testet beräknas till 0,959. Ett så högt p-värde innebär att sannolikheten att man har fel om man förkastar nollhypotesen så stor som 96%. Så stor felrisk som 96% är inte acceptabelt och slutsatsen blir att man inte kan förkasta nollhypotesen, dvs. att området är förorenat över riktvärdet på 15 mg/kg.

Exempel 3-9. Hypotestest för ett stickprov mot en referenshalt

Samma stickprov med avseende på arsenik skall jämföras med det jämförvärde som beräknats utifrån bakgrundshalter i närområdet i Exempel 3-4. Jämförvärdet är 56 mg/kg. Stickprovet från det misstänkt förorenade området innehåller 25 enskilda prover enligt exemplet ovan. Ett Chen-test utfördes (i Excel) för att testa om medelhalten i det misstänkt förorenade området överskrider jämförvärdet på 56 mg/kg. Nollhypotesen formulerades enligt:

H₀: Medelhalten av arsenik i området ≥ 56 mg/kg
H_A: Medelhalten av arsenik i området < 56 mg/kg

Resultatet av hypotestestet blir att nollhypotesen bör förkastas (*p*-värdet är nära noll dvs. sannolikheten att förkasta en sann nollhypotes är mycket liten). Slutsatsen blir alltså att området inte bedöms ha en medelhalt som överskrider bakgrundshalten.

Hypotestest: stickprov mot stickprov

Den tredje principen för jämförelse innebär att stickprovet från det förorenade området jämförs mot ett annat stickprov med hjälp av hypotesprövning. Detta kan vara användbart när man vill jämföra data från undersökningsområdet med data som representerar bakgrundshalter.

Vid jämförelse mellan två stickprov är man vanligtvis intresserad av hur mycket medelhalten i det potentiellt förorenade området överskrider medelhalten för bakgrunden. Hur stor skillnaden måste vara för att området skall anses utgöra en hälso- eller miljörisk finns det inga egentliga riktlinjer för (annat än riktvärden). Ett angreppssätt är att definiera en icke-acceptabel haltskillnad som kan leda till att någon form av handlingsåtgärd måste utföras. Denna skillnad benämns som en ”väsentlig skillnad”, *S* (eng. *substantial difference*). Värdet på *S* kan vara noll eller ett positivt värde baserat på riskbedömningen eller någon riktlinje. Om man inte vet viken värde på *S* som är rimligt kan man utföra tester på mer än ett värde på *S*, eller tom. utföra en känslighetsanalys.

På samma sätt som beskrivits i avsnittet ovan, formuleras en nollhypotes och en alternativhypotes:

H₀: medelhalten i undersökningsområdet är större än eller lika med (medelhalten hos bakgrunden + *S*).

Alternativhypotesen blir då:

H_A: medelhalten i undersökningsområdet är mindre än (medelhalten hos bakgrunden + *S*).

Informationen man får från hypotesprövningen är av samma typ som beskrivits ovan: En kvantifiering av risken att ha fel om nollhypotesen förkastas (typ I-fel) i form av ett *p*-värde.

Denna typ av hypotestest är mindre vanlig eftersom man ofta inte samlar in ett större antal prover för att beskriva bakgrundshalten. I de fall det är relevant att göra detta kan den här typen av hypotestest dock vara användbar. Om lokala bakgrundsdata är insamlade så är det bästa att utföra ett hypotestest där två stickprover jämförs med varandra. Minsta rekommenderade mängd data för bakgrundsstickprovet är 8 till 10 prover och detsamma gäller för stickprovet från det misstänkt förorenade området (USEPA, 2007). Om områdena är stora bör man öka antalet prover och USEPA (2007) rekommenderar då istället att stickprovet innehåller minst 10 till 15 enskilda prover.

Val av hypotestest för jämförelse av två stickprov

Följande rekommendationer ges i USEPA (2007) när man ska välja testmetod för att jämföra två stickprov mot varandra: för normalfördelad data rekommenderas Student's two-sample t-test när variansen i stickproven är densamma. Om variansen i stickproven skiljer sig åt finns Satterthwaite two-sample t-test att tillgå. För lognormalfördelad data är man dock hänvisad till icke-parametriska tester: Wilcoxon-Mann-Whitney (WMW) och kvantiltestet. Även de två förstnämnda hypotestesten för normalfördelad data kan användas om man utför testet på logaritmerade data. För dataset med mätvärden under detektionsgränsen kan Gehan-testet användas. Alla dessa tester beskrivs i USEPA (2007), med fördelar och nackdelar angivna samt anvisningar för beräkningarna. Metoderna finns även implementerade i ProUCL.

För situationer som kräver icke-parametriska metoder rekommenderar ProUCL att både WMW och kvantiltestet utförs. Dessa två metoder testar olika saker: WMW testar medelhalten/medianen i de två stickproverna medan kvantiltestet undersöker hur de två högra svansarna skiljer sig åt. Slutsatsen från de båda testen kan bli olika och i så fall bör alltid slutsatsen dras att området är förorenat, oavsett vilket av testen som säger det.

Exempel 3-10. Hypotestest med två stickprov

Två stickprov har samlats in och analyserats med avseende på arsenik. Det ena stickprovet ($n = 12$) är tagna slumpvis i närområdet till det misstänkt förorenade området med syftet att undersöka bakgrundshalter, och det andra stickprovet ($n = 25$) är slumpmässigt insamlat på det misstänkt förorenade området. Data redovisas nedan.

Prov från misstänkt förorenat område och ett närområde (bakgrund).

Nr	Bakgrund	Misstänkt förorenat område	Bakgrund (logaritmerade halter)	Misstänkt förorenat område (logaritmerade halter)
	As (mg/kg)	As (mg/kg)		
1	34,1	27,7	3,529	3,321
2	2,2	31,2	0,788	3,440
3	6,4	28,6	1,856	3,353
4	7,2	39,0	1,974	3,664
5	18,3	13,8	2,907	2,625
6	11,8	29,9	2,468	3,398
7	6,6	22,2	1,887	3,100
8	10,0	64,7	2,303	4,170
9	5,6	13,0	1,723	2,565
10	18,3	12,8	2,907	2,549
11	9,7	13,8	2,272	2,625
12	7,4	27,8	2,001	3,325
13		27,8		3,325
14		32,3		3,475
15		32,5		3,481
16		22,2		3,100
17		12,0		2,485
18		34,7		3,547
19		11,5		2,442
20		12,3		2,510
21		22,5		3,114
22		17,1		2,839
23		25,6		3,243
24		20,9		3,040
25		16,8		2,821

Fördelningarna testades mot antagandet att de är normalfördelade, men från Shapiro-Wilks test i ProUCL dras slutsatsen att data inte följer en normalfördelning. Istället kunde slutsatsen dras att data är lognormalfördelade. Därför används Students t-test och Satterthwaittes test på logaritmerade data. Nollhypotesen definierades som:

H₀: Medelhalten av arsenik i det misstänkta området är högre än eller lika med medelhalten av arsenik i närområdet.

Stickprovets varians verkar vara lika, vilket innebär att Student's t-test kan användas. Resultatet av testen är att nollhypotesen inte kan förkastas (p-värdet är nära 1 i båda fallen). Slutsatsen är att medelhalten av arsenik från det misstänkt förorenade området bedöms överskrida bakgrundshalten.

För att illustrera principen kan även de icke-parametriska testerna i ProUCL användas: WMW (Wilcoxon-Mann-Whitney) testet och kvantiltestet. Nollhypotesen för WMW-testet formuleras som:

H₀: Medelhalten/medianen av arsenik i det misstänkta området är högre än eller lika med medelhalten/medianen av arsenik i närområdet.¹³

Resultatet från WMW-testet är att nollhypotesen inte kan förkastas. Slutsatsen dras därför att området är förorenat över bakgrundshalten.

ProUCL rekommenderar att kvantiltestet också utförs, speciellt i de fall WMW-testet indikerar att området inte är förorenat. I det här fallet kan alltså kvantiltestet hoppas över. Utförs trots allt ett kvantiltest indikerar detta att området *inte* är förorenat över bakgrundshalten. Detta baseras på stickprovets högra svansar. Svaren kan tyckas motsägelsefulla, men de två metoderna testar olika saker så i statistisk mening är detta inte problematiskt.

3.5 Akuttoxiska föroreningar

Om en förorening riskerar att vara akuttoxisk är det av intresse att bedöma hur stor andel av området som kan uppvisa halter som överskrider en akuttoxisk referenshalt. Hur stor andel eller volym som kan accepteras uppvisa högre halter än de akuttoxiska referenshalterna finns det idag inga riktlinjer för. Däremot kan ett resonemang föras om vad som är en acceptabel sannolikhet att påträffa halter som överskrider den akuttoxiska referenshalten.

Notera att sannolikheten som beräknas avser sannolikheten att påträffa akuttoxiska halter i en godtyckligt vald punkt som motsvaras av provvolymens storlek. Sannolikheten kommer alltså att påverkas av hur stor volym jord som proverna representerar.

För att beräkna hur stor andel av området som uppvisar halter över en akuttoxisk referenshalt hänvisas till de metoder som beskrivs i kapitel 4.

¹³ Icke-parametriska tester testar ofta medianen istället för medelhalten. Detta är viktigt att notera eftersom medianen ofta är betydligt lägre än medelhalten, speciellt vid väldigt skeva fördelningar.

3.6 Bedömning av sannolikheten att missa en hotspot

I många fall är det viktigt att identifiera lokala *hotspots* inom undersökningsområdet. Detta kan exempelvis vara av intresse för ämnen som uppvisar akuttoxiska egenskaper. En *hotspot* som inte upptäcks kan i förlängningen medföra oacceptabla hälsorisker. Då är det relevant att lyfta fram hur stor sannolikheten är att provtagningen har missat en *hotspot*.

Sannolikheten att man vid provtagningen inte upptäckt en *hotspot* (av en bestämd storlek och form) är relaterad till områdets storlek, hur många prover som tagits samt hur provpunkterna placerats i förhållande till varandra. Ju glesare mellan provpunkterna, desto större är sannolikheten att man inte har provtagit i ett område med förhöjda halter, dvs. att man missat *hotspoten*.

Vilken sannolikhet som kan accepteras att inte hitta en *hotspot*, samt vilken storlek denna *hotspot* kan ha, har en koppling till typen av ämne och vilka risker som kan accepteras. Detta blir en diskussionsfråga från fall till fall eftersom det inte finns några riktlinjer för detta.

Verktyget SADA (Stewart & Purucker, 2008) har i version 5 en funktion för att beräkna sannolikheten att missa en *hotspot* med en viss storlek och form, givet ett valt provtagningsmönster. Metoden bygger på Monte Carlo-simuleringar, där en *hotspot* med angiven storlek och form placeras ut slumpmässigt inom det definierade området upprepade gånger. Sannolikheten för att en *hotspot* skall upptäckas är det antal gånger som en eller flera provpunkter ligger inom den utslumpade *hotspotens* area, dividerat med det totala antalet simuleringar. Denna metod skiljer sig mot traditionell sök teori som bygger på att sannolikheten att hitta en *hotspot* räknas ut analytiskt, se Gilbert (1987).

4 Steg 2: Bedömning av andel förorenade massor

I följande avsnitt beskrivs tre olika principer för att bedöma hur stor andel av det aktuella området som har halter som överskrider en referenshalt. De tre principerna bygger på:

- a) en normalfördelningsplot av data,
- b) en definierad statistisk standardfördelning, eller
- c) en betafördelning.

Fördelen med att använda en normalfördelningsplot av data är att man inte behöver göra något antagande om att data är fördelad enligt en viss statistisk standardfördelning (t.ex. normalfördelning eller lognormalfördelning). Istället plottar man helt enkelt de data man har och läser av andelen direkt ur grafen. Metoden kan dock innebära svårigheter vid små stickprov, dvs. då antalet mätvärden är litet.

Metoden att använda en definierad statistisk standardfördelning kan vara fördelaktig om man har relativt små stickprover men ändå kan göra rimliga modellantaganden.

Att använda betafördelningen tillför helt andra möjligheter i skattningen av hur stor andel av ett område som är förorenat över en viss referenshalt. Med betafördelningen krävs inte heller några antaganden om att uppmätt data måste följa en viss statistisk standardfördelning, vilket kan vara en tydlig fördel i många situationer. Metoden tar också hänsyn till att man har större osäkerhet i sin skattning desto mindre stickprovet är. Slutligen bör påpekas att metoden kan användas både med och utan subjektiva inslag, dvs. man kan utnyttja möjligheten att föra in sin expertkunskap eller avstå, allt beroende på problemställning och önskemål.

4.1 Bedömning av andel med normalfördelningsplot

Ett relativt enkelt sätt att bedöma hur stor andel av ett område som är förorenat över en viss nivå är att använda en normalfördelningsplot. Ett exempel på en sådan redovisas i exempel 4-1 **Fel! Hittar inte referenskälla.** nedan. Observera att man inte behöver göra något antagande om att data följer en viss statistisk fördelning (data behöver alltså inte vara normalfördelade eller lognormalfördelade för att metoden ska fungera). Metoden bygger på att man plottar stickprovet i ett diagram med koncentrationen på x-axeln och kvantilerna/percentilerna¹⁴ av stickprovet på y-axeln. Genom att gå in direkt

¹⁴ Kvantiler är på samma sätt som percentiler ett lägesmått sådant att givna proportioner av fördelningens värden är mindre än (eller större än) talen. Exempelvis är 90-percentilen det värde vid vilket 90% av observationerna är mindre än värdet. Uttrycker man de relativa frekvenserna som p i stället för i procent, kallas storheterna kvantiler: 0,9-kvantilen = 90-percentilen.

i diagrammet vid den aktuella referenshalten på x-axeln kan sannolikheten att den halten skall överskridas läsas av på y-axeln. Detta värde representerar sannolikheten (p) att för en slumpmässigt vald punkt ur målpopulationen underskrida referenshalten. Detta är detsamma som att $1-p$ är den andel av området som kan uppvisa halter över aktuell referenshalt.

Metoden kan dock innebära svårigheter vid små stickprover, dvs. få enskilda prover, eftersom avståndet mellan punkterna i diagrammet då blir stort, vilket kan kräva interpolation mellan punkterna.

4.1.1 Metodik

Första steget är att plotta data. Föroreningshalterna plottas på x-axeln och på y-axeln plottas motsvarande percentiler eller kvantiler. Y-axeln är inte linjär utan följer en gradering som motsvarar normalfördelningen. Därför är det enklast att plotta data med hjälp av något statistikprogram. Det är även möjligt att plotta data i kalkylprogrammet Excel men då används en linjär skala på y-axeln. Om data är skevt fördelade kan det vara lämpligt att logaritmera x-axeln. Man får då en så kallad lognormalfördelningsplot.

Nästa steg är att i grafen helt enkelt läsa av hur stor andel av stickprovet som ligger under ett visst riktvärde. Om antalet data är litet kan man behöva interpolera mellan datapunkterna i grafen. Andelen som överskrider riktvärdet är lika med 1 minus andelen under riktvärdet.

Observera att resultatet från grafen bara gäller för den skala som proverna representerar (*support* – se vidare avsnitt 7.7). Om data exempelvis representerar 5 dm³ jord men efterbehandlingen ska göras i enheter på 100 m³ kan det bli problem. Andelen som överskrider riktvärdet minskar nämligen när skalan ökar. Orsaken är att i större jordvolym skär en utjämning när man betraktar medelhalten (det är normalt medelhalten i en efterbehandlingsenhet som beslutet om bortgrävning baseras på). Detta medför att den andel som läses av i en normalfördelningsplot i praktiken kan vara en överskattning. Denna effekt motverkas dock av att även ”ren” jord avlägsnas när stora efterbehandlingsenheter grävs bort.

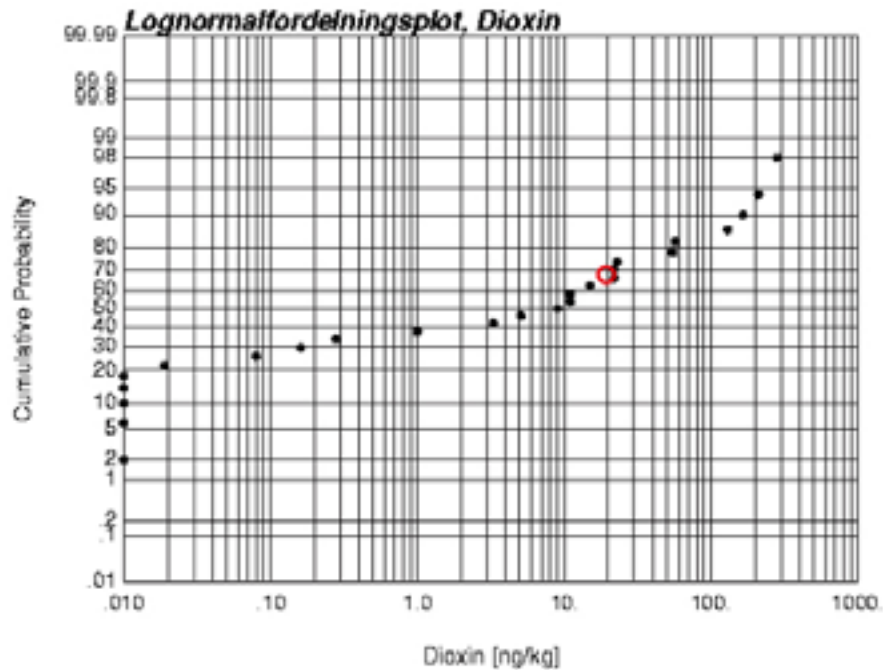
Exempel 4-1 Bedömning av andel med en normalfördelningsplot

Vid ett fd. sågverk är jorden förorenad av dioxin. Riktvärdet är satt till 20 ng/kg och frågan är hur stor andel av området som är förorenad över riktvärdet. Samtliga provpunkter har placerats ut slumpmässigt och varje prov representerar nivån 0 - 1 m under markytan.

Steg 1: Data plottas i en lognormalfördelningsplot, se nedan. De 25 datapunkterna följer inte en rät linje, vilket indikerar att data inte är lognormalfördelade. En lognormal fördelningsmodell kan därför inte användas för att skatta andelen av området som är förorenad över riktvärdet (det behövs inte heller eftersom vi har relativt gott om data).

Steg 2: Andelen som är under 20 ng/kg läses av på grafens y-axel. Grafen indikerar att ca 67 % av området har en halt under 20 ng/kg, vilket betyder att ca 33 % av området överskrider 20 ng/kg.

Observera att andelen 33 % bara gäller för den skala som proverna representerar (*support*).



Exempel på lognormalfördelningsplot av en verklig datamängd från ett dioxin-förorenat område. Grafen indikerar att ca 67% av området (röd cirkel) har halter under riktvärdet 20 ng/kg, dvs. att 33% av området är över riktvärdet.

4.2 Bedömning av andel med statistisk fördelning

För att avgöra vilken statistisk fördelning som ett stickprov kommer ifrån kan man utföra statistiska tester. När man på detta sätt identifierat den lämpligaste fördelningen skattar man även fördelningens parametrar. Då är hela fördelningen definierad. Med hjälp av fördelningen kan man sedan beräkna den percentil (i fördelningsmodellen) vid vilken en viss referenshalt överskrids. Om riktvärdet överskrids vid 90-percentilen kan detta översättas som att 10% av ytan är förorenad med halter över riktvärdet¹⁵.

Här beskrivs kortfattat metoder som är lämpliga att använda om data följer en normalfördelning, en lognormalfördelning eller en gammafördelning.

Normalfördelning

De parametrar som bestämmer normalfördelningens utseende är medelvärdet och standardavvikelsen. För normalfördelad data är den bästa skattningen av dessa parametrar det medelvärde och den standardavvikelse som beräknas från stickprovet. Både medelvärde och standardavvikelse beräknas enkelt i Excel.

¹⁵ Notera att percentilernas värden bara gäller för den aktuella provskalan, dvs. provernas storlek. Vid samlingsprov över större jordvolymer kommer den statistiska fördelningen, och därmed percentilerna, att bli annorlunda än för små provvolymer.

Exempel 4-2. Bedömning av andel med en normalfördelning

Stickprovet för arsenik som beskrivs i exempel 3-1 har befunnits vara normalfördelat. Stickprovets medelvärde och standardavvikelse är 56 mg/kg respektive 18 mg/kg. Riktvärdet för mindre känslig markanvändning är 40 mg/kg.

