

Provtagningsstrategier för förorenad jord

RAPPORT 5888 • JULI 2009



Kunskapsprogrammet


**HÄLLBAR
SANERING**



Provtagningsstrategier för förorenad jord

Jenny Norrman, Statens Geotekniska Institut

Pär-Erik Back, SWECO

Fredric Engelke, Statens Geotekniska Institut

Landon Segó, Pacific Northwest National Laboratory

Ola Wik, Statens Geotekniska Institut

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 978-91-620-5888-3.pdf

ISSN 0282-7298

Elektronisk publikation

© 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 ”Provtagningsstrategier för förorenad jord” som har genomförts inom Hållbar Sanering. Rapporten beskriver en metodik för att planera provtagning av jord inom förorenade områden, där syftet av provtagningen är av överordnad karaktär och där man ställer krav på att hantera de stora osäkerheter som föreligger i samband med karaktärisering av förorenad jord.

Två nya metoder för att beräkna hur många prover som skall samlas in för att uppnå en viss säkerhet i resultatet har utvecklats inom projektet. Metoden som bygger på hypotestestning har tagits fram av Landon Segó (PNNL) och metoden som bygger på ett acceptabelt konfidensintervall har tagits fram av Pär-Erik Back (SWECO).

Arbetet i föreliggande rapport har utförts vid Statens Geotekniska Institut (SGI) i samarbete med SWECO och Pacific Northwest National Laboratory (PNNL). Rapporten har författats av Jenny Norrman (SGI), Pär-Erik Back (SWECO), Fredric Engelke (SGI), Landon Segó (PNNL) och Ola Wik (SGI).

Under arbetets gång har vi fått värdefulla kommentarer från Anders Bank (Envipro Miljöteknik). Ett seminarium genomfördes med Anders Bank, 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 uppskattade kommentarer för det fortsatta arbetet.

Slutligen har rapporten granskats av Anders Bank, Thomas von Kronhelm, Johan Wigh (Länsstyrelsen i Skåne), Thomas Holm (SWECO) samt Mikael Stark (SGI). Deras kommentarer har varit mycket värdefulla.

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		6
SUMMARY		8
1	INLEDNING	11
1.1	Bakgrund	11
1.2	Syfte	11
1.3	Rapportens upplägg och avgränsning	12
1.4	Definitioner och begrepp	13
2	GENOMFÖRANDE AV EFTERBEHANDLINGSPROJEKT	14
2.1	Projekt mål	14
2.2	Utredningsstrategi	14
2.2.1	Konceptuell modell	16
2.2.2	Provtagningsstrategi	16
2.2.3	Provtagningsplan	17
2.2.4	Genomförande av provtagning	17
2.2.5	Datautvärdering	17
2.2.6	Datakvalitetskedjan i efterbehandlingsprojekt	18
2.3	Rapportering	19
3	METODIK FÖR ATT UTFORMA EN PROVTAGNINGSTRATEGI	20
3.1	Steg 1. Definiera provtagningens syfte	21
3.2	Steg 2. Bestäm hur förhandskunskap ska hanteras	22
3.2.1	Kunskap om området	23
3.2.2	Kunskap om aktuella föroreningar	23
3.2.3	Hantering av förhandskunskapen	24
3.3	Steg 3. Definiera rumsliga och tidsmässiga avgränsningar	25
3.4	Steg 4. Bestäm provtagningsskalan	28
3.5	Steg 5. Välj angreppssätt	29
3.5.1	Sannolikhetsbaserat angreppssätt	29
3.5.2	Bedömningsbaserat angreppssätt	30
3.5.3	Sökbaserat angreppssätt	31
3.5.4	Avvägningar vid val av angreppssätt	31
3.6	Steg 6a: Antal prov med sannolikhetsbaserat angreppssätt	32
3.6.1	Välj statistisk parameter	32
3.6.2	Gör en uppskattning av variabiliteten	33
3.6.3	Bestäm önskad säkerhet i resultat	35
3.6.4	Välj provtagningsmönster	37
3.6.5	Bestäm antalet prover	38
3.7	Steg 6b: Antal prov med bedömningsbaserat angreppssätt	43
3.7.1	Formulera hypotes om föroreningsituationen	44

3.7.2	Bestäm typ av prov	44
3.7.3	Bestäm antal prover	44
3.7.4	Bestäm provtagningspunkternas lägen	44
3.7.5	Definiera hur data ska tolkas	45
3.8	Steg 6c: Antal prov med sökbaserat angreppssätt	45
3.8.1	Ange objektets form och storlek	46
3.8.2	Välj modell och provtagningsmönster	46
3.8.3	Bestäm önskad säkerhet i resultaten	47
3.8.4	Bestäm antalet prov	48
4	ALTERNATIVA ANGREPPSSÄTT	49
4.1	Datavärdesanalys	49
4.2	Dynamisk provtagning, TRIAD	51
4.2.1	Kombinerade dataset	54
4.2.2	Sekventiell provtagning	55
4.2.3	Anpassad samlingsprovtagning	55
4.2.4	Rankbaserad provtagning	56
4.2.5	Anpassad klusterprovtagning	57
4.2.6	Geobayesiansk metod	57
4.3	Beviskedjor - Lines of evidence	58
5	DISKUSSION OCH REKOMMENDATIONER	59
6	REFERENSER	61
BILAGOR	64	
A	Exempelsamling	64
B	Ordlista	77
C	Provtagningsmönster	82
D	Hypotestest, styrka och antalet prov	85
E	Antal prov från en lognormalfördelad population baserat på konfidensintervall	93
F	Determining the sample size required for testing the mean of lognormal data	95
G	Hjälpgrafer för att avgöra lämpligt antal prover	112

Sammanfattning

För en långsiktigt hållbar hantering av förorenade områden är det viktigt att efterbehandlingsåtgärder sätts in där de verkligen behövs. Bedömningen av var och i vilken utsträckning sanering är nödvändig baseras till stor del på resultaten av provtagningar. Eftersom en provtagning innehåller flera typer av osäkerheter kan felaktiga bedömningar göras om dessa osäkerheter inte hanteras korrekt.