I Excel beräknas sannolikheten (p) att halten i en slumpvis vald punkt är lägre än riktvärdet med följande funktion:

=NORMFÖRD(40;56,1;18,4;SANT)

Resultatet blir 0,19. Andelen av området med halter över riktvärdet beräknas som $1-p$, i det här fallet $1 - 0,19 = 0,81$, dvs 81% av området har halter som överskrider 40 mg/kg. Detta kan översättas till volym förorenad jord, förutsatt att områdets storlek är känd.

Lognormalfördelning

För lognormalfördelad data är skattningarna av fördelningens parametrar, medelvärde och standardavvikelse, desamma som för normalfördelad data, men beräkningarna görs på logaritmerade data (med den naturliga logaritmen, \ln). Om man dessutom logariterar den referenshalt man vill jämföra mot så kan man direkt använda sig av normalfördelningen i Excel på samma sätt som redovisades i exempel 4-2. Annars använder man lämpligen lognormalfördelningen i Excel, se exempel 4-3 nedan.

Exempel 4-3. Bedömning av andel med en lognormalfördelning

Stickprovet för arsenik som beskrivs i exempel 3-8 har befunnits vara lognormalfördelat. Stickprovets medelvärde och standardavvikelse på logskalan är 2,61 respektive 0,94 vilket beräknas genom att logaritmera stickprovets mätvärden. Riktvärdet för mindre känslig markanvändning är 15 mg/kg.

Först logaritmeras riktvärdet 15, vilket blir 2,71. I Excel beräknas sannolikheten (p) att halten i en slumpvis vald punkt är lägre än riktvärdet med följande funktion:

=NORMFÖRD(2,71;2,61;0,94;SANT)

Andelen av området med halter över referenshalten 15 mg/kg beräknas som $1-p$, i det här fallet $1 - 0,54 = 0,46$, dvs 46% av området har halter som överskrider 15 mg/kg.

En variant är att använda funktionen för lognormalfördelningen i Excel tillsammans med referenshalten i ursprungsskalan:

=LOGNORMFÖRD(15;2,61;0,94)

Resultatet av denna beräkning blir detsamma. Andelen 46% kan översättas till volym förorenad jord, förutsatt att områdets storlek är känd.

Gammafördelning

Hur parametrarna skattas för en gammafördelning beskrivs i USEPA (2007) och kan beräknas i ProUCL. Det kan vara värt att notera att gammafördelningens parametrar betecknas α och β i Excel men k och θ i ProUCL.

4.3 Bedömning av andel med betafördelning

Att skatta andelen av ett område med halter över ett riktvärde med hjälp av betafördelningen bygger på helt andra principer än de två tidigare beskrivna metoderna. I metoderna ovan är variabeln som beskrivs av fördelningen en *halt*. Används betafördelningen är istället variabeln som beskrivs just *andelen* av området som överskrider en viss referenshalt. Den variabeln kan kallas för x .

Det som betafördelningen kan tillföra jämfört med de andra metoderna är möjligheten att beskriva osäkerheten för x . Utöver en skattning av andelen av området som är förorenat över en referenshalt, så erhålls även osäkerheten i skattningen. Denna osäkerhet kan uttryckas på liknande sätt som ett konfidenstervall, men kallas i detta fall för *kredibilitetsintervall*.

En annan finess med betafördelningen är möjligheten att föra in subjektiva skattningar, eller som det kanske snarare bör benämnas: expertkunskap. *Bayesiansk* statistik tillåter att man för in subjektiva skattningar och att dessa kan uppdateras med mätdata. Ofta görs en distinktion mellan klassisk statistik och Bayesiansk statistik, där den stora skillnaden ligger just i möjligheten att föra in subjektiva skattningar. Ju mer mätdata relativt sett, desto mer kommer resultatet att likna ett resultat som endast bygger på mätdata. Ibland kallas subjektiv data för mjuk data (eng. *soft data*) och mätdata för hård data (eng. *hard data*), vilket kan relateras till hur lätt det är att kvantifiera kunskapen. Mätdata är kvantitativ i sig, men förhandskunskap om ett förorenat område, t.ex. information om historik och föroreningsprocesser tillsammans med tidigare erfarenheter, är inte lika lätt att kvantifiera. En sådan kvantifiering av expertkunskap gör det alltså möjligt att inom Bayesiansk statistik använda den informationen i statistiska beräkningar. Kvantifieringen kan vara svår, och det finns mycket arbete gjort på hur översättningen kan göras.

Vidare, vilket redan har nämnts, tillåter betafördelningen att beräkningarna uppdateras på ett enkelt sätt när de första mätningarna eller ytterligare mätningar blir tillgängliga.

Att använda betafördelningen tillför helt andra möjligheter i skattningen av hur stor andel av ett område som är förorenat över en viss referenshalt. Den kräver heller inga antaganden om att uppmätt data måste följa en viss statistisk standardfördelning, vilket kan vara en tydlig fördel när en sådan inte kan identifieras eller antas. Betafördelningen tar också hänsyn till att det finns en större osäkerhet i skattningen ju mindre stickprovet är. Slutligen bör påpekas att metoden även kan användas helt utan subjektiva inslag, dvs. att möjligheten att föra in expertkunskap inte måste utnyttjas.

4.3.1 Betafördelningen

Betafördelningen är en kontinuerlig fördelning för en slumpmässig variabel x , och som definieras av parametrarna α och β , samt max och minvärdena för det intervall som x definieras över (här: $[0,1]$)¹⁶.

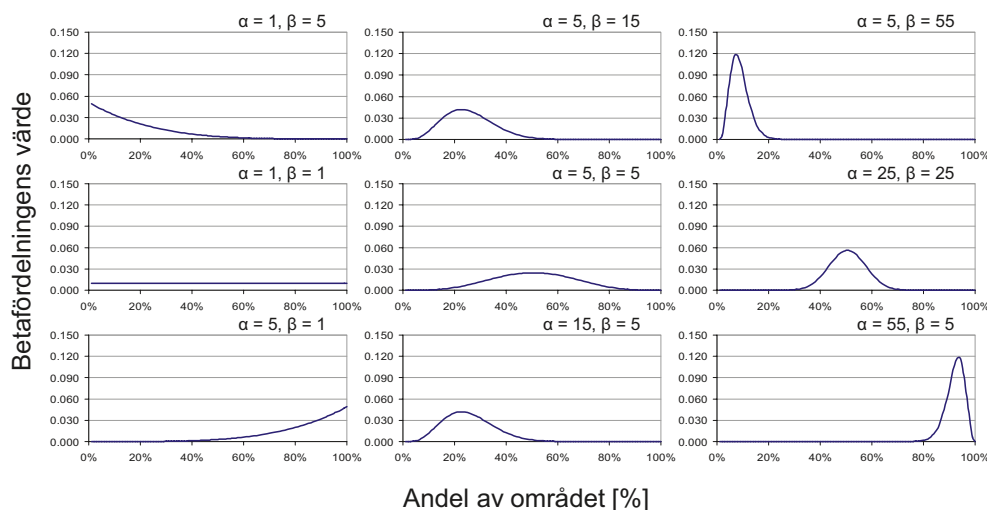
¹⁶ Fördelningen beskrivs utförligare i Bilaga C.

Variabeln x är i detta fallet andelen av området som är förorenad med halter över riktvärdet. Antag att stickprovet består av n oberoende observationer, av vilka f är antalet observationer över riktvärdet och g är antalet observationer under riktvärdet (dvs. $n = g + f$). Då gäller att fördelningens parametrar bestäms av

$$\alpha = f + 1 \quad \text{och} \quad \beta = g + 1,$$

förutsatt att förhandskunskap om x (dvs. förorenad andel av området) modelleras som $\alpha = \beta = 1$.

Parametrarna α och β styr betafördelningens läge och form. Ett flertal exempel av betafördelningar med $\min = 0$ och $\max = 1$, men med olika α och β illustreras i Figur 4-1. För stora stickprov blir osäkerheten i skattningen mindre än för små stickprov. I figuren visas också en fördelning där $\alpha = \beta = 1$, vilket alltså är ett sätt att modellera att det inte finns någon förhandskunskap alls om området. Antalet prover (observationer) från området är då noll och sammantaget är osäkerheten i skattningen som allra störst. Faktum är att betafördelningen är likvärdig med en uniform fördelning, dvs. den ser ut som en rät linje där alla värden mellan 0 och 1 är lika troliga. Ett alternativt sätt att modellera att man saknar förhandskunskap om området är att använda $\alpha = \beta = 0,5$, vilket gör att betafunktionen ser ut som en u-formad kurva. Fördelen med det sistnämnda sättet är att detta i mindre grad påverkar fördelningen när observationer tillförs.



Figur 4-1. Exempel på betafördelningar med olika värden på parametrarna α och β . X-axeln anger den förorenade andelen av området i procent och y-axeln anger betafördelningens värde, dvs. sannolikheten att en viss andel av området är förorenad.

Ju mer data som finns tillgängliga desto snävare blir fördelningen. Istället för att prata om konfidensintervall som man gör i samband med osäkerheter i skattning av medelhalt, så pratar man i dessa sammanhang om ett kredibili-

tetsintervall. Detta skrivs på liknande sätt: UCL för *Upper Confidence Limit* respektive UCL för *Upper Credibility Limit* och skillnaden är mest filosofisk.

I Figur 4-1 är det tydligt att ju större stickprov desto smalare och toppigare blir betafördelningen. Betrakta två av fördelningarna i Figur 4-1, t.ex. den där $\alpha = 5$ och $\beta = 15$ ("A") och den där $\alpha = 5$ och $\beta = 55$ ("B"). I fördelning "A" ges det mest troliga värdet av betafördelningen (fördelningens *mode*, se Bilaga C) av $4/(4+14) = 0,22$. Det 90%-iga kredibilitetsintervallet ligger mellan 0,11 and 0,42. I "B" är fördelningens *mode* 0,07 med ett kredibilitetsintervall mellan 0,03 och 0,15. Bredden på kredibilitetsintervallet är betydligt mindre i fallet B än i A, dvs. osäkerheten i skattad andel är lägre.

4.3.2 Användning av betafördelningen

Betafördelningens värde för valfri percentil kan enkelt beräknas i Excel, t.ex. gränserna för det 90%-iga kredibilitetsintervallet. Den undre gränsen beräknas i Excel som =BETAINV(0.05; α ; β) i Excel och motsvarande för den övre gränsen =BETAINV(0.95; α ; β).

Exempel 4-4. Bedömning av andel med hjälp av betafördelning

Stickprovet med arsenikdata som beskrivs i Exempel 3-1 Beskrivande statistik innehåller totalt 12 observationer. Två av dessa ligger under riktvärdet på 40 mg/kg och tio stycken har halter som överskrider riktvärdet. Detta innebär att $f = 10$ och $g = 2$, vilket ger att parametrarna för betafördelningen är $\alpha = 11$ och $\beta = 3$.

I Excel beräknas det mest troliga värdet på den andel (x) av området där halterna överskrider riktvärdet. Beräkningen görs med uttrycket för *mode* i Bilaga C:

$$\hat{x} = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{11 - 1}{11 + 3 - 2} = 0,83$$

Detta innebär att 83% av området bedöms vara förorenat över riktvärdet. För att räkna ut ett 90%-igt kredibilitetsintervall används funktionen "=BETAINV(sannolikhet; α ; β)" i Excel. De två gränserna för intervallet beräknas som "=BETAINV(0,05; α ; β)" respektive "=BETAINV(0,95; α ; β)". Resultatet blir 59% respektive 93%. Detta innebär att med sannolikheten 0,95 så är den förorenade andelen massor lägre än 93% av ytan, och med sannolikheten 0,05 så överskrider den förorenade andelen massor 59% av ytan. Den aktuella betafördelningens utseende visas i Figur 4-2 (A). Med hjälp av uppgifter om områdets storlek kan dessa resultat översättas till mängd förorenad jord.

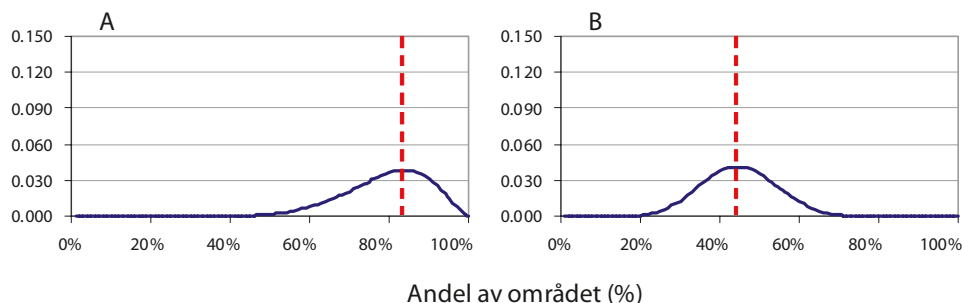
Exempel 4-5. Bedömning av andel med hjälp av betafördelning

Stickprovet för arsenik som beskrivs i exempel 3-8 innehåller totalt 25 observationer. Fjorton av dessa ligger under riktvärdet på 15 mg/kg och elva stycken har halter som överskrider 15 mg/kg. Detta innebär att $f = 11$ och $g = 14$, vilket ger att parametrarna för betafördelningen är $\alpha = 12$ och $\beta = 15$.

I Excel beräknas det mest troliga värdet på den andel (x) av området där halterna överskrider riktvärdet, på motsvarande sätt som i föregående exempel:

$$\hat{x} = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{12 - 1}{12 + 15 - 2} = 0,44$$

Beräkningen visar att 44% av området bedöms vara förorenat över riktvärdet. Ett 90%-igt kredibilitetsintervall beräknas på samma sätt som i föregående exempel. De två kredibilitetsgränserna för intervallet blir då 29% respektive 60%. Med 90% sannolikhet ligger den verkliga andelen i intervallet 29-60%. Den aktuella betafördelningens utseende visas i Figur 4-2. Figur 4-2 (B).



Figur 4-2. Betafördelningens utseende för exempel 4-4 (A) och exempel 4-5 (B). De streckade linjerna markerar det mest troliga värdet (mode) i respektive fördelning.

4.3.3 Att inkludera förhandskunskap i en betafördelning

Betafördelningen kan definieras enbart med hjälp de prover man har tagit på ett område utan någon prior kunskap (förhandskunskap), men den kan också definieras utifrån en subjektiv bedömning som görs av utredaren och som baseras på hennes/hans kunskap om området och erfarenheter från liknande områden

För en subjektiv skattning av betafördelningens parametrar kan två olika tillvägagångssätt användas (Rosén et al. 2008b):

- Skatta hur många indikatorer¹⁷ som den tillgängliga informationen motsvarar, samt hur många av dessa som talar för att området är förorenat (f) respektive rent (g).
- Skatta den mest troliga, den lägsta rimliga ($P_{0,5}$) och den högsta rimliga ($P_{9,5}$) förorenade andelen av området, baserat på tillgänglig information.

Det andra tillvägagångssättet är förmodligen det mest intuitiva, men percentilerna måste då översättas till värden på α och β . Detta görs enkelt i en programvara som t.ex. Crystal Ball (Oracle, 2007), i annat fall måste ett ekvationssystem lösas för att beräkna priorvärdena på α och β . Det finns varianter på de två tillvägagångssätten, t.ex. att det mest troliga värdet anges i kombination med att man skattar 90-percentilen. Fördelen med detta, förutom att det kan vara mer intuitivt i vissa applikationer, är att endast en ekvation måste lösas (Norberg, 2008).

¹⁷ En indikator är något som pekar på att området är rent eller förorenat. Ett enskilt prov kan motsvara en indikator.

Exempel 4-6. Inkludera förhandskunskap i en betafördelning

Området som refereras till i exempel 3-1 inspekterades och inventerades noggrant före provtagningen. Man bedömer att informationen som framkom motsvarar ca 4 prover i området och man skattar att tre av dessa bör representera halter över riktvärdet på 40 mg/kg. Det fjärde provet bör representera en halt lägre än riktvärdet. Det mest troliga är alltså att $\frac{3}{4}$ av ytan uppvisar halter över riktvärdet. Baserat *endast* på förhandskunskap kan betafördelningens parametrar då beräknas genom att ansätta att det mest troliga värdet (*mode*) = $\frac{3}{4}$ ¹⁸. Uttrycket för det mest troliga värdet (*mode*) är (se även Bilaga C):

$$\hat{x} = \frac{\alpha - 1}{\alpha + \beta - 2} = 0,75$$

Förhandskunskapen motsvarar 4 prover vilket ger att $\alpha + \beta = 4$. Dessa två ekvationer ger sammantaget att:

$$\alpha = 2,5 \quad \text{och} \quad \beta = 1,5.$$

Utifrån förhandskunskapen antas att 75% av området är förorenat över riktvärdet. Gränserna för ett 90%-igt kredibilitetsintervall beräknas på samma sätt som i tidigare exempel, för den nedre gränsen: " $=\text{BETAINV}((0,05; \alpha; \beta))$ ". Detta ger resultatet 24%. Kredibilitetsintervallets övre gräns blir 94% (notera att den mest troliga andelen ligger utanför intervallet, vilket beror på fördelningens skeva form). Betafördelningen kan ses i Figur 4-3 nedan.

Uppdatering av förhandskunskapen

Uppdateringen av priorskattningarna av α - och β -parametrarna görs helt enkelt genom att addera nya observationer:

$$\alpha' = \alpha + f \quad \text{och} \quad \beta' = \beta + g.$$

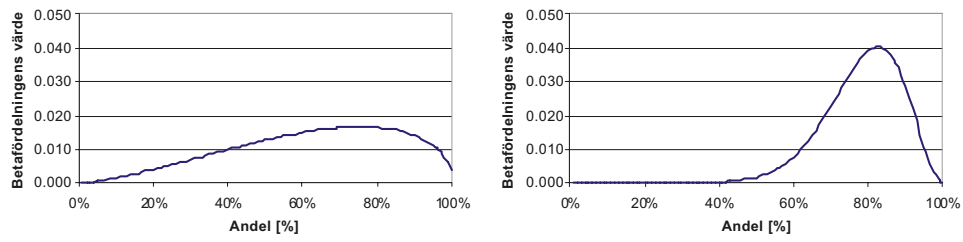
Exempel 4-7. Uppdatering av förhandskunskapen

Den priorskattning av betafördelningens parametrar som gjordes i exempel 4-6 uppdateras med resultaten från provtagningen som resulterade i $f = 10$ och $g = 2$. Parametrarna för priorfördelningen var $\alpha = 2,5$ och $\beta = 1,5$. Parametrarna för den uppdaterade betafördelningen blir då $\alpha' = 2,5 + 10 = 12,5$ och $\beta' = 1,5 + 2 = 3,5$. Det mest troliga värdet på andelen x beräknas som tidigare:

$$\hat{x} = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{12,5 - 1}{12,5 + 3,5 - 2} = 0,82$$

Gränserna för kredibilitetsintervallen beräknas till 60% respektive 92%. Genom att lägga till förhandsinformation har det troligaste värdet på andelen ökat och bredden på kredibilitetsintervallet minskat. Den aktuella betafördelningens utseende visas i Figur 4-3.

¹⁸ I princip skulle man även kunna ansätta att fördelningens medelvärde är lika med 0,75 med ekvationen given i Bilaga C. Detta skulle ge att parametrarna α och β blir 3 respektive 1. Fler exempel finns i Norberg, 2006.



Figur 4-3. Kurvan till vänster visar betafördelningen baserad på priorskattningen i exempel 4-6. Kurvan till höger visar betafördelningen efter uppdatering med nya data, dvs. priorskattningen + provtagningen (exempel 4-7).

5 Steg 3: Bedömning av rumslig korrelation

Rumslig korrelation kan föreligga vid förorenade områden på grund av att föroreningars depositions- och spridningsprocesser i markmiljön ofta kan ha ett rumsligt uppträdande. Detta medför att det kan finnas ett släktskap mellan uppmätta halter i närliggande provpunkter, vilket benämns som rumslig korrelation. Närvaron av rumslig korrelation möjliggör användningen av rumsliga interpolationsmodeller, vilka i sin tur kan användas för att avgränsa områden för efterbehandling eller för beräkningar av volym förorenade massor. Rumslig korrelation kan kvantitativt bekräftas genom de metoder som beskrivs nedan, men ofta är en visuell analys av ett korrelogram eller ett variogram tillräckligt.

Den viktigaste frågeställningen som bör besvaras i Steg 3 är om det är rimligt att gå vidare med ett antagande om rumslig korrelation. Den bedömningen bör grundas på en konceptuell idé om föroreningssituationen och kompletteras med visuell analys av data och eventuellt även statistiska tester. Endast om det finns goda skäl att anta en rumslig korrelation på en skala som är rimlig för interpolationens syfte, är det befogat att gå vidare till Steg 4: Interpolation.

5.1 Hänsyn till förekomst av rumslig korrelation vid utvärdering av data

Om det föreligger rumslig korrelation så behöver detta inte nödvändigtvis påverka tidigare beskrivna steg i ramverket vad gäller statistiska jämförelser mellan data (en representativ halt, t.ex. UCLM) och riktvärde (Brus & de Gruijter, 1997). Däremot skulle det kunna orsaka problem vid hypotesprövning och beräkning av stickprovsstorlekar, i alla fall enligt vissa författare. Eftersom prover som samlats in inom korrelationslängden¹⁹ förväntas vara mer lika så är det inte troligt att de är oberoende, åtminstone inte ur halt-synpunkt. Detta innebär enligt Griffith (2005) att varje enskilt prov inte kan tillgodoräknas en hel frihetsgrad för signifikanstester. Metoder som är vanligt förekommande för att testa och skatta nödvändig stickprovsstorlek för att kunna fatta beslut med viss bestämd felrisk (USEPA, 2006), skulle då påverkas av detta, vilket skulle kräva korrigerings till ett korrekt antal frihetsgrader.

I den statistiska litteraturen inom miljöområdet är man oense i vilken grad rumsliga korrelationer är ett problem eller ej, se exempelvis Brus & de Gruijter (1997). Några valmöjligheter finns dock om man anser att rumslig korrelation är ett problem och man önskar kompensera för detta. Ett sätt är

¹⁹ Korrelationslängden är det avstånd inom vilket det förekommer rumsligt beroende (släktskap) mellan föroreningshalterna.

att ta bort data från vissa provpunkter så att inga observationer används där provpunkterna ligger inom korrelationslängden i förhållande till varandra. Detta har dock tydliga nackdelar eftersom man då minskar styrkan i sina beräkningar och kastar bort värdefull insamlad data. En annan möjlighet är att korrigera (dvs. minska) antalet frihetsgrader för ett test, men återigen så innebär detta att man sänker styrkan i sina beräkningar, kanske mer än önskat av beslutsfattaren. Ett tredje alternativ är helt enkelt att samla in mer data för att kompensera för den grad av korrelation man observerat i den ursprungliga datamängden.

Rumsliga funktioner som är relevanta för miljöbedömningar inkluderar lokala och globala mått på rumslig korrelation i data, hypotestest som hanterar korrelation samt metoder för att utvärdera antagande om stationaritet. Antagande om stationaritet utgår från att data följer en likadan sannolikhetsfördelning överallt i rummet och är ett grundantagande för att geospatiala interpolationsmetoder skall kunna användas.

5.2 Metoder för bedömning av rumslig korrelation

5.2.1 Experimentella variogram

Ett vanligt sätt att bedöma om det föreligger rumslig korrelation är att visuellt analysera ett experimentellt semivariogram. Detta är inget statistiskt test i formell mening utan innebär att data plottas på ett speciellt sätt och att plotten kan användas för att bedöma om det är ett rimligt antagande att det finns rumslig korrelation. Variogram avbildar kovariansen²⁰ för en variabel. Variabeln av intresse är i denna rapport halten av ett ämne. Alla datapunkter delas in i par som kopplas till avståndsklasser. Avståndsklasserna plottas på x-axeln och variansen inom varje avståndsklass på y-axeln. Variogram plottar semivariansen²¹ som ett mått på autokorrelationen²².

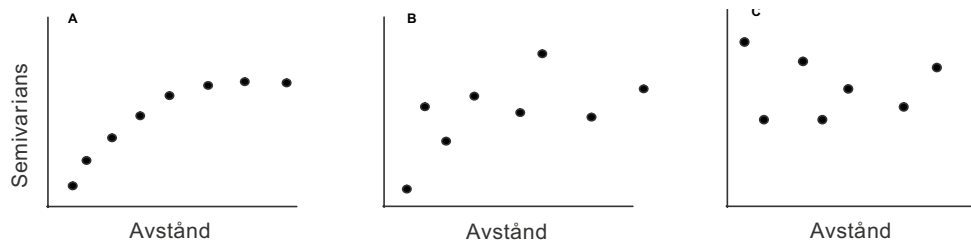
Exempel 5-1. Experimentella variogram

I ett experimentellt variogram plottas semivariansen på y-axeln mot x-axelns avståndsklasser. Dessa avståndsklasser kallas för laggar (*eng. lags*). Ett idealt experimentellt variogram liknar diagram A nedan där semivariansen hos datapar är liten på små avstånd och successivt ökar med avståndet. Experimentella variogram uppvisar dock ofta *inte* det monotont ökande utseende som ofta visas i bokexempel utan utseendet är ofta betydligt mer svårtolkat, som i diagram B och C nedan. Vidare kan variogrammen påtagligt ändra utseende beroende på hur avståndsklasserna subjektivt väljs.

²⁰ Kovariansen är ett statistiskt mått på samvariationen mellan två variabler.

²¹ Semivariansen är variansen mellan två punkter separerade av ett visst avstånd. Detta avstånd kallas *lag*.

²² Autokorrelation innebär att ett samband föreligger mellan observationer i tiden eller rummet.



Exempel på experimentella variogram, med varierande grad av rumsligt beroende.

En visuell inspektion av experimentella variogram är till stor nytta för att avgöra om det finns någon rumslig korrelation i data. Variogram som tyder på att det inte finns någon rumslig korrelation alls uppvisar ingen skillnad i varians med olika avstånd, såsom är fallet i diagram C i figuren.

5.2.2 Korrelogram

Korrelogram kan också användas för att utvärdera rumslig korrelation hos en variabel. All data är indelad i par som definieras i avståndsklasser på x-axeln och ett mått på korrelationen för varje avståndsklass på y-axeln. Korrelogram plottar ett mått på autokorrelationen på y-axeln, t.ex. Geary's C (Geary, 1954) eller Moran's I (Moran, 1950), se exempel 5-2 nedan.

5.2.3 Konfidensintervall och signifikanstester

Konfidensintervall och signifikanstest finns utvecklade för de olika typer av plottar som omnämns ovan (se exempel 5-2 nedan). Test som beräknar variansen under antagande om slumpmässighet (dvs. att ingen rumslig korrelation föreligger) kan avgöra om det föreligger en signifikant nivå av rumslig struktur i data för de valda avståndsklasserna (Cliff & Ord, 1981). Det faktum att samma data används i flera av avståndsklasserna påverkar dock oberoendet och kräver en så kallad *Bonferronikorrektion* för att korrigera sannolikhetsnivån för antalet avståndsklasser. Mer om detta finns i Oden (1984).

Exempel 5-2. Moran's I korrelogram och signifikanstest

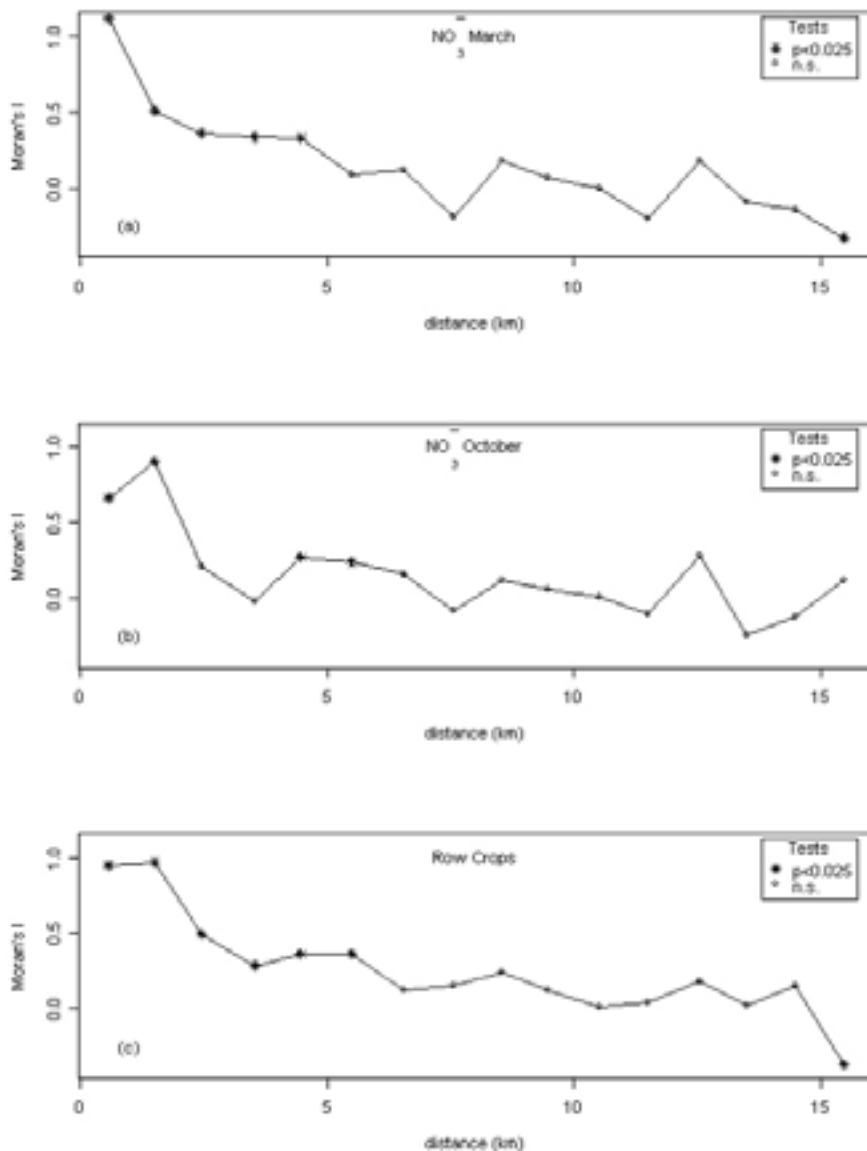
Exemplet här är från Golden et al. (2009) där man undersökt den rumsliga korrelationen för halten av nitrat i Cayuga Lakes avrinningsområde (New York). Jordbruk orsakar ofta nitratläckage och det rumsliga uppträdet av nitrat kan stämma väl överens med procentandelen av sk. radgrödor i markanvändningsdata.

Moran's I korrelogram genererades för nitrathalter för två enskilda provtagningstillfällen i mars och oktober. Moran's I statistika beräknades för varje avståndsklass, där varje klass är 1 km. De övre diagrammen i figuren visar den rumsliga strukturen för dessa provtagningstillfällen för de 15 första avståndsklasserna. Stickprovsstorlekarna i dessa avståndsklasser varierade mellan 28 och 92.

För varje avståndsklass har ett tvåsidigt signifikanstest utförts för att undersöka huruvida nollhypotesen, att ingen rumslig korrelation föreligger, kunde förkastas. Korrelationer beräknade med Moran's I varierar vanligen mellan 1 (positiv korrelation) och -1 (negativ korrelation), även om värden utanför detta intervall är möjliga.

I det övre diagrammet i figuren visas att vid provtagningstillfället i mars kunde man observera en signifikant positiv rumslig korrelation i de fem första avståndsklasserna och att en minskning av korrelationen är tydlig med ökande avstånd. För provtagningstillfället i oktober (diagrammet i mitten) observerades att fyra av de sex första avståndsklasserna uppvisar en signifikant positiv rumslig korrelation.

Det nedersta diagrammet i figuren visar ett Moran's I korrelogram för procentandelen radgrödor, vilket uppvisar en liknande grad och utbredning av rumslig korrelation som nitrathalterna vid de två provtagningstillfällena i mars och oktober. De första sex avståndsklasserna uppvisar en signifikant positiv rumslig korrelation. Återigen uppvisar data en nedgång av positiv korrelation efter de första 5-6 km, varefter man inte hittar någon signifikant positiv rumslig korrelation.



Exempel på korrelogram där Moran's I plottats. Den översta bilden visar korrelogrammet för nitrathalter mätta under mars, den mittersta nitrathalter mätta under oktober och den undre procentandelen radgrödor i landskapet. Punkter i diagrammen markerade som stjärnor betecknar att nollhypotesen (dvs. att det *inte* föreligger någon rumslig korrelation) kunde förkastas, dvs. stjärnorna är en indikering av rumslig korrelation.

5.2.4 Manteltest för rumslig korrelation

Manteltestet är ett annat tillvägagångssätt för att testa rumslig korrelation (Mantel, 1967). Tillämpat på förorenade områden så konstruerar Manteltestet två matriser där en innehåller all möjliga $n(n-1)/2$ datapar insamlade på området med ett visst avstånd, och den andra innehåller skillnaden i föroreningshalt för varje datapar. Produkten av matriserna är standardiserad från -1 till 1 och summeras för att beräkna Mantel-statistikan. Nollhypotesen är oberoende av de två matriserna och kan testas genom en t-testapproximation för typiska stickprovsstorlekar eller genom ett randomiseringstest när stickproverna är små eller väldigt stora.

5.2.5 Stationaritetstester

De tester som omnämns ovan, kräver att antagandet om *stationaritet* är uppfyllt inom undersökningsområdet. En stokastisk process är stationär om dess egenskaper är oberoende av var i det definierade området man befinner sig. Den vanliga formen är "svag" stationaritet, vid vilken medelvärde och varians för föroreningen inte varierar på ett *systematiskt* sätt inom området. För att uppfylla detta antagande är det viktigt att välja avgränsningen av området på ett sätt som avspeglar depositions- och spridningsprocesserna för föroreningen.

För större områden är beräkningar av global variografi (såsom Geary's C och Moran's I) otillräckligt för att utvärdera stationaritet eftersom dessa beräknar medelvärden över delområden och kan på så sätt dölja lokala skillnader (Anselin, 1995). Local Indexes of Spatial Association (LISA) är användbart för att identifiera tydliga skillnader i värden på Geary's C och Moran's I för mindre lokala områden (Anselin, 1995). Genom att plotta och analysera LISA-kartor kan man identifiera lokala områden med ett distinkt rumsligt uppträdande, vilket på så sätt kan förbättra indelningen av ett stort område i delområden innan man analyserar variografien och utför sin interpolation. Den här typen av kartor kan genereras i exempelvis programvaran SADA (2008).

6 Steg 4: Interpolation

Interpolation är ett viktigt verktyg för att exempelvis kunna avgränsa delområden med förhöjda halter, beräkna miljö- och hälsorisker och även för att kunna visualisera undersökningsresultat. Frågeställningarna om rumslig korrelation i Steg 3 måste dock vara besvarade innan man går vidare och utför interpolation av sina data. Endast i de fall en rumslig korrelation kan påvisas och förklaras på tillräckligt stora avstånd bör interpolation användas.

För att interpolation skall kunna utföras och generera ett rimligt resultat måste en lämplig interpolationsmetod väljas och den matematiska modell som ligger till grund för interpolationen valideras. Interpolationen kan även ge värdefull vägledning till var eventuella kompletterande provtagningar bör utföras.

6.1 Interpolationsmetoder

De enklaste interpolationsmetoderna använder sig av deterministiska rumsliga prediktioner, t.ex. avståndsviktad medelvärdesinterpolation, lokala polygoner eller triangulering samt spline-interpolation (Lam, 1983). Dessa metoder skattar värdet i punkter som inte provtagits genom en viktning grundad på punktens placering och mätvärden i närliggande provpunkter. Dessa metoder saknar dock beskrivning av osäkerheten i skattningen, men de kan tillämpas över hela området (globalt) eller i ett lokalt ”rörligt fönster” inom området. Den här typen av metoder kan producera bra resultat för vissa datamängder, ibland t.o.m. bättre resultat än mer sofistikerade metoder.

Geostatistiska interpolationsmetoder är av en annan typ som nyttjar geostatistik, en form av rumslig statistik som kan ta hänsyn till variabilitet och rumslig korrelation i data. Med hjälp av detta kan man göra skattningar av den rumsliga fördelningen. Geostatistik används ibland som verktyg i samband med riskbedömning och för utformning av åtgärdslösningar (Thayer et al., 2003). De vanligaste geostatistiska metoderna är olika typer av *kriging* men det finns även metoder som bygger på *geostatistisk simulering*. Både kriging och geostatistisk simulering är metoder som använder sig av rumsliga parametrar från en variogrammodell.

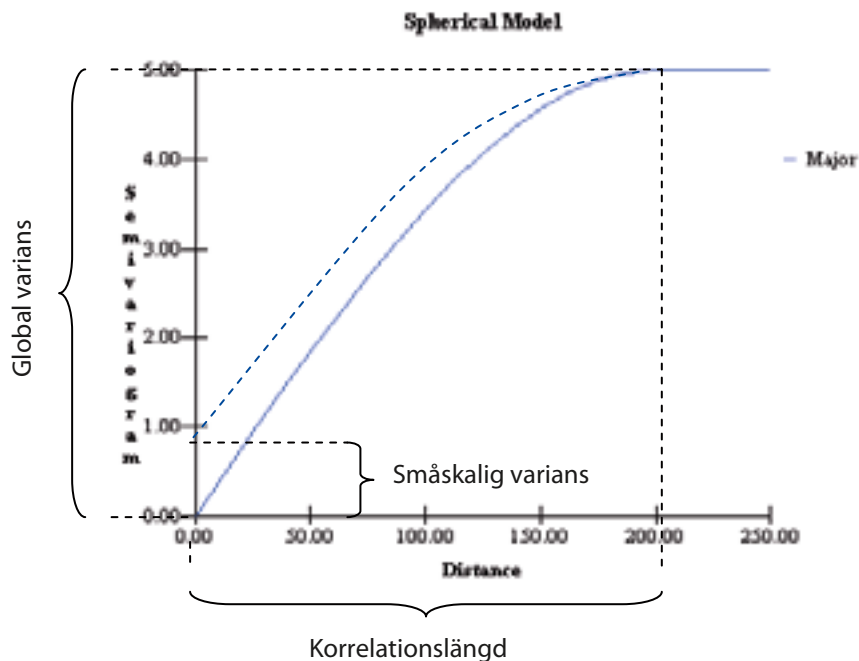
6.1.1 Variogrammodellen

En variogrammodell behövs för att kunna utföra mer avancerade interpolationer som t.ex. kriging. Modellen baseras på ett experimentellt variogram (se avsnitt 5.2.1) för att bedöma hur den rumsliga korrelationen ser ut. En variogrammodell passas till det experimentella variogrammet så att överensstämmelsen blir så bra som möjligt. I ramverket beskrivs framtagning och inspektion av det experimentella variogrammet som ett steg skilt från variogramanalysen (dvs. passningen till en modell). I praktiken görs dock dessa två steg samtidigt i de flesta programvaror som utför variogramanalys.

De flesta variogrammodeller definieras av tre parametrar:

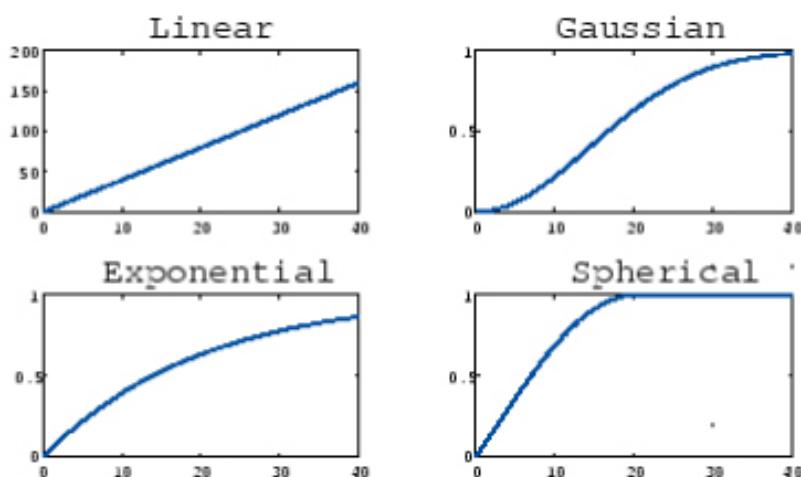
1. Korrelationslängd (eng. *range*),
2. Global varians (eng. *sill*)
3. Småskalig varians (eng. *nugget effect*),

De tre parametrarna illustreras i variogrammodellen i Figur 6-1. Man behöver dessutom ange vilken typ av matematisk modell som är lämplig som utgångspunkt för variogrammodellen, t.ex. *linjär*, *sfärisk*, *exponentiell* eller *gaussisk* modell, se Figur 6-2. Ibland kan det vara nödvändigt att kombinera flera modeller för att kunna efterlikna det experimentella variogrammet, så kallade *nested variograms*. Här kan också nämnas att den småskaliga variansen (*nugget effect*) ibland brukar hanteras som en egen variogrammodell och inte som en parameter.



Figur 6-1. Figuren visar en sfärisk variogrammodell, samt de tre parametrar som bestämmer formen: Sill är den globala variansen, korrelationslängden (*range*) är det avstånd inom vilket rumslig korrelation förekommer och nuggeteffekten beskriver variabiliteten på mycket korta avstånd.

I flera programvaror som hanterar variogram finns automatiska passningsrutiner tillgängliga, dvs. programvaran letar upp en variogrammodell som programmet bedömer bäst passar det experimentella variogrammet. Ett problem med automatiska passningsrutiner i programvaror som hanterar variogram är dock att det alltid går att anpassa en modell till data, även om modellen är orealistisk. Därför bör man vara extra försiktig om automatiska passningsrutiner används. Variogrammodellen kan dock testas med olika typer av valideringsförfaranden, se avsnitt 6.2.



Figur 6-2. Exempel på olika typer av variogrammodeller. Vid kriging och geostatistik simulering används den valda modellen för att hitta en optimal viktning av värden i varje definierad punkt rummet.

Data från förorenade områden har ofta en skev fördelning, varför data behöver logaritmeras innan variogramanalysen utförs för att resultaten skall bli rimliga. Att ta fram en variogrammodell som väl beskriver den rumsliga korrelationen kräver en hel del kunskap, erfarenhet och kanske även talang.

Variogram som tyder på att det inte finns någon rumslig korrelation alls uppvisar ingen skillnad i varians med olika avstånd. Ett sådant variogram brukar kallas för *pure nugget effect*-variogram, dvs. ett variogram där den globala variansen (*sill*) är lika med variansen på mycket små avstånd (*nugget effect*). Korrelationslängden är därmed är noll och variogrammodellen blir en horisontell linje vid sill-variansen.

6.1.2 Kriging

Kriging är en interpolationsteknik som bygger på att man utnyttjar variogrammodellen av den rumsliga korrelationen, se ovan. Matematiskt är tekniken en linjär "minsta kvadrat"-metod som löser ett ekvationsystem genom att använda variogrammodellen. Genom att använda variogrammodellen kan de "bästa" värdena (dvs. halterna) beräknas för alla definierade punkter, även sådana där inga data finns. Punkterna definieras genom att området delas in i ett rutnät, vilket innebär att en bedömning måste göras av hur stora cellerna i rutnätet bör vara.

Kriging utförs under antagande om en Gaussisk process, vilket kräver normalfördelade data. För tillämpning på förorenad jord innebär detta att man oftast måste transformera data till en normalfördelning innan man utför kriging, samt transformera tillbaka resultatet efteråt. Det finns dock parameterfria krigingmetoder som inte kräver denna typ av transformation, s.k. indikator kriging. Sådana metoder kräver dock istället att data indelas i klasser (koncentrationsklasser) innan kriging utförs.

Resultaten som man får från en kriginganalys är:

1. Beräknade (predikterade) värden i alla definierade punkter inom undersökningsområdet, även de som inte provtagits, samt
2. en variansskattning av prediktionsfelet i varje punkt.

Variansskattningen kan tolkas som osäkerheten i en viss punkt (cell). Varje cell har sin egen variansskattning. Variansskattningen kan användas för att beräkna en sannolikhet att överskrida en referenshalt i en viss cell.

Flera typer av kartor kan genereras från en kriginganalys, t ex:

1. Kartor över den troligaste halten i varje punkt, dvs. 50-percentilen i fördelningen,
2. kartor över halter som inte kommer att överskridas med en viss sannolikhet (t.ex. 0,95), eller
3. kartor över sannolikheten att överskrida en viss referenshalt, t ex ett riktvärde i varje punkt.

Den vanligaste krigingkartan är den som visar den troligaste halten i varje enskild punkt. Här måste dock betonas att en sådan karta INTE ger en realistisk bild över föroreningshalternas variation i området. I krigingprocessen sker nämligen en utjämning av variansen i data, vilket leder till att krigingkartan får ett utjämnat utseende med lägre variabilitet än i verkligheten. För att få kartor med realistisk variabilitet krävs geostatistisk simulering.

6.1.3 Geostatistisk simulering

Geostatistisk simulering kan vara till stor nytta för att undersöka variabiliteten inom ett förorenat område. Vid geostatistisk simulering används samma typ av information som vid kriginganalys, dvs. en variogrammodell med ett antal rumsliga parametrar, med vars hjälp man konstruerar en sannolikhetsfördelning för halten i varje definierad punkt inom undersökningsområdet.

Skillnaden mot kriging är att vid geostatistisk simulering så görs en hypotetisk provtagning i punkter som saknar mätdata genom att slumpmässigt dra ett värde från de genererade sannolikhetsfördelningarna. En sådan hypotetisk provtagning i alla punkter resulterar i en så kallad *realisering*. Detta upprepas så att en mängd realiseringar skapas. Varje realisering är lika sannolik och utgör en realistisk bild av föroreningsituationen, förutsatta att variogrammodellen är realistisk. Geostatistisk simulering resulterar således inte i en enda karta, utan i en stor mängd kartor. Dessa kan sedan användas för att ta fram en sannolikhetsbaserade kartbilder vilka kan användas för att bedöma osäkerheter och risker. Ett intressant användningsområde är avgränsning av en förorening, där avgränsningen kan göras enligt en önskad säkerhetsnivå (man kan avgränsa föroreningen så att sannolikheten är exempelvis minst 50% att riktvärdet överskrids).

Kartorna som genereras med geostatistisk simulering bibehåller den övergripande rumsliga variabiliteten, men genom den slumpmässiga provtagningen undviks de utjämnings effekter (s.k. *smoothing effects*, variansreduktion) som

man ofta ser i krigingkartor. På så vis avspeglas haltvariationer på ett mer realistiskt sätt med simulerade kartor än med krigingkartor.

Om alla genererade realiseringskartorna läggs ovanpå varandra, för att identifiera det mest sannolika värdet i varje punkt, så erhålls i princip samma bild som vid kriging.

6.2 Modellvalidering

Interpolation kräver en modellutvärdering för att välja metod och rätt parameterutformning.

Korsvalidering är ett sätt att jämföra skillnaden mellan interpolerad data och uppmätt data (Legendre & Legendre, 1998). Genom att utföra korsvalidering kan olika modellantaganden för en interpolationsmetod jämföras eller olika interpolationsmetoder kan jämföras med varandra. Jämförelsen görs genom att i tur och ordning plocka bort verkliga mätvärden, interpolera fram ett värde för den punkten och därefter observera skillnaden mellan det värde som interpolationsmetoden ger och det faktiska mätvärdet i punkten. Ju närmare det uppmätta värdet som det interpolerade värdet hamnar, desto bättre kan metoden sägas vara. Vanliga kriterier vid utvärderingen är bland annat medelfelet och absoluta medelfelet.

Korsvalidering håller dock efterhand på att ersättas med mer sofistikerade modeller som bygger på *maximum likelihood* principer se t.ex. Pardo-Igúzquiza (1998).

6.3 Strategi för kompletterande provtagning

Det finns ett antal sekundära provtagningsstrategier, dvs. strategier för var kompletterande provpunkter bör placeras. Strategierna förutsätter att en rumslig korrelation finns och provpunkterna placeras ut med hänsyn till den geospaciala modell som arbetats fram på basis av tidigare provtagning. Den optimala placeringen av kompletterande provpunkter kan göra med olika syften, t ex tätare provtagning i delområden med för få prover, provtagning i delområden där de högsta halterna förväntas, provtagning i områden där variansen (osäkerheten) är störst eller provtagning där modellen är som mest osäker på att överskrida en referenshalt.

Det finns ett flertal olika kriterier för att avgöra när fler prover inte tillför mer information. Många är av statistisk typ, dvs. tills att osäkerheten är reducerad tillräckligt mycket, men det finns också mer ekonomiska kriterier som kopplar till den ekonomiska konsekvensen av att ta fel beslut.

I en ideal situation så skulle sekundära provtagningsstrategier tillämpas i realtid, dvs. att placeringen av ett nytt prov beräknades, data samlades in, resultatet lades in direkt i den geospaciala modellen och bästa placering av nästa provpunkt beräknades. Detta tillvägagångssätt är dock ovanligt. Istället behöver man oftast ta fram placeringen för ett antal olika punkter på en gång

för att kunna genomföra en samlad provtagningsomgång där proverna samlas in tillsammans och sedan analyseras. Då måste man istället simulera provtagningen av de nya provpunkterna, dvs. den första placeringen väljs ut och ett modellerat analysresultat används i modellen för att ta fram nästa provtagningspunkt, vilket repeteras fram till att antalet prover är tillräckligt. Även om tillförlitligheten reduceras vid simulerade värden så är tillvägagångssättet ändå värdefullt när man är tvungen att ta fram flera provplaceringar på en gång.

SADA (2008) har implementerat fem olika typer av sekundära provtagningsstrategier, där varje typ av strategi har olika mål: (1) *Adaptive Fill*, (2) *Estimate Rank*, (3) *Variance Rank*, (4) *Percentile Rank*, samt (5) *Uncertainty Rank*. Den första metoden behöver inte baseras på en geostatistisk modell utan kan användas för vilken interpolationsmodell som helst.

Målet med *Adaptive Fill* är att fylla ut områden som inte är provtagna lika tätt som andra områden, dvs. nya punkter placeras där befintliga provpunkter ligger långt från varandra. Metoden är enkel att tillämpa eftersom den inte kräver någon geostatistisk bearbetning av data.

Placeringen av nya provpunkter som baseras på *Estimate Rank* går ut på att placera dessa där högst halter predikteras i den geospatiala modellen. Det här är en typ av bekräftande strategi, dvs. provtagningen syftar att bekräfta förekomsten av *hotspots* och den tar ingen hänsyn till osäkerheten i skattningen av halter.

Strategin *Variance Rank* baseras på vanlig kriging där även variansen predikteras i alla gridpunkter tillsammans med koncentrationen. Dessa två mått definierar en fördelning av möjliga halter i varje punkt. Nya provpunkter placeras där modellvariansen är som störst. Detta resulterar därmed i en provtagningsstrategi som syftar till att reducera modellvariansen.

Percentile Rank baseras på att nya provpunkter placeras i punkter som har en potential att ha mycket höga halter. De geostatistiska skattningarna producerar, som tidigare nämnts, förutom den mest troliga halten även ett spann för hur halten kan variera. I den här strategin väljs de punkter ut där exempelvis 90-percentilen av predikterad halt har de högsta värdena.

Uncertainty Rank utgår, liksom *Variance* och *Percentile Rank*, från ett geostatistiskt tillvägagångssätt, men till skillnad från dessa båda metoder används ett beslutskriterium. Nya punkter placeras där modellen är som mest osäker huruvida halten överskrider ett specifikt haltkriterium (referenshalt) eller ej, dvs. där sannolikheten att överskrida referenshalten ligger runt 0,50. Den här strategin är framförallt tillämpbar för att avgränsa områden med halter över exempelvis ett riktvärde.

I de fyra sistnämnda strategierna kan ett andra kriterium användas som komplettent. Rent matematiskt kan nämligen den bästa placeringen av en ny provpunkt innebära att den nya punkten hamnar alldeles intill en redan provtagen punkt, vilket inte alltid är önskvärd praktiken. Som följd av detta kan ett minsta-avståndskriterium vara bra att använda som ett kompletterande kriterium. Ett sådant krav tvingar nya provpunkter att hamna på minst ett visst avstånd från alla andra punkter och ger således en bättre spridning av provpunkterna inom området.

7 Praktiska aspekter vid utvärdering av data

7.1 Statistiska populationer

Vid en statistisk undersökning studerar man en population bestående av element. En grundregel är att man ska ange vad ett element avser och vilka element som ingår i populationen (Blom, 1989). Vid undersökning av förorenad jord utgörs elementen av föroreningshalter med en definierad jordvolym. Populationen å sin sida utgörs av alla sådana föroreningshalter som förekommer i en avgränsad jordvolym, eller som kan kopplas till ett visst föroreningsutsläpp (eller annan händelse som givit upphov till föroreningen).

Ett förorenat område är i regel heterogent ur föroreningssynpunkt med delområden som kan vara opåverkade av föroreningar medan andra delar kan vara kraftigt förorenade. Det är därför lämpligt att gruppera mätdata efter egenskapsområden (delområden), se avsnittet om rumsliga och tidsmässiga avgränsningar i Norrman et al. (2009). Helst bör sådana egenskapsområden definieras före provtagningen, åtminstone preliminärt. Avgränsningen av egenskapsområden kan exempelvis baseras på:

- Historik om verksamheten på områden.
- Tidigare undersökningar.
- Geologiska förhållanden.
- Förväntade föroreningars uppträdande (spridningsmönster och spridningsvägar).
- Markanvändning, t.ex. byggnader.
- Visuella intryck (färg, vegetation m.m.).
- Saneringsaspekter.

Även uppmätta halter kan beaktas när man avgränsar ett egenskapsområde, om avgränsningen görs efter provtagningen. Här är det dock på sin plats med en varning: man bör aldrig basera en indelning i egenskapsområden enbart på uppmätta halter, åtminstone om stickprovet är litet – då kan stora fel begås. Orsaken är att det alltid finns en slumpeffekt i mätdata och en avgränsning som enbart grundas på data kan leda till helt felaktiga slutsatser.

Ibland kan det dock vara befogat att revidera den preliminära indelningen i egenskapsområdena efter provtagningen men då bör även andra aspekter än uppmätta halter vägas in, t.ex. tidigare verksamheter, geologiska förhållanden eller andra aspekter i den konceptuella modellen över området. Ett exempel på en sådan situation är då den statistiska analysen resulterar i märkliga statistiska fördelningar. Detta kan vara en indikation på att data kommer från fler än en population. På detta sätt kan en preliminär indelning av egenskapsområden bekräftas/vederläggas av data så att en slutlig indelning kan göras.

7.2 Hantering av datakluster

Det är vanligt att datamängder från förorenade områden innehåller datakluster, dvs. mätdata från provpunkter som förekommer i grupper genom förtätad provtagning. Detta kan vara ett stort problem vid den statistiska utvärderingen. Datakluster uppkommer ofta genom att så kallad *riktad provtagning* använts för att lokalisera delvolymmer med höga halter (*hotspots*) och där kompletterande prov sedan tagits, exempelvis med syftet att avgränsa föroreningen. Eftersom provpunkterna inte valts slumpmässigt kommer därmed den statistiska beskrivningen av området att få ett systematiskt fel, dvs. statistiken blir inte representativ för området. Vanligtvis blir det beräknade medelvärdet av mätdata då högre än områdets verkliga medelhalt. Detta kan i sin tur leda till att riskerna i området överskattas. Om beräknade medelvärden används för att uppskatta föroreningsmängderna kan även dessa överskattas.

Det bästa sättet att upptäcka om datakluster förekommer är genom att ta reda på vilket syfte de olika provpunkterna hade när de placerades ut. Om syftet med vissa punkter är att bekräfta en misstänkt förorening, i kombination med andra punkter som syftar till avgränsning, kan man vara ganska säker på att datakluster förekommer. Provpunkternas placering på en karta kan också vara till stor hjälp.

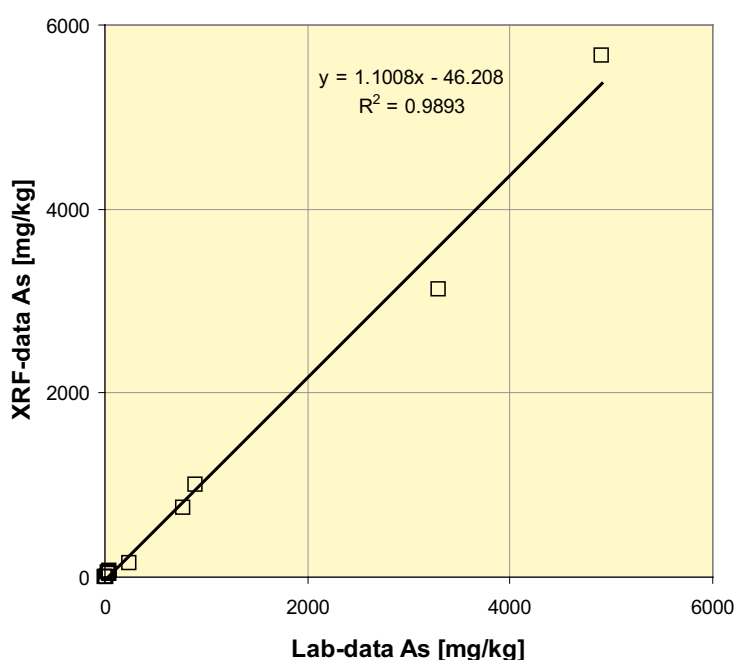
Det är i många fall möjligt att korrigera för datakluster, åtminstone delvis. Problemet kan minskas genom s.k. *declustering* (Myers, 1997; Deutsch & Journal, 1998), vilket innebär att de olika mätvärdena ges olika vikt vid den statistiska utvärderingen. Dessa vikter kan användas för att ta fram viktade medelvärden, viktade percentiler etc. Viktningen kan göras med olika metoder och beräkningarna kan utföras i exempelvis Excel eller med speciella programvaror.

7.3 Hantering av olika typer av data

En vanlig situation vid utvärdering av mätdata från förorenade områden är att data är av olika typer. Exempel på faktorer som kan bidra till detta är att provtagningarna kan ha genomförts vid olika tidpunkter, olika analysmetoder kan ha använts, data kan komma från olika laboratorier med olika rapporteringsgränser samt att delar av datamängden kan ha samlats in med annan mätteknik än vad som används på laboratoriet såsom med fältinstrument. Generella råd hur detta ska hanteras är svåra att ge eftersom situationerna kan skilja sig åt från fall till fall. I vissa fall kan det ändå vara rimligt att hantera data som ett enda stickprov, i andra fall bör istället separata statistiska analyser göras för de olika datamängderna. Ett angreppssätt kan vara att i ett första skede utföra separata statistiska analyser på de olika datamängderna. Om resultaten då blir likartade kan det vara möjligt att slå samman datamängderna.

Om datamängderna slås samman måste man beakta att variabiliteten inom respektive datamängd kan skilja sig åt beroende på olika noggrannhet. I vissa fall kan systematiska fel förekomma och då är det önskvärt att man justerar för sådana. Ett exempel på detta är när både laboratoriedata och data från fältinstrument som XRF förekommer. Då kan det vara möjligt att genom regressionsanalys hitta ett samband mellan laboratoriedata och fältmätningarna, och på så sätt justera fältdata. Detta kräver dock att båda mätmetoderna utförs på samma prover. Exempel på regressions samband mellan laboratoriedata och fältmätningar redovisas i **Fel! Hittar inte referenskölla.**

Regression arsenikdata



Figur 7-1 Enkel linjär regression mellan laboratoriedata och fältdata (XRF-data), totalt 10 mätvärden från ett förorenat område. För att justera fältdata kan regressionsekvationen eventuellt användas men sambandet vid låga halter bör först studeras närmare.

Här bör påpekas att vanlig regressionsanalys minimerar felet i y-led, vilket innebär att den osäkra variabeln (fältmätningarna) bör plottas på y-axeln och variabeln med större säkerhet (laboratoriedata) på x-axeln. Om man istället gör det omvända kommer man att få ett annat och mindre tillförlitligt regressions samband mellan fältdata och mätdata.

7.4 Data under detektionsgränsen

Ett vanligt inslag i analysresultat från förorenade områden är att vissa mätvärden ligger under detektionsgränsen²³, s.k. *non detects* (NDs). Det finns metoder och rekommendationer framarbetade för hur sådana datamängder skall hanteras i statistiska sammanhang. Detta omfattar bl.a. metoder för att beräkna medelvärde, UCL för medelhalten (UCLM), jämförvärden för bakgrundshalter samt identifiering av s.k. *outliers*, se USEPA (2007). De flesta statistiska metoder är dock svåra att använda om datamängden innehåller en stor andel NDs.

Tre typer av metoder för att hantera NDs kan särskiljas (Helsel, 1990; USEPA, 2007):

1. Enkla ersättningsmetoder (substitutionsmetoder).
2. Parametriska metoder.
3. Icke-parametriska metoder.

Ytterligare en metod är att utesluta alla mätvärden som ligger under detektionsgränsen men nackdelen är att värdefull information då går förlorad. Denna metod bör därför aldrig användas. De enkla ersättningsmetoderna är vanligast och innebär att alla mätvärden under detektionsgränsen sätts lika med:

- A. detektionsgränsen (DG),
- B. halva detektionsgränsen (DG/2), eller
- C. noll (0).

USEPA (2007) ger emellertid rekommendationen att inte använda dessa metoden när t.ex. UCLM95 skall beräknas för att de kan ge tveksamma resultat. I många fall kan dessa enkla metoder ändå vara tillräckliga, i varje fall om antalet NDs är litet. Vilken ersättningsmetod man väljer är delvis en smaksak men man bör vara medveten om att medelhalten överskattas med metod A men underskattas med metod C. Den sistnämnda metoden kan inte heller användas om man vill kunna logaritmera halterna, som exempelvis i en lognormalfördelningsplot. USEPA (2007) anger att metod B varit den vanligast fram till dess att ProUCL och andra programvaror gjort det möjligt att använda mer avancerade och bättre metoder.

Den andra gruppen av metoder är sk. parametriska metoder, vilka bygger på antagande om en viss statistisk fördelningsmodell. I ProUCL (2008) finns metoder för att extrapolera data som ligger under detektionsgränsen så att de följer en definierad statistisk fördelning (i ProUCL kan det göras för normal-, lognormal- och gammafördelningarna). Dessa metoder betecknas ROS-metoder (Regression on Order Statistics Estimation Methods). För att kunna använda dessa måste det dock finnas tillgång till tillräckligt många data över detektionsgränsen för att man ska kunna avgöra vilken statistisk fördelning som data följer. I statistikprogramvaror finns även andra parametriska meto-

²³ Kan också vara laboratoriets rapporteringsgräns. Beteckningen detektionsgräns används både för detektionsgräns och rapporteringsgräns i rapporten.

der, bl.a. maximum likelihood-metoder, som kan användas för beräkningar med datamängder som innehåller NDs. Sådana finns exempelvis i statistikprogrammet Minitab.

Icke-parametriska metoder kräver inget antagande om en viss statistisk fördelning. Enligt USEPA (2007) är sådana metoder att föredra om andelen NDs är större än 40%-50%. Ett begränsat antal icke-parametriska metoder finns implementerade i ProUCL.

I ProUCL (2008) finns det möjligheter att anpassa valet av statistisk metod baserat på om datasetet innehåller mätvärden under detektionsgränsen för att utföra "goodness-of-fit"-tester, tester för *outliers*, hypotesprövning, beräkning av UCLM95 samt jämförvärden för bakgrundshalter. Det finns även möjlighet att beräkna och spara extrapolerade värden på NDs med hjälp av ROS-metoder för att kunna exportera dessa till andra programvaror för andra beräkningar.

7.5 Duplikat

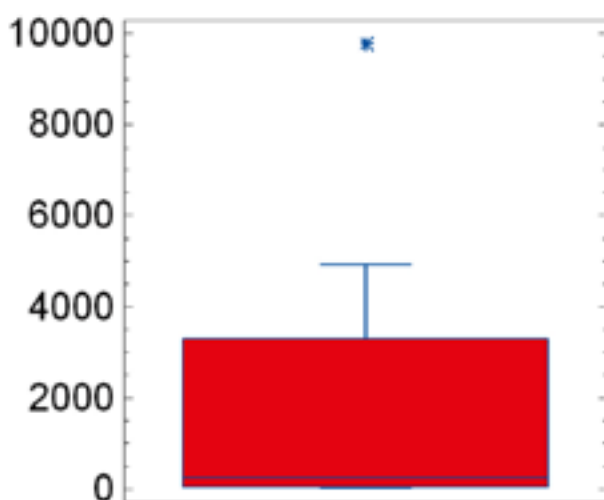
Som ett moment i kvalitetskontrollen analyseras ibland dubbelprover, s.k. duplikat. Ibland förekommer det även att samma provpunkt analyserats i flera undersökningar. Detta medför att för vissa provpunkter och djupnivåer kan det finnas två eller fler mätvärden. Skillnaden mellan mätvärdena ger då ett mått på variationen, eller osäkerheten, för den aktuella provpunkten. Dessa osäkerheter omfattar flera typer av provtagningsosäkerheter samt analysosäkerheter, se Back (2003).

Vid statistisk utvärdering av data finns det flera sätt att hantera duplikat. Det enklaste är att slumpmässigt välja ett av mätvärdena och bortse från duplikatet (eller duplikaten) i utvärderingen. Ett annat sätt är att betrakta provpunkterna med duplikat som datakluster. Genom att vikta de olika mätvärdena kan då en representativ datamängd skapas, se avsnitt 7.2.

7.6 Outliers

Inom statistiken är *outlier* är ett engelskt uttryck för ett mätvärde som numeriskt avviker från resten av datamängden (vedertaget svenskt begrepp saknas). Förekomst av *outliers* kan vara en indikation på att mätvärdet egentligen hör till en annan population, dvs. att en felaktig avgränsning av målpopulationen gjorts. Andra förklaringar är mätfel eller ren slump. Statistik som beräknas på datamängder med *outliers* kan ge vilseledande resultat. Det finns ingen entydig matematisk definition på när man ska betrakta ett mätvärde som en *outlier* utan i slutändan är det ett subjektivt val. Om man känner den statistiska fördelningen för målpopulationen är det dock möjligt att beräkna om ett mätvärde avviker signifikant från fördelningen. I statistiska programvaror brukar *outliers* identifieras genom någon typ av matematisk algoritm, bl.a. i så kallade *Box-plottar* (se Figur 7-2). I programvaran ProUCL finns funktioner för att testa om mätvärden är *outliers*.

Box-and-Whisker Plot



Figur 7-2 Exempel på *Box-and-Whisker* plot av ett stickprov från ett förorenat område (halt i mg/kg). Trolig *outlier* har markerats med en stjärna.

Vid förorenade områden kan det vara svårt att avgöra orsaken till en *outlier* eftersom man i de flesta fall kan förvänta sig en stor variation i data. Vid små datamängder med endast några enstaka mätvärden kan det vara näst intill omöjligt att bedöma orsaken till en *outlier*.

Sammantaget innebär detta att man aldrig bör utesluta data från den statistiska utvärderingen endast baserat på att mätvärdena är *outliers*. Överhuvudtaget är det kontroversiellt att utesluta *outliers* eftersom det kan leda till helt felaktiga slutsatser; vid förorenade områden en underskattning av hälso- och miljörisker. Om det finns faktorer som talar för felaktig avgränsning av målpopulationen eller mätfel kan man istället komplettera den statistiska utvärderingen med ett alternativt scenario där ett eller flera mätvärden utesluts. Man bör då tydligt redovisa båda scenarierna samt motiven. I de enstaka fall där *outliers* faktiskt måste uteslutas ur datamängden måste man alltid tala om att så skett och förklara varför.

7.7 Data med olika support

Med *support* menas den volym som ett prov representerar samt dess form och orientering. I Norrman et al. (2009) används begreppet provtagnings-skala som svensk term för detta. Observera att ett provs *support* inte avser själva provets volym och form utan den volym och form som provet representerar, vilket inte är samma sak. Exempelvis kan vikten på ett samlingsprov vara några kg medan provet kan representera en jordvolym på tiotals ton jord. Här bör påpekas att det inte är användaren som avgör vilken volym provet representerar utan snarare vilken provtagningsmetod som används och hur variabiliteten i marken ser ut. Ett extremfall är om helt homogena

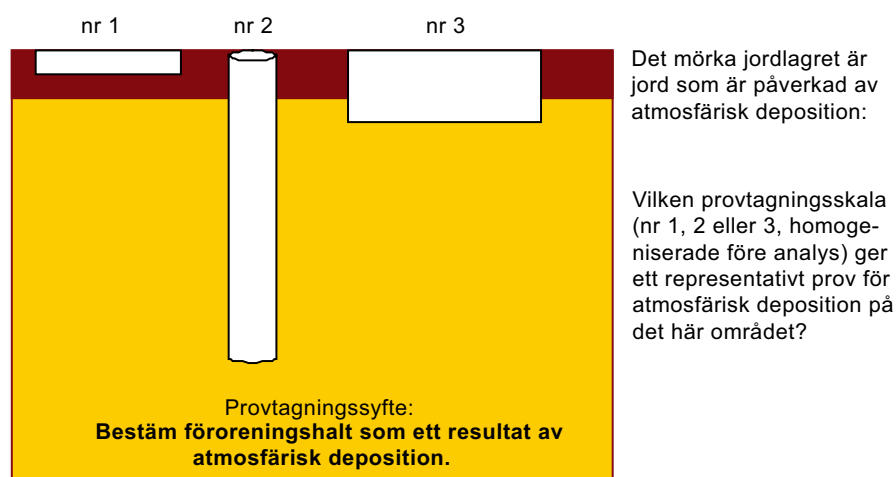
förhållanden föreligger — då behövs bara ett prov och *supporten* är lika stor som undersökningsområdet. Det motsatta extremfallet är då jordprovet bara representerar själva provvolymen. Samlingsprov som tas har en *support* någonstans mellan dessa ytterligheter.

Jordprover från förorenade områden kan representera olika skalor, exempelvis följande (listade med ökande support):

- Fältmätningar med exempelvis XRF-instrument, representerande en volym på några cm³.
- Ytjord provtagen med spade, representerande några dm³.
- Jordprov taget med skruvborr representerar i storleksordningen 5 dm³.
- Samlingsprov från provgröp grävd med grävmaskin representerande någon eller några m³.
- Samlingsprov från flera provpunkter, exempelvis en efterbehandlingsvolym, representerande i storleksordningen 10-500 m³.

Notera att inte bara volymen utan även formen på volymerna ovan skiljer sig åt. Tre olika former av *support* visas grafiskt i Figur 7-3 nedan.

Provtagningskala inbegriper provets rumsliga utbredning



Figur 7-3. Exempel på olika provtagningsskalor (*support*) vid provtagning av jord. Den mest lämpliga skalan beror på provtagningens syfte. Anpassad efter NARPM (2005).

Vid undersökning av förorenad jord bör man alltid ange vilken *support* data representerar. Detta gäller oavsett syftet med undersökningen och är en nödvändighet för att data i senare skeden ska kunna värderas eller användas korrekt.

När datamängden består av data med olika *support* måste detta hanteras. Ju större skillnaden är mellan olika *support*, desto större blir felet om man blandar de olika typerna av data. Följande gäller då man har två datamängder med mycket olika *support*:

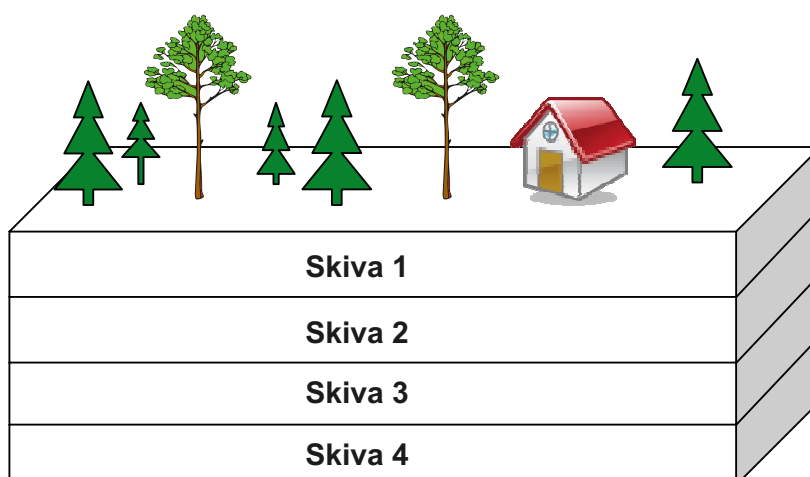
- Undvik att blanda data med olika *support* om det är variabiliteten som är intressant.

- Om det är medelhalten som ska bestämmas kan data med olika *support* under vissa förutsättningar hanteras som en datamängd. Ett sådant fall är när ett samlingsprov tagits av flera provpunkter som slumpats ut över ett större område och detta mätvärde hanteras tillsammans med mätdata från enskilda provpunkter. I detta fall bör samlingsprovet ges en större vikt än övriga eftersom det representerar flera provpunkter. Om man har två datamängder med mycket olika *support* kan det vara bättre att beräkna statistiken för vardera datamängd för sig. Därefter kan man jämföra resultaten och diskutera vilken *support* som är mest relevant för det aktuella problemet. Om de två datamängderna kommer från samma egenskapsområde kan förvänta sig följande statistiska resultat:
 - Medelvärdet för de två stickproven bör bli någorlunda lika men man kan ändå förvänta sig en viss skillnad.
 - Varians och standardavvikelse kommer att bli lägre för data med större *support*.
- Om antalet data i de två datamängderna är någorlunda stort är det möjligt att transformera data från en *support* till en annan. I praktiken innebär detta att man kan räkna om föroreningshalter som uppmätts med t.ex. XRF i liten skala till halter som motsvarar en större jordvolym, t.ex. en provgrop. Omräkningen leder till att variansen i data minskar. Därefter kan de två datamängderna hanteras som en, eftersom de representerar samma volym. Detaljer om hur sådana transformeringar går till redovisas av bl.a. Myers (1997).

7.8 Data från olika provtagningsdjup

Data som representerar olika djup i jordprofilen förekommer i nästan alla undersökningar av förorenade mark. Hur sådana data ska hanteras beror på hur man betraktar sitt problem. Nedan beskrivs tre olika angreppssätt, två som bygger på klassisk statistik utan att den rumsliga korrelationen beaktas, samt ett som bygger på geostatistik.

Provtagning av förorenad jord kan betraktas som ett tredimensionellt (3D) provtagningsproblem. Detta är det mest korrekta synsättet. I praktiken är det dock ofta enklare att hantera djupet (z-koordinaten) något annorlunda än x- och y-koordinaterna, av praktiska skäl. En förenkling som ofta kan göras är att dela upp området i horisontella ”skivor” (strata) av jord, där varje skiva har en viss tjocklek eller följer en viss jordart (man kan likna detta vid ett antal pannkakor som lagts på varandra), se Figur 7-4. Därefter hanteras varje skiva tvådimensionellt. Detta är ett mycket vanligt angreppssätt vid undersökning av förorenad jord. Ett bra argument för denna strategi är att man många gånger (dock inte alltid) kan förvänta sig att föroreningshalterna avtar mot djupet, vilket gör att man vill kunna särskilja olika djup i den statistiska utvärderingen.



Figur 7-4 Indelning av den 3-dimensionella jordvolymen i 2-dimensionella skivor av en viss tjocklek. Med detta angreppssätt görs en separat statistisk analys för varje skiva.

Om det tvådimensionella angreppssättet ("pannkaksstrategin") används måste varje datavärde hänföras till en av skivorna. Problem uppkommer om ett mätvärde representerar en volym som överlappar mellan två skivor. Detta är en indikation på att provtagning och utvärdering inte planerats tillräckligt då undersökningen inleddes. Hur sådana överlappningsproblem ska lösas måste bedömas från fall till fall. När samtliga mätdata kopplats till en viss skiva kan statistiken för denna skiva beräknas, exempelvis medelhalten. På detta sätt kan föroreningsituationen bestämmas för vardera skivan. Notera att osäkerheten i bedömningarna kan variera kraftigt mellan olika skivor beroende på antalet data i respektive skiva.

En alternativ strategi är att slå samman alla "pannkakor" till en enda skiva. Hela jordvolymen karakteriseras då som en enhet med hjälp av hela datamängden (stickprovet). Notera dock följande: för att statistiken för hela jordvolymen ska bli representativ krävs att provpunkterna placerats ut slumpmässigt i både x-, y- och z-riktningarna. Med andra ord ska provtagningen vara lika omfattande längst ner i jordvolymen som vid markytan för att denna strategi ska ge rättvisande resultat. Strategin kan också vara tveksam om föroreningshalterna avtar kraftigt mot djupet.

En nackdel med ovanstående två angreppssätt är att man bortser från den korrelation som kan finnas mellan olika provpunkter. För att ta hänsyn till detta krävs tredimensionella geostatistiska metoder, t.ex. kriginginterpolation i 3D. Om man i sin analys kommit fram till att det är lämpligt att utföra interpolation så finns det idag många programvaror som klarar detta för tre dimensioner. För att få ut mesta möjliga information från sin datamängd är detta angreppssätt att föredra.

7.9 Projekt med få data

Förorenad mark-projekt kan vara av mycket olika omfattning beroende på problemställning och i vilket skede undersökningarna görs. Det är inte ovanligt att datamängderna är mycket små. Frågan är hur detta ska hanteras vid en statistisk utvärdering.

För att en statistisk utvärdering ska ge ett bra resultat krävs först och främst att data kommer från slumpmässigt valda provpunkter inom egenskapsområdet. Givetvis blir noggrannheten i resultaten större om antalet data (stickprovet) är stort. Det finns dock inget som hindrar att en statistisk utvärdering görs även för små stickprov. Det som krävs är att antalet data är minst några stycken per egenskapsområde. Om data är få kommer t.ex. medelvärdet att bli en osäker skattning av den verkliga medelhalten i området men statistiken är ändå relevant. Med andra ord är små projekt sällan ett bra argument för att låta bli att göra en statistisk utvärdering av data. Det finns många exempel på att statistisk utvärdering använts framgångsrikt även för projekt med små datamängder. Det som snarare kan skapa problem är om stickprovet i hög grad är riktat mot förväntat höga föroreningshalter så att datamängden inte är representativ för området, se avsnitt 7.2 och 7.10.

7.10 Data med osäker representativitet

Många verkliga datamängder från förorenade områden har en osäker representativitet. Orsakerna till detta kan vara flera, varav några vanliga är följande:

- En felaktig indelning i egenskapsområden har gjorts. Alternativt har ingen indelning i egenskapsområden gjorts, trots att flera olika föroreningspopulationer finns i området. Båda dessa fall innebär att samtliga data kan vara helt korrekt i sig själva men tillsammans kan de ge en felaktig bild av området.
- Riktad provtagning har utförts, vilket leder till icke-representativ statistik för ett egenskapsområde.
- Flera jordprover har tagits från olika djup i en provgrop men bara ett prov har analyserats, baserat på förväntad hög föroreningshalt. Även detta är en typ av riktad provtagning som leder till icke-representativ statistik.
- Mätdata representerar olika jordvolymmer (olika *support*).
- Äldre och nyare datamängder har blandats, vilket kan ge dålig representativitet om olika analysmetoder använts (de kan ha förändrats över tiden eller olika laboratorier kan ha använt olika metoder).
- Samlingsprov har tagits som överlappar mer än ett egenskapsområde eller flera djupintervall. Detta kan innebära att mätdata representerar flera olika populationer.

- Endast samlingsprov som representerar stora jordvolymen har tagits, vilket gör att eventuella *hotspots* inte kan upptäckas.
- Provtagningen kan ha introducerat stora osäkerheter eller fel, bl.a. systematiska fel. Se Back (2003) för en detaljerad beskrivning av sådana osäkerheter och fel.

Flera av ovanstående aspekter har diskuterats tidigare i detta kapitel. I vissa fall kan effekterna av den osäkra representativiteten minskas med olika metoder, se t.ex. avsnitt 7.2 och 7.7. I andra fall måste man leva med osäkerheten men tydligt påpeka den när resultaten redovisas. En framgångsrik strategi kan vara att försöka väga samman de olika bristerna i sin undersökning och göra en bedömning om de statistiska parametrarna som medelhalt och varians överskattas eller underskattas. Det kan då eventuellt visa sig att osäkerheterna är acceptabla, om de inte påverkar slutsatserna av utredningen. I enstaka fall kan det dock, under olyckliga omständigheter, vara befogat att inte göra någon statistisk utvärdering utan istället kassera datamängden, särskilt om man har anledning att misstänka stora systematiska fel.

7.11 Kvalitetskontroll och rimlighetsbedömningar

Beräknad statistik för en datamängd måste alltid kvalitetskontrolleras. Detta kan göras med hjälp av överslagsberäkningar eller med hjälp av andra statistiska verktyg. Rimligheten på resultaten måste alltid kontrolleras. Exempel på frågor man bör ställa sig när man granskar resultaten är:

- Är medelvärdet rimligt med hänsyn till de data jag har?
- Är standardavvikelsen rimlig med hänsyn till medelvärdet och spridningen i data?
- Är beräknade volymer och mängder rimliga? Är storleksordningen den rätta?
- Är interpolerade kartor över föroreningsbilden rimliga? Finns det någon fysisk process som kan förklara den rumsliga korrelationen och den rumsliga variationen? Är interpolerade halter rimliga i de områden där inga prover tagits?

Sådana rimlighetsbedömningar är enkla sätt att upptäcka grova fel i resultaten.

8 Fallstudie med beräkningsexempel

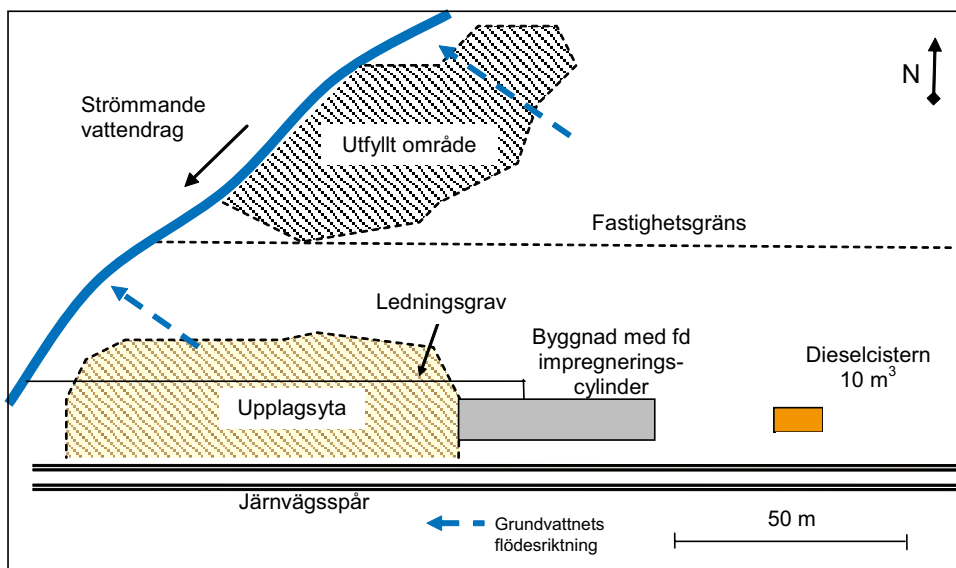
Fallstudien som används här är samma område som använts i Norrman et al. (2009) varför texten för områdesbeskrivningen i stort sett är densamma. Här fokuseras dock endast på upplagsytan som identifierats som ett egenskapsområde. För att följa beräkningarna fallstudien kan man behöva ta hjälp av huvudtexten i rapporten.

8.1 Områdesbeskrivning för fallstudien

Inom undersökningsområdet har tryckimpregnering bedrivits från 1940-talet och fram till mitten 1990-talet. Impregneringsmedlet har varit Bolidensalt (koppar-, krom-, och arseniksalt). Tryckimpregnering har bedrivits med fullcellsmetoden och enligt uppgift alltid på samma plats. Anläggningen är numera nedmonterad och bortforslad från området, kvar finns endast en byggnad. Det impregnerade virket lastades på järnvägsvagnar för transport och i vissa perioder mellanlagrades virket på en upplagsyta. En invallad drivmedelscistern (diesel/brännolja) ovan mark för verksamhetens truckar och fordon finns fortfarande kvar på fastigheten.

Jordarterna inom fastigheten utgörs av genomsläppligt isälvsmaterial, framförallt sand och grus. Området är relativt flackt och sluttar svagt ner mot en å, se plankarta i Figur 8-1. Grundvattenytan återfinns två till tre meter under markytan.

Undersökningsområdet har en grusad markyta. Ett större utfyllt område finns norr om fastigheten, i anslutning till en större å som rinner i sydvästlig riktning. Fyllnadsmassornas ursprung och sammansättning är okända.



Figur 8-1. Plankarta över området för fallstudien.

Verksamhetshistoriken indikerar att det finns fem olika egenskapsområden som kan urskiljas med avseende på verksamhet och typ av föroreningar:

- 1) Dieseltornen.
- 2) Impregneringsanläggningen.
- 3) Upplagsytan.
- 4) Eventuellt nedgrävt impregneringssalt.
- 5) Utfyllnadsområdet.

I Tabell 8-1 sammanfattas föroreningsituationen för upplagsytan avseende typ av förorening samt en hypotes kring föroreningsituationen.

Tabell 8-1 Sammanfattning av typ av förorening och hypotes om föroreningsituationen.

Delområde	Typ av förorening	Hypotes kring föroreningsituationen
Upplagsytan, ca 2100 m ² stor	Koppar, krom och arsenik. Känd men heterogen (och diffus) källa över hela ytan.	Spill/dropp från impregnerat virke bör kunna påträffas fläckvis i de översta marklagren. Om betydande mängder spill/dropp har förekommit inom upplagsytan kan förorening även påträffas på större djup. Eftersom utlakning av förorening går långsamt förväntas marginell påverkan på grundvattnet.

8.2 Provtagning och analysresultat

Nedan följer en förkortad beskrivning av den föreslagna arbetsgången som finns redovisad i Norrman et al. (2009) för att ta fram en provtagningsplan.

8.2.1 Provtagningens syfte

Det övergripande syftet med provtagningen är att bedöma om upplagsytan är förorenad över riktvärdena samt att bedöma hur stor mängd förorening som finns. Representativa halter skall tas fram för området och dessa ska användas i en riskbedömning samt för mängdberäkningar.

I ett nästa steg är man intresserad av att identifiera delområden som är mer förorenade än andra genom någon form av interpolation. Ett delsyfte blir således också att undersöka om det finns någon rumslig korrelation mellan provtagningspunkterna så att interpolation kan utföras.

8.2.2 Önskad säkerhet

En metod baserad på hypotestestning används för att beräkna storleken på stickprovet utifrån önskad säkerhet, för vidare beskrivning, se Norrman et al. (2009).

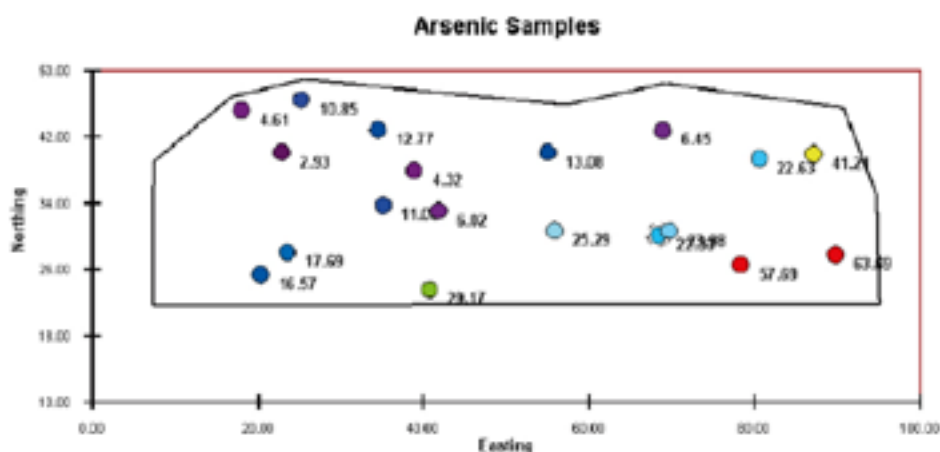
De parametrar som anger önskad säkerhet och måste väljas för att beräkna stickprovet är α (väljs till 0,05), β (väljs till 0,2), samt effektstorleken (Δ). Effektstorleken väljs som 50 % av riktvärdet (AL), där riktvärdet för arsenik är 15 mg/kg TS; dvs. $\Delta = AL \times 0,5 = 7,5$ mg/kg. Detta ger, tillsammans med en antagen variationskoefficient på 1,5, att stickprovets storlek bör vara minst 19 prover.

8.2.3 Provtagning och analysresultat

Här väljs en systematisk slumpmässig provtagning. Placering och provtagningsmönster osv. utförs enligt Norrman et al. (2009). Alla prover analyseras avseende arsenik. Tabell 8-2 redovisar mätvärdena och Figur 8-2 deras placering inom upplagsytan.

Tabell 8-2 Analysresultat från upplagsytan.

Provnr.	As (mg/kg)	Provnr.	As (mg/kg)
1	4,61	10	63,69
2	12,77	11	10,85
3	13,08	12	4,32
4	6,45	13	41,21
5	22,63	14	17,69
6	16,57	15	11,08
7	29,17	16	57,69
8	25,29	17	2,93
9	23,98	18	6,02
		19	22,33



Figur 8-2. Analysresultat för arsenik i jord från upplagsytan. Halterna i mg/kg visas till höger om varje provpunkt.

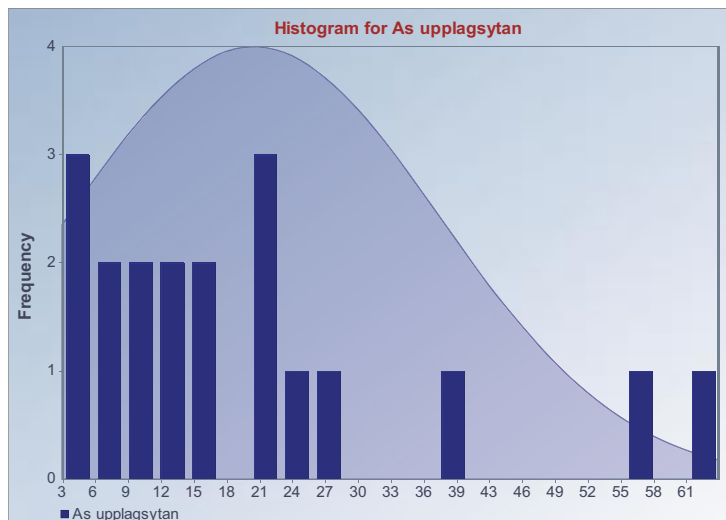
8.3 Steg 1: Bedömning av föroreningsgrad

8.3.1 Beskrivande statistik

De 19 analyserna av arsenik lades in i ProUCL för att ta fram beskrivande statistik över stickprovet, se Tabell 8-3. Data plottades i ett histogram för att undersöka om data är skevt fördelade, se Figur 8-3.

Tabell 8-3. Sammanfattande statistik för stickprovet.

Variabel	Antal	Min	Max	Medel	Median	Varians	Std	Skevhets	Kurtosis	CV
As	19	2,93	63,69	20,65	16,57	297	17,23	12,93	1,421	1,572
Antal enskilda prov	Minsta värdet i stick- provet	Största värdet i stick- provet	Arit- metiskt medel- värde	Medianen (dvs. 50-per- centilen)	Varians	Standard- avvikelse	Skevhets <0=vänster >0=höger 0=symmetri	Toppighet >3= spetsig <3= flack	Variations- koefficient = standard- avvikelse/ medelvärde	



Figur 8-3 Histogramplot i ProUCL för arsenik-data från upplagsytan.

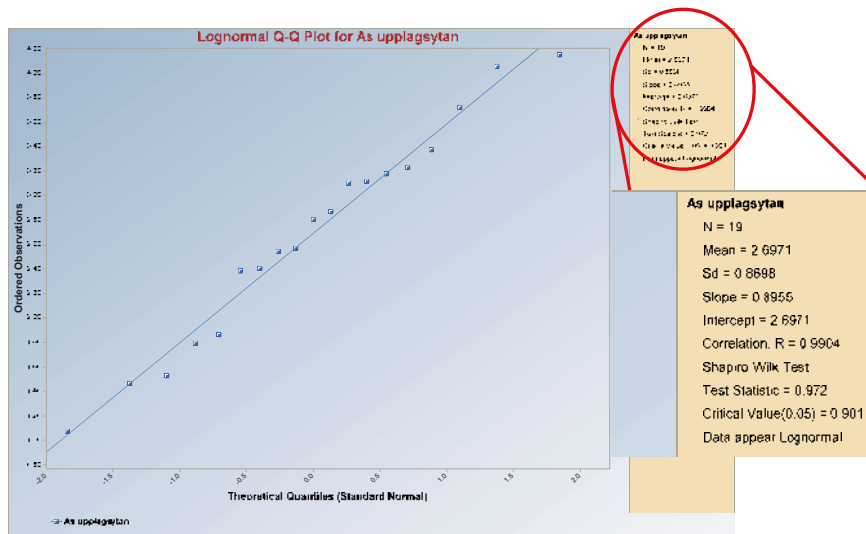
Genom att ta fram beskrivande statistik får man en första bild av data, hur spridningen ser ut och datas fördelning. Här kan man misstänka att data inte följer en normalfördelning utan verkar ha en skev fördelning, både utifrån histogram-plotten och från måttet på skevhet.

Från den beskrivande statistiken i Tabell 8-3 ses att både stickprovets medelvärde och median ligger över riktvärdet. Därför kan man redan här konstatera att området bör bedömas som förorenat över riktvärdet. Trots detta genomförs beräkningarna av UCLM95 och hypotestestet för att se över osäkerheterna.

8.3.2 Goodness-of-fit test

Datamängden för arsenik i fallstudien plottas i en lognormal Q-Q-plot, med hjälp av ProUCL, för att undersöka om data följer en lognormalfördelning, (Figur 8-4). Testen om data är lognormalfördelad görs med hjälp av Shapiro-Wilks test i ProUCL. Nollhypotesen formuleras som att data är lognormalfördelade och testet utförs på signifikansnivå 95%.

I grafen ges teststatistikan och det kritiska värdet för testet (*Critical Value*) på signifikansnivån 95%. På den nivå accepteras en felrisk på 5%, dvs. ett p-värde på max 0,05. Här är teststatistikan 0,972 och det kritiska värdet 0,901. Eftersom värdet från testet överstiger det kritiska värdet kan vi inte förkasta nollhypotesen att data är lognormalfördelade, och därmed förutsätter vi att det är rimligt att vår målpopulation är lognormalfördelad.



Figur 8-4. Lognormal Q-Q-plot i ProUCL. Resultatet av Q-Q-ploten och Shapiro-Wilks test i ProUCL. Här presenteras stickprovets logaritmerade medelvärde och standardavvikelse, samt resultat från testet.

8.3.3 Beräkning av UCLM95

Först jämförs riktvärdet för arsenik (15 mg/kg) med övre konfidensgränsen för den skattade medelhalten i området, dvs. UCLM95. Beräkningarna görs i ProUCL.

Data har konstaterats vara lognormalfördelade och vi antar att målpopulationen även är lognormalfördelad. Med hjälp av Land's metod beräknas UCLM95 för datamängden för arsenik till 35,7 mg/kg (jämför med det aritmetiska medelvärdet som är 20,7). UCLM95 överskrider klart riktvärdet på 15 mg/kg, vilket bör tolkas som att risken inte är acceptabel i relation till riktvärdet. I eventuella beräkningar för att kvantifiera riskerna bör det övervägas att använda UCLM95-värdet, om denna säkerhet önskas.

8.3.4 Hypotestest: stickprov mot riktvärde

Stickprovet av arsenik användes också för att utföra en hypotestest. Tidigare har konstaterats att data och målpopulationen kan antas vara lognormalfördelade, vilket innebär att Chen-testet rekommenderas. Chen-testet saknas i ProUCL, men finns beskrivet i USEPA (2006) och beräkningarna kan ganska enkelt utföras i Excel.

Om nollhypotesen H_0 formuleras enligt:

H_0 : medelhalten av As på upplagsytan är större eller lika med 15 mg/kg,

och alternativhypotesen H_A enligt:

H_A : medelhalten av As på upplagsytan är mindre än 15 mg/kg,

så erhålls ett p-värde för testet på 0,963. Detta innebär att risken att ha fel om nollhypotesen förkastas är så stor som 0,963 eller 96%. Utifrån testet kan man därför med stor säkerhet säga att medelhalten i området överskrider riktvärdet.

8.3.5 Bedömning av andel som överskrider referenshalt för akuttoxicitet

Då föroreningar med akuttoxiska egenskaper förekommer är det lämpligt att bedöma hur stor andel av ett område eller en jordvolym som har akuttoxiska halter. Denna andel kan också ses som sannolikheten att man slumpvis råkar välja en punkt med akuttoxisk halt, exempelvis då ett barn äter jord. För denna bedömning kan alla de tre metoder som beskrivs i kapitel 4 användas. Alla dessa metoder har dock en viss osäkerhet eftersom bedömningen bör göras i datamängdens periferi. Osäkerheten är särskilt stor om det inte finns några mätvärden som är högre än referenshalten för akuttoxicitet.

I detta exempel används metoden som bygger på antagandet om en lognormalfördelad population. Parametrarna för fördelningsmodellen beräknas från stickprovet och blir:

Medelvärde på logskalan: 2,70
Standardavvikelse på logskalan: 0,87

Med hjälp av dessa parametrar kan sannolikheten att överskrida referenshalten för akuttoxicitet beräknas. Som referenshalter för arsenik används 100 mg/kg (ingen förväntad akut effekt) respektive 2000 mg/kg (dödlig effekt), se exempel 3-5. Beräkningarna görs t.ex. i Excel genom att använda sig av uttrycket:

"=1-LOGNORMFÖRD(referenshalt för akuttoxicitet; medelvärde på logskalan; standardavvikelse på logskalan)".

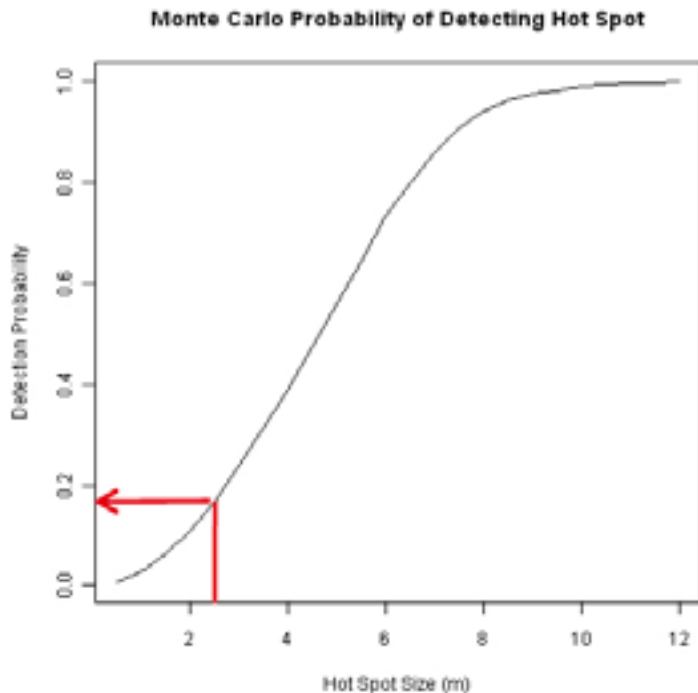
Andelen förorenade massor över 100 mg/kg beräknas på detta sätt till 1% av området. Områdets storlek är ca 2100 m², vilket innebär att en sammanlagd area på ca 20 m² förväntas ha halter över 100 mg/kg. Detta kan också uttryckas som att sannolikheten är 1% att ett barn som äter jord på området väljer en punkt med halter över 100 mg/kg. Huruvida detta är acceptabelt eller ej bör diskuteras med inblandade aktörer eftersom inga kriterier för detta finns.

Vid bedömning av andelen förorenade massor över 2000 mg/kg beräknas denna till $8,6 \times 10^{-11}$ % av området, dvs. en försumbar andel.

8.3.6 Sannolikheten för okänd hotspot

Man vill även bestämma sannolikheten att hitta en *hotspot*, givet det mönster som provpunkterna bildar inom området. Figur 8-5 visar ett diagram av hur sannolikheten att detektera en hotspot beror av hotspotens storlek för det aktuella provtagningsmönstret vid upplagsytan (se Figur 8-2). Figuren har genererats genom Monte Carlo-simulering där en cirkulär hotspots radie varit 0,5 m men stegvis ökats med en halv meter upp till maximalt 12 m. För varje steg har 10 000 realiseringar simulerats fram. Hotspoten placeras ut med en slumpmässigt vald mittpunkt inom det definierade området (upplagsytan) 10 000 gånger och om en eller flera av provpunkterna hamnar inom den simulerade hotspoten så antas att den detekteras utan detektionsfel. Sannolikheten att detektera hotspoten är antalet realiseringar där hotspoten detekteras delat med det totala antalet realiseringar. Beräkningarna utfördes i SADA v.5 (beta-

version). Sannolikheten att upptäcka en antagen hotspot på 20 m² (vilket för en cirkulär hotspot motsvarar en radie på ca. 2,5 m) är ca 0,15, se Figur 8-5. Om detta är acceptabelt kanske är tveksamt men en slutsats blir att om kravet finns att hitta en liten hotspot med stor sannolikhet, så krävs väldigt många prover, förmodligen orimligt många.

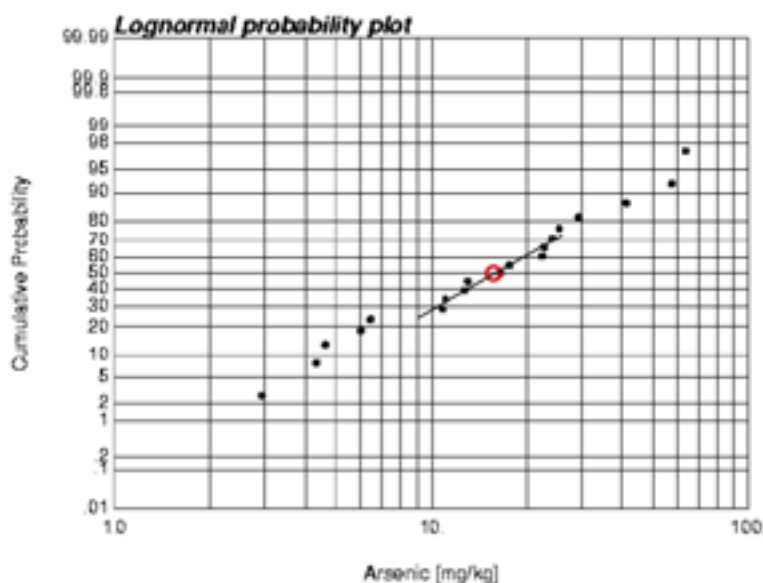


Figur 8-5 Diagram över hur sannolikheten att detektera en *hotspot* beror av hotspotens storlek för det aktuella provtagningsmönstret vid upplagsytan. För en liten hotspot är sannolikheten låg att den ska detekteras, men med ökande storlek så ökar även sannolikheten för detektion. Den röda pilen markerar ungefärligen sannolikheten för att en hotspot på 20 m² ska detekteras.

8.4 Steg 2. Bedömning av andelen förorenade massor

8.4.1 Bedömning av andel med normalfördelningsplot

Data från fallstudien plottas i en normalfördelningsplot. Eftersom man kan förvänta sig att data är relativt skevt fördelade med enstaka höga värden väljer man att logaritmera x-axeln, dvs. man får en lognormalfördelningsplot (Figur 8-6). Om data följer en rät linje i grafen tyder detta på att data är lognormalfördelade. Här läses dock andelen över riktvärdet av direkt i grafen. En hjälplinje har lagts in i grafen för att underlätta avläsningen mellan de närliggande datapunkterna. Y-axeln visar att andelen under riktvärdet är ca 50 %, vilket innebär att andelen över riktvärdet också är ca 50 %.



Figur 8-6 . Lognormalfördelningsplot för 19 arsenikdata. Den röda cirkeln markerar riktvärdet 15 mg/kg. Grafen visar att ca 50% av data överskrider riktvärdet.

8.4.2 Bedömning av andel med statistisk fördelning

Bedömningen av andelen massor med halt av arsenik som överskrider 15 mg/kg beräknas på samma sätt som i avsnitt 8.3.5, med skillnaden att referenshalten för akuttoxiciteten här ersätts med riktvärdet.

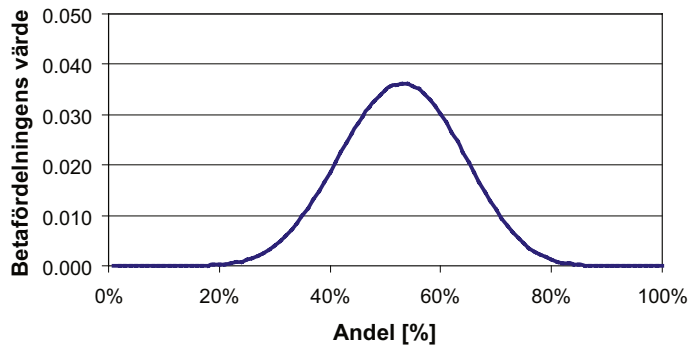
Vid bedömning av andelen förorenade massor över 15 mg/kg beräknas en andel om 49% av området fram. Områdets totala storlek är 2100 m², vilket skulle innebära att ca 1050 m² skulle uppvisa halter över 15 mg/kg. Notera dock att detta gäller för den *support* som data representerar, se avsnitt 7.7.

8.4.3 Bedömning av andel med betafördelning

Parametrarna för betafördelningen blir $\alpha = 11$ (tio observationer över riktvärdet) och $\beta = 10$ (nio observationer under riktvärdet). I Excel beräknas det mest troliga värdet på andelen (x) av området där halten överskrider riktvärdet med hjälp av uttrycket för "mode" (se Bilaga C):

$$\hat{x} = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{11 - 1}{11 + 10 - 2} = 0,53$$

Detta innebär att 53% av området bedöms vara förorenat över riktvärdet. De två gränserna för det 90%-iga kredibilitetsintervallet beräknas till 35% respektive 70%. Den aktuella betafördelningens utseende visas i Figur 8-7.



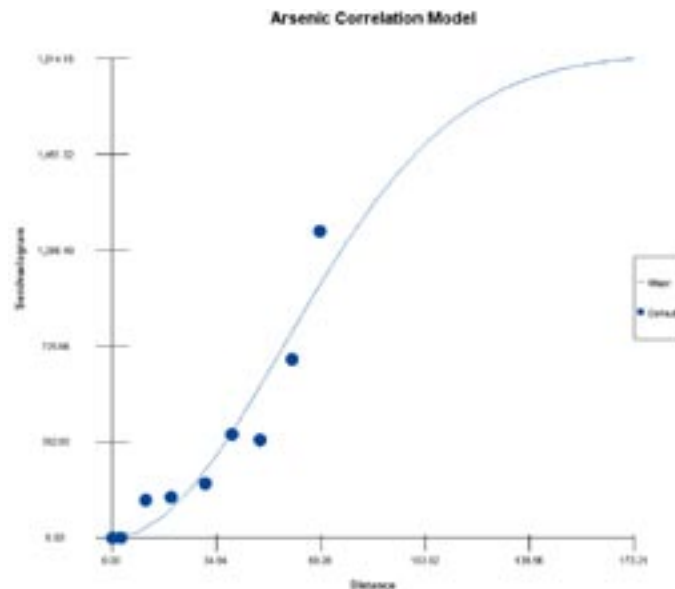
Figur 8-7. Betafördelningen för stickprovet från upplagsytan.

På basis av dessa skattningar görs bedömningen att det kan vara fördelaktigt att avgränsa de förorenade massorna före saneringen.

8.5 Steg 3. Bedömning av rumslig korrelation

Med hänsyn till det sätt som föroreningarna har hamnat på området området är det troligt att det finns en rumslig korrelation. Det behandlade virket har körts in på området från den östra sidan och framförallt lagrats och lastats längs med järnvägsspåren söder om upplagsytan. Ett experimentellt variogram konstruerades för att undersöka den rumsliga korrelationen, se Figur 8-8.

Med hjälp av det experimentella variogrammet och den konceptuella modellen drogs slutsatsen att rumslig korrelation föreligger inom området och att det är värt att gå vidare med en geostatistisk interpolation för att ta fram en rumslig modell som kan användas för att exempelvis avgränsa föroreningar med halter över ett haltkriterie (åtgärds mål).



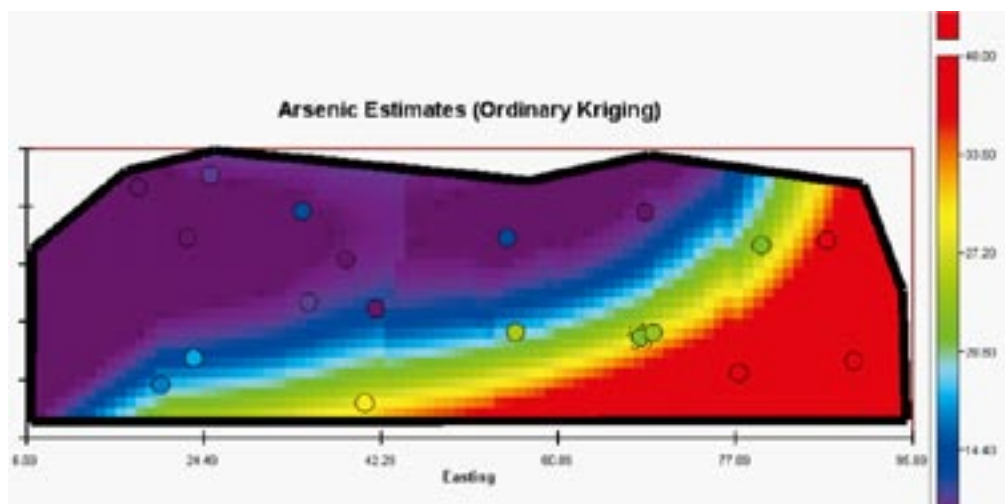
Figur 8-8. Experimentellt variogram (prickar) samt passad variogrammodell (heldragen kurva) för stickprovet från upplagsytan. Variogrammet är konstruerat med logaritmerad data eftersom data har befunnits följa en lognormalfördelning.

8.6 Steg 4. Interpolation

Variogramanalysen utfördes i SADA. Det vanliga förfarandet vid en variogramanalys är att pröva sig fram till en indelning i avståndsklasser som är lämplig genom att titta på det experimentella variogrammet. Därefter föreslår man en modelltyp och låter programvarans autorutin göra en passning till kurvan. Parametrarna kan även handjusteras. Modellen som bäst passade data i föreliggande exempel ges av följande parametrar i en Gaussisk modell:

Korrelationslängd (<i>range</i>):	80 m
Global variation (<i>sill</i>):	1830
Småskalig variation (<i>nugget effect</i>):	0

Interpolationen gjordes med vanlig kriging där området delades in i celler om 1×1 m. Figur 8-9 visar resultatet från kriginganalysen med skattningar av den troligaste halten i varje cell.

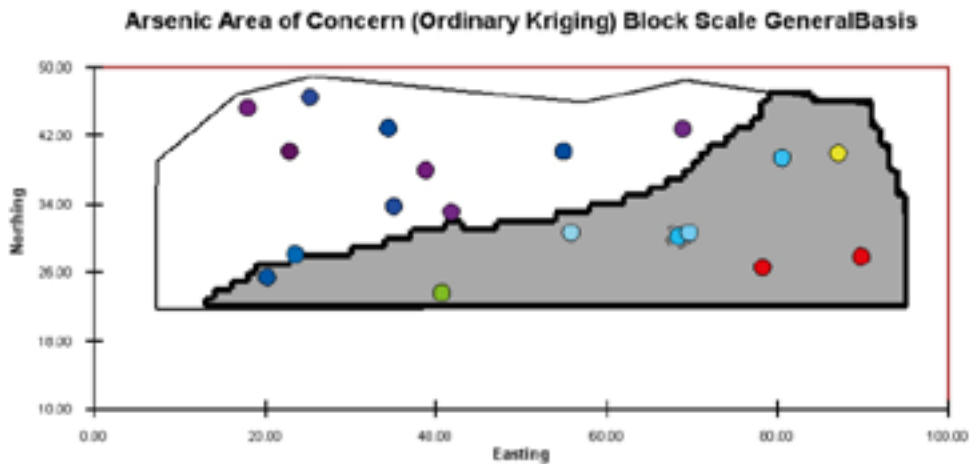


Figur 8-9. Skattning av den troligaste halten i varje punkt över hela området mha. interpolation med kriging (ordinary kriging).

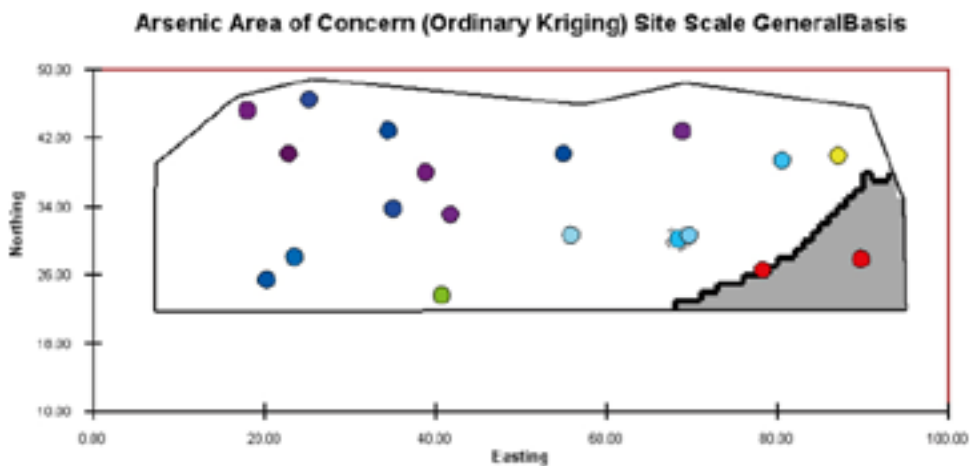
Baserat på kriginganalysen användes två olika angreppssätt för att generera kartor över de celler som bör åtgärdas baserat på ett åtgärds mål satt till 15 mg/kg, dvs samma som riktvärdet:

1. Karta där åtgärds målet appliceras på cellskala (Figur 8-10)
2. Karta där åtgärds målet appliceras på hela upplagsytan (Figur 8-11).

Det är intressant att notera den stora skillnaden i saneringsområdenas storlek för de två fallen, se Figur 8-10 och Figur 8-11 nedan, trots att åtgärds målet utgörs av samma halt, 15 mg/kg.



Figur 8-10. Tvådimensionell avgränsning av ytor som kräver efterbehandling baserat på att medelhalten i varje enskild cell ska uppfylla åtgärds målet 15 mg/kg. Storleken på området som behöver saneras är 1039 m².



Figur 8-11. Tvådimensionell avgränsning av ytor som kräver efterbehandling baserat på att medelhalten i hela området ska uppfylla åtgärds målet 15 mg/kg. Storleken på området som behöver saneras är 204 m².

Som komplement till de två angreppssätten ovan så testades ytterligare ett: Angreppssätt 2 i kombination med ett åtgärds mål på cellskala där referenshalten för akuttoxicitet inte får överskridas i cellerna, dvs. ett åtgärds mål på 100 mg/kg tillämpades för varje enskild cell. I det här fallet kunde dock inte några celler med halter över 100 mg/kg identifieras genom interpolationen, varför en sådan utformning av efterbehandlingsinsatserna skulle se likadan ut som i Figur 8-11.

9 Diskussion

I den här rapporten har fokus framförallt legat på att kvantifiera typ I-fel, dvs. risken att felaktigt bedöma föroreningsgraden i ett område som under en referenshalt i ett område som egentligen kräver efterbehandling. Vilken storlek på typ I-fel som är acceptabel har att göra med vem som fattar beslut, vilka kostnader och nyttor som är förknippade med en sanering och hur dessa värderas, samt vilka miljö- och hälsorisker som kan föreligga på området. Valet av representativ halt för det aktuella området är också en viktig del i en utredning. I den här rapporten har fokus framförallt varit på den 95 %-iga övre konfidensgränsen för den skattade medelhalten i området (UCLM95), men även andra mått på den representativa halten kan vara av intresse i olika situationer, se t ex Arnér (2009).

Betafördelningen är ett verktyg där subjektiv kunskap formellt kan läggas in i en statistisk analys. Detta verktyg klarar dock inte att hantera den spatiala komponenten hos data, dvs. det är ingen rumslig analys av data som utförs med hjälp av betafördelningen. Ansatser har dock gjorts för att kunna använda subjektiva inslag i rumsliga modeller (t.ex. i SADA), men arbete återstår för att göra sådana modeller användbara. Anledningen till att modeller som kan ta hänsyn till subjektiva data är ett intressant utvecklingsområde är att de personer som arbetar med utredningar av förorenade områden ofta sitter inne med en stor portion erfarenhet och expertkunskap som inte tillvaratas i klassiska (frekventistiska) statistiska analyser. Kombinationen av expertkunskap och klassisk statistik skulle kunna resultera i mycket effektivare verktyg än de som idag används i branschen.

Den skala som ett åtgärds mål ska tillämpas på är en viktig aspekt. Om åtgärds mål som tagits fram för att gälla för ett helt område tillämpas på SEV-skalan kan detta leda till översanering med resthalter långt under det åtgärds mål som skulle uppnås för platsen. Detta medför givetvis kostnader som skulle kunna undvikas.

För att statistiska metoder och verktyg skall börja användas i större utsträckning än idag krävs en allmän höjning av kunskapsnivån i Sverige inom branschen. En ökad kunskap kan bidra till att det blir lättare att kommunicera statistiska analyser mellan t.ex. konsulter, problemägare och tillsynsmyndigheter. Att höja kunskapsnivån om statistiska metoder och verktyg bör kunna bidra till att kvalitén generellt sett höjs på utredningar och beslutsunderlag för förorenade områden.

10 Referenser

- Anselin L, 1995. Local indicators of spatial association - LISA. *Geographical Analysis*, 27:93-115.
- Arnér M, 2009. Information om Naturvårdsverkets vägledningsmaterial för förorenade områden 4 mars 2009 (PowerPoint-bilder om riskbedömning).
- Back P.E, 2003. On Uncertainty and Data Worth in Decision Analysis for Contaminated Land. Department of Geology, Chalmers University of Technology, Göteborg. (Licentiatuppsats).
- Blom G, 1989. Sannolikhets teori och statistik teori med tillämpningar. Studentlitteratur, Lund.
- Brus D.J, de Gruijter J.J, 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). *Geoderma*, 80:1-44.
- Cliff A.D, Ord J.K, 1981. *Spatial Processes: Models and Applications*. London, Pion.
- Deutsch C.V, & Journel A.G, 1998. *GSLIB. Geostatistical software library and user's guide*. Second edition. Oxford University Press, New York.
- Geary R.C, 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5:115-145.
- Gilbert R, 1987. *Statistical methods for environmental pollution monitoring*. John Wiley & Sons, Inc. New York.
- Golden H.E, Boyer E.W, Brown M.G, Purucker S.T, Germain R.H, 2009. Spatial variability of nitrate concentrations under varying seasonal conditions in tributaries to Cayuga Lake Watershed, New York, USA. *Journal of the American Water Resources Association*, (in revision).
- Grandin U, 2003. Dataanalys och hypotesprövning för statistikanvändare. PM. Naturvårdsverket. http://www.naturvardsverket.se/upload/02_tillstandet_i_miljon/Miljoovervakning/handledning/dataanalys_och_hypotesprovn.pdf (2008-05-26).
- Griffith D.A, 2005. Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American geographers*, 95(4):740-760.
- Helsel D.R, 1990. Less than obvious. Statistical treatment of data below the detection limit. *Environ. Sci. Technol.*, 24(12):1766-1774.
- Lam N. S.-N, 1983. Spatial interpolation methods: a review. *The American Cartographer* 10: 129-149.

- Legendre P. & L. Legendre, 1998. Numerical Ecology, 2nd edition. Elsevier, Amsterdam.
- Mantel N, 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2):209-220.
- Moran P.A, 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37:17-37.
- Myers J.C, 1997. Geostatistical error management. Quantifying uncertainty for environmental sampling and mapping. Van Nostrand Reinhold, New York.
- NARPM, 2005. EPA National Association of Remedial Project Managers, annual Training Conference 2005 (<http://www.epanarpm.org/narpm2005/home.htm>). Managing Site Heterogeneity with Triad Sampling Programs (<http://www.epanarpm.org/narpm2005/pdf/ManagingSiteHeterogeneity.pdf>)
- Norberg T, 2006. Introduktion till Bayesiansk uppdatering. Kurslitteratur till kurs i matematisk statistik för V2, Chalmers. <http://www.math.chalmers.se/Stat/Grundutb/CTH/mve265/0809/>. 2009-05-28.
- Norberg T, 2008. Personlig kommunikation. 08-02-20.
- Norrman J, Back P.-E, Engelke F, Segó L. & O. Wik, 2009. Provtagningsstrategier för förorenad jord. Hållbar Sanering, Rapport 5888. Naturvårdsverket, Stockholm.
- NV, 1997a. Generella riktvärden för förorenad mark, beräkningsprinciper och vägledning för tillämpning. Rapport 4638. Naturvårdsverket, Stockholm.
- NV, 1997b. Bakgrundshalter i mark. Halter av vissa metaller och organiska ämnen i jord i tätort och på landsbygd. Rapport 4640. Naturvårdsverket, Stockholm.
- NV, 1999. Metodik för inventering av förorenade områden. Rapport 4918. Naturvårdsverket, Stockholm.
- NV, 2007. Riktvärden för förorenad mark - Modellbeskrivning och vägledning. Bilaga 1 Sammanställning av indata till riktvärdesmodellen. Remissversion 2007-10-19.
- NV, 2008. Beskrivande statistik och presentation. <http://www.naturvardsverket.se/sv/Tillstandet-i-miljon/Miljoovervakning/Handledning-for-miljoovervakning/Utformning-av-program-och-statistik/>. 2008-01-08
- Oden N.L, 1984. Assessing the significance of a spatial correlogram. *Geographical Analysis*, 16:1-16.
- Oracle, 2007. Crystal Ball 7.3 Reference Manual.
- Pardo-Igúzquiza, E., 1998. Maximum likelihood estimation of spatial covariance parameters. *Mathematical Geology*, 30(1):95-108.

- ProUCL, 2008. Statistical Software ProUCL 4.0 for Environmental Applications For Data Sets with and without Nondetect Observations. <http://www.epa.gov/esd/tsc/software.htm> (2008-08-29)
- Rosén L, Back P-E, Soutukorva Å, Söderqvist T & L, Grahn, 2008a. Kostnads-nyttoanalys som verktyg för prioritering av efterbehandlingsinsatser. Hållbar Sanering, Rapport 5836. Naturvårdsverket, Stockholm.
- Rosén L, Englov P & L, Grahn, 2008b. Förorenade områden: System för riskklassificering och prioritering (SRP). Standard BVS 585.82. Banverket
- SADA, 2008. Spatial Analysis and Decision Assistance. Software. University of Tennessee, UT. <http://www.tiem.utk.edu/~sada/index.shtml> (08-06-29)
- Starzec P, Purucker T & R Stewart, 2008. Osäkerheter i riskbedömning och beslutprocess. Hållbar Sanering, Rapport 5804. Naturvårdsverket, Stockholm.
- Stewart R.N, Purucker S.T, 2008. SADA: A freeware decision support tool integrating GIS, sample design, spatial modeling, and risk assessment. Environmental Modelling and Software (in review).
- Thayer W.C, Griffith D.A, Goodrum P.E, Diamond G.L, Hassett J.M, 2003. Applications of geostatistics to risk assessment. Risk Analysis, 23(5):945-960.
- USEPA, 2002. Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan. EPA QA/G-5S. EPA/240/R-02/005, December 2002, Office of Environmental Information, US EPA, Washington DC. <http://www.epa.gov/quality/qs-docs/g5s-final.pdf> (08-06-26)
- USEPA, 2004. ProUCL Version 3.0 User's Guide. EPA/600/R04/079, April 2004. Office of Research and Development, US EPA. <http://www.epa.gov/esd/tsc/images/proucl3apr04.pdf> (08-11-06)
- USEPA, 2006. Data Quality Assessment: Statistical Methods for Data Practitioners. EPA QA/G-9S. EPA/240/B-06/003, February 2006, Office of Environmental Information, US EPA, Washington DC. <http://www.epa.gov/quality/qs-docs/g9s-final.pdf> (08-10-27)
- USEPA, 2007. ProUCL Version 4.0 Technical Guide. EPA/600/R-07/041, April 2007, Office of Research and Development, US EPA, Washington DC. <http://www.epa.gov/esd/tsc/images/proucl4technical.pdf> (08-10-18)
- White J. W, 1999. Hazards of short-term exposure to arsenic contaminated soil. Office of Environmental Health Assessment Services, Washington State Department of Health.

Bilaga A. Statistisk programvara

I den här bilagan nämns kortfattat tre programvaror som är utvecklade specifikt för förorenad mark och som alla tre är gratis att ladda ner från nätet. ProUCL och SADA har framförallt använts för beräkningarna i exemplen i rapporten. VSP användes framförallt i rapporten *Provtagningsstrategier för förorenad jord* (Norrman et al. 2009).

ProUCL

ProUCL är en programvara som utför flera av de statistiska analyser som rekommenderas i US EPAs datakvalitetsmanual (DQO). Grundläggande analys av data kan utföras (t.ex. histogram, Q-Q-plottar) och *goodness-of-fit*-tester, hypotesprövning samt beräkning av UCLM ingår i paketet. Programvaran är mycket lättanvänd. Den är utvecklad under ledning av Technical Support Center vid US EPA (<http://www.epa.gov/esd/tsc/software.htm>).

SADA – Spatial Analysis and Decision Assistance

SADA är en programvara utvecklat för visualisering, rumslig analys, statistisk analys, hälsoriskbedömning, miljöriskbedömning, kostnadsbedömningar, provtagningsdesign och beslutsanalys för förorenade områden. Programmet kan laddas ner gratis och har utvecklats av the Institute of Environmental Modeling vid University of Tennessee (<http://ecology.tiem.utk.edu/~sada/index.shtml>).

VSP – Visual Sample Plan

VSP är en programvara som har utvecklats med syftet att vara ett enkelt verktyg för att utforma ett optimalt, tekniskt genomförbart provtagningsprogram för att karakterisera yttjord, bebyggda ytor, vattenområden eller liknande. Programvaran utnyttjar statistiska och matematiska algoritmer samt presenterar resultat på ett användarvänligt sätt i form av kartor. Programmet har utvecklats av Battelle Memorial Institute vid Pacific Northwest National Laboratory (<http://dgo.pnl.gov/index.htm>).

Bilaga B. Ordlista

Action Limit (AL)	<i>eng.</i> används ofta av US EPA. Betecknar att det är vid det värdet/den halten måste åtgärder av något slag vidtas. I den här rapporten används riktvärdet som svensk motsvarighet.
Akut toxicitet	Observerbar effekt av ett gift på en organism, där exponeringstiden och observationsperioden är kort.
Arbetsplan	Utredningsstrategin dokumenteras i en arbetsplan.
Autokorrelation	Innebär att ett samband (släktskap) finns mellan observationer i tiden eller rummet.
Bakgrundshalt	Naturlig halt plus antropogent diffust tillskott.
Datakvalitets- målsprocessen (DQOP)	Ett systematiskt angreppssätt för att definiera de kriterium som en datainsamlingsstrategi ska uppfylla. Syftet är att datainsamlingen ska ha rätt kvalitet till rimlig kostnad. Metodiken har utvecklats av US EPA.
Datavärdesanalys	En metodik för att bestämma provtagningens omfattning som bygger på att provtagningskostnaden balanseras mot den potentiella ekonomiska nyttan av provtagningen. Nya data värderas ekonomiskt och detta värde vägs mot provtagningskostnaden. Bara provtagningsprogram där värdet är större än kostnaden bör genomföras.
Delområde	En del av ett förorenat område som definieras som en egen målpopulation. Avgränsning av ett delområde baseras bl.a. på förhandskunskap om föroreningsituationen.
Delvolym	Som ovan, men ordet ”delvolym” betonar att man betraktar ett tredimensionellt problem och inte enbart en area
Dynamisk provtagning	Ett samlingsbegrepp för strategier som inbegriper användning av realtidsmätningar för att låta resultaten av mätningarna styra provtagningen i fält.
Effektstorlek	Den förändring eller skillnad man vill kunna upptäcka, här mellan verklig medelhalt och det utifrån stickprovet uppmätta medelvärdet. Effektstorleken kan också beskrivas som den minsta skillnad man anser vara av vikt att upptäcka

Efterbehandling	Åtgärder som varaktigt syftar till att eliminera eller minska den nuvarande och framtida påverkan på hälsa och miljö från föroreningar i mark, grundvatten, sediment och från deponier, byggnader och anläggningar, samt att begränsa inverkan på landskapet.
Egenskapsområde	Se <i>delområde</i> .
Element	En statistisk term på de enheter som utgör <i>populationen</i> . Vid undersökning av förorenad jord utgörs ett element av en viss definierad jordvolym, t.ex. jordvolymen från en skruvborr eller i en provgrop.
Fyllning	Av människan påförda lösa massor som kan bestå av jord, byggavfall, schaktmassor, spån, slig etc.
Fördjupad riskbedömning	Utredning om riskerna med ett förorenat område, risker med dagens situation, bedömning av hur rent det behöver bli samt åtgärdsbehov. Kvantitativa beskrivningar av olika risker samt plats specifika riktvärden används vanligen.
Förenklad riskbedömning	Utredning om riskerna med ett förorenat område, risker med dagens situation, bedömning av hur rent det behöver bli samt åtgärdsbehov. Vanligen används generella riktvärden vid bedömningen.
Förfrågningsunderlag	Handlingar som tas fram som kalkylerbart underlag för upphandling av utredningar, efterbehandlingsåtgärder eller andra aktiviteter som behövs i efterbehandlingsarbetet.
Generellt riktvärde	Riktvärde som gäller för många, men inte alla, objekt i landet. Anger en nivå under vilken risk för oönskade effekter på människor eller miljö inte föreligger.
Goodness-of-fit-test	Ett statistiskt test som undersöker om data följer en viss definierad statistisk fördelning. Det finns flera olika tester utvecklade.
Grey region	Engelskspråkigt begrepp som används av US EPA, <i>effektstorlek</i> .
Hotspot	Engelskt uttryck för ett starkt förorenat och till yta/volym begränsat område i mark eller sediment.

Hypotesprövning	En teknik för att göra uttalanden om en population baserat på ett stickprov av populationen. Hypotestprövningen utförs med hjälp av ett <i>hypotestest</i> .
Hypotestest	En viss statistisk metod som kan användas vid <i>hypotesprövning</i> . Olika hypotestester används beroende vilka antaganden som kan göras om den underliggande populationen som stickprovet kommer från.
Inferens	I statistiska sammanhang: att dra slutsatser om målpopulationen baserat på stickprov.
Konceptuell modell	En begreppsmodell där tillgänglig information används för att beskriva objektet som skall undersökas. Modellen utgör en hypotes som kan förkastas och revideras under projektets gång.
Konfidensgrad	Anger sannolikheten för att det sanna värdet för den uppmätta storheten ligger inom konfidensintervallet. Betecknas ibland även konfidensnivå.
Konfidensintervall	En statistisk term som anger graden av osäkerhet för ett försök eller mätvärde. Anges ofta i form av en punktskattning med felmarginal, till exempel 30 ± 3 , dvs. ett intervall inom vilket t.ex. medelvärdet av en variabel kan ligga. Gränserna för konfidensintervallet benämns undre respektive övre konfidensgräns och beteckas LCL respektive UCL.
Kovarians	Ett statistiskt mått på samvariationen mellan två variabler
Kvantiler	Ett lägesmått sådant att givna proportioner av fördelningens värden är mindre än (eller större än) talen: 0.9-kvantilen är densamma som 90-percentilen. Se även <i>percentiler</i> .
LCLM	<i>Eng. lower confidence limit of the mean</i> . Den undre gränsen av ett konfidensintervall för medelhalten. Inom förorenad mark är oftast denna mindre intressant än UCLM.
Markanvändning	Det ändamål för vilket ett mark- eller vattenområde utnyttjas eller kommer att utnyttjas.
Medelhalt	I denna rapport används medelhalten för att beteckna den verkliga men okända medelhalten i en målpopulation.

Medelvärde	I denna rapport används medelvärdet för att beteckna <i>skattningen</i> av medelhalten, dvs. det medelvärde man observerar från sitt stickprov och som man kan använda för att skatta den verkliga medelhalten.
Minsta detekterbara skillnad	Se <i>effektstorlek</i> .
Målpopulation	Se <i>population</i> .
Naturlig halt	Den halt som skulle finnas utan antropogen påverkan.
Percentiler	Ett lägesmått i en statistisk fördelning sådant att en given andel av fördelningens värden är mindre än vald percentil. Exempelvis är 90-percentilen det värde vid vilket 90% av datamängdens eller fördelningens värden överskrids.
Platsspecifikt riktvärde	Riktvärde framtaget för ett objekt baserat på de speciella förutsättningar som råder vid objektet.
Population	Den totala mängden av <i>element</i> som man önskar dra slutsatser om genom sin undersökning. Vid undersökning av förorenad jord avgränsas populationen ofta som ett delområde eller egenskapsområde.
Prediktionsintervall	Ett prediktionsintervall är ett intervall som talar om vad en enskild mätning av en variabel kan komma att bli.
Provtagningsmönster	Det sätt på vilket provpunkter placeras i rummet.
Provtagningsplan	Den färdiga produkt som beskriver provtagningsstrategin och hur provtagningen ska gå till.
Provtagningsskala	<i>Eng. support</i> . Den volym som ett prov representerar, inkluderandeprovets form och orientering. Provtagningsskalan kan bl.a. ökas genom att ta samlingsprov.
Provtagningsstrategi	Strategin för hur en provtagning läggs upp.
Provtagningsvolym	Den jordvolym som fysiskt tas ut som prov. Provtagningsvolymen är alltid mindre eller lika med <i>provtagningsskalan</i> .
Punktkälla	En föroreningskälla som förorenar miljön (mark, luft, sediment eller vatten) lokalt

Påverkansområde	Område där mark, grundvatten, ytvatten eller sediment är konstaterat förorenade eller på sikt riskerar att bli förorenade via spridning från föroreningskällan.
Riktad provtagning	En typ av bedömningsbaserad provtagning där provpunkternas lägen väljs subjektivt. Ett vanligt fall är att man riktar provpunkterna mot jordvolymmer där man förväntar sig höga föroreningshalter.
Riktvärde	Den halt av förorening där inga oönskade effekter på människor eller miljö förväntas. Se <i>generellt riktvärde</i> respektive <i>platspecifikt riktvärde</i> .
Riskbedömning	De risker, med avseende på människors hälsa eller miljön, som ett förorenat område kan ge upphov till identifieras och kvantifieras. De nivåer som inte utgör risker för människa eller miljö identifieras.
Skanningstest/metod	Metod för snabb översiktlig undersökning.
Semivariansen	Variansen mellan två punkter separerade av ett visst avstånd. Detta avstånd kallas <i>lag</i> .
Slumpmässig	Avser ett val där slumpen och inte människan påverkar valet.
Standardavvikelse	Ett mått på hur mycket de olika värdena i en population avviker från medelvärdet. Standardavvikelsen (σ) är en egenskap hos en sannolikhetsfördelning och definieras som kvadratroten ur variansen för fördelningen.
Statistisk styrka	Sannolikheten att förkasta en nollhypotes som är felaktig. Definieras som $1-\beta$ där β är sannolikheten att inte förkasta en nollhypotes som är felaktig. β kallas även typ II-fel.
Stickprov	En samling slumpmässigt utvalda prov som skall representera den målpopulation man vill undersöka.
Stratifierad provtagning	Provtagning där den volym som ska provtas har delats in i delvolymmer med syfte att varje delvolym ska vara mer homogen än hela volymen. Syftet är att ta fram mer representativ statistik.
Support	Se <i>provtagningskala</i> .

Tillsynsmndighet	Myndighet (dvs. kommun, länsstyrelse eller Naturvårdsverket) med ansvar för tillsynen av ett objekt, vilket innebär att ta emot anmälningar, ställa krav på verksamheten, följa upp kraven och att ge vägledning till dem som utför arbetena, allt enligt miljöbalken.
UCLM	<i>Eng. upper confidence limit of the mean.</i> Den övre gränsen av ett konfidensintervall för medelhalten.
UCLM ₉₅	<i>Eng. 95% upper confidence limit of the mean.</i> Den övre 95%-iga gränsen av ett konfidensintervall för medelhalten.
Utredningsstrategi	Strategin för planeringen och upplägget av utredningen.
Varians	Varians och <i>standardavvikelse</i> är exempel på spridningsmått för en statistisk fördelning, det vill säga ett mått på hur utspridd fördelningens värden är kring väntevärdet.
Åtgärds mål	Funktionsmål som anger hur området kan användas efter åtgärderna (övergripande åtgärds mål), samt kvantitativa mål (mätbara åtgärds mål) som sätts med hänsyn till de risker objekt medför och till åtgärdernas omfattning baserat på vad som är tekniskt möjligt, miljömässigt motiverat och ekonomiskt rimligt.

Bilaga C. Betafördelningen

En standard tvåparametrisk betafördelning för en slumpmässig variabel x på intervallet $[0, 1]$ ges av:

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

Här är $B(\alpha, \beta)$ den vanliga Beta funktionen:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

där Γ är Euler Gamma funktionen. Medelvärdet bestäms av:

$$\mu = \frac{\alpha}{\alpha + \beta}$$

Variansen är:

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}$$

och *mode* (d.v.s. det mest troliga värdet) är:

$$\hat{x} = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Variabeln x är den förorenade andelen av området. Antag att vi har n oberoende observationer, av vilka f är antalet observationer över riktvärdet och g är antalet observationer under riktvärdet. Då gäller att

$$\alpha = f + 1 \quad \text{och} \quad \beta = g + 1,$$

om priorkunskapen om x modelleras som $\alpha = \beta = 1$. Norberg (2008) och Rosén et al. (2009) argumenterar för att när man inte har någon priorkunskap alls, så avspeglas detta bättre genom att ansätta att $\alpha = \beta = 0,5$ för att detta har minst inverkan på slutresultatet. I så fall ges α och β av:

$$\alpha = f + 0,5 \quad \text{och} \quad \beta = g + 0,5.$$

Parametrarna α och β styr betafördelningens plats och form.

Referenser

Norberg, 2008. Personlig kommunikation. 08-02-20.

Rosén L, Norberg T & J Norrman, 2009. Kurs i statistisk dataanalys och tolkning av resultat: Tillämpningar inom förorenad mark. Hållbar Sanering, Rapport 5897. Naturvårdsverket, Stockholm. Under publicering.

Metodik för statistisk utvärdering av miljötekniska undersökningar i jord

RAPPORT 5932

NATURVÅRDSVERKET
ISBN 978-91-620-5932-3
ISSN 0282-7298

Rapporten beskriver man kan strukturera statistisk utvärdering av tillgänglig information från förorenade områden, både förhandskunskap och insamlad data. Syftet är att lyfta fram statistiska metoder som kan användas för vissa delar i en riskbedömning, men också som underlag för kostnadsuppskattningar i en åtgärdsutredning. Rapporten gör inte anspråk på att ge ett komplett arbetsschema för utredningar av förorenade områden, utan snarare en hjälp till att förstå var olika statistiska metoder kan komma till användning.

Naturvårdsverket har inte tagit ställning till innehållet i rapporten. Författarna svarar ensamma för innehåll, slutsatser och eventuella rekommendationer.

Kunskapsprogrammet Hållbar Sanering samlar in, bygger upp och sprider kunskap om förorenade mark- och vattenområden. Genom Hållbar Sanering kan myndigheter, forskare och företag söka bidrag för utredningar, seminarier och utvecklingsprojekt som täcker kunskapsluckor på kort och lång sikt. Hållbar Sanering styrs av en programkommitté som består av representanter från Banverket, Göteborgs stad, KTH, Linköpings Universitet, Länsstyrelsen i Kalmar, Naturvårdsverket, Norges Teknisk- Naturvetenskaplige Universitet; SGI, SLU, Sydkraft SAKAB och Umeå Universitet.