Syftet med det projekt som redovisas i föreliggande rapport har varit att utveckla och beskriva en arbetsmetodik som kan användas för planering av provtagning av förorenad jord. Den föreslagna arbetsmetodiken utgår från att de frågeställningar man vill besvara med provtagningen styr hur provtagningen bör utformas. Ett annat viktigt syfte med projektet har varit att lyfta fram statistiska metoder som kan användas i planeringen för att hantera de stora osäkerheter som finns vid provtagning av förorenad jord.

Rapporten är indelad i tre huvudsakliga delar: huvudrapport, exempel-samling och ett antal teoretiska beskrivningar i bilagor.

Huvudrapporten är inriktad på att kortfattat beskriva den föreslagna metodiken för att planera provtagningar. Metodiken innehåller följande sex generella moment: (1) Definiera provtagningens syfte, (2) Bestäm hur förhandskunskap ska hanteras, (3) Definiera rumsliga och tidsmässiga avgränsningar, (4) Bestäm provtagningskala och typ av prov, (5) Välj angreppssätt och (6) Bestäm antal prov och placering. De tre olika angreppssätt som beskrivs är: (i) sannolikhetsbaserat angreppssätt, (ii) bedömningsbaserat angreppssätt och (iii) sökbaserat angreppssätt. Både ett sannolikhetsbaserat och ett sökbaserat angreppssätt tillåter att man ställer krav på säkerhet i resultaten för att beräkna antalet prov. Alla angreppssätt har det gemensamt att en noggrann konceptuell modell är grunden för en lyckad provtagningsstrategi.

Det som framförallt är nytt för det sannolikhetsbaserade angreppssättet är att det ges praktiska metoder för att beräkna antalet prover som behövs för att uppnå givna krav på säkerhet i resultaten, både för normal- och lognormalfördelad data. Det introduceras grafer i rapporten som kan användas dels för att dimensionera antalet prover utifrån ställda krav på säkerhet, men också för att avläsa vilken säkerhet man kan förvänta sig med ett givet antal prover, dvs. en låst budget. För ett bedömningsbaserat angreppssätt betonas vikten av att sätta upp en hypotes för hur föroreningen ser ut samt att på förhand definiera hur data bör tolkas beroende på analysvar. För ett sökbaserat angreppssätt betonas vikten av att använda sig av kompletterande metoder för att antalet prover skall bli rimligt.

Metodiken skall inte ses som en enkel kedja av steg som man successivt arbetar sig igenom utan ett iterativt arbetssätt rekommenderas där man går tillbaka en eller flera gånger och ser över och eventuellt reviderar de antaganden och skattningar som gjorts.

Metodiken riktar sig främst mot ”traditionella projekt”, dvs. projekt där man planerar sin provtagning på kontoret, etablerar sig i fält och samlar in

data för att därefter återvända till kontoret för att utvärdera och analyserad data. Detta är fortfarande det vanligaste sättet att arbeta och syftet med föreliggande arbete är att skapa en struktur för att planera den här typen av traditionella projekt, där man tydligt definierar syftet med provtagningen och vad man vill uppnå. Metodiken kan vara en grund för konsulten att kommunicera krav på säkerhet med beställare och tillsynsmyndigheter.

Ett antal alternativa strategier att utforma provtagningen som inte fångas av den föreslagna metodiken går också igenom. Datavärdesanalys, där man väger kostnader och nyttor, finns beskrivet med ett enkelt exempel. Även ett mer dynamiskt arbetssätt (TRIAD) med analys av insamlade prover direkt i fält går igenom, tillsammans med korta beskrivningar av dynamiska strategier som är lämpliga vid ett sådant arbetssätt. Så kallade beviskedjor är ytterligare en strategi som även berörs kortfattat.

I exempelsamlingen beskrivs tillämpning av metodiken som beskrivits i huvudrapporten för fem delområden relativt kortfattat. Exempelen som beskrivs är: (A) avgränsning av förorening från en dieselcistern ovan mark, (B) avgränsning och uppskattning av mängd vid en impregneringsanläggning, (C) uppskattning av medelhalt i en gammal upplagsyta, (D) sökning efter nedgrävt impregneringssalt, och (E) uppskattning av medelhalt i ett fyllnadsområde. Exempelsamlingen kan vara ett stöd för att tillämpa metodiken.

Summary

For a sustainable management of contaminated sites, it is important that remediation measures are applied where they are needed. The assessment of the extent of remediation is based largely upon the results from environmental sampling. Since the data contains several types of uncertainties, misjudgements can be made if these uncertainties are not treated correctly.

The objective of this report is to describe a methodology to plan sampling in soil at contaminated sites, where the objective of the sampling is the superior characteristic and where demands are put upon the management of the large uncertainties present when characterising contaminated soil.

The report is divided into three main parts: the main report, a description of five examples and a number of more theoretical descriptions as attachments.

The main report shortly describes the proposed methodology for planning sampling. The methodology contains the following six general steps: (1) Define the objective of sampling, (2) Decide how prior knowledge should be treated, (3) Define spatial and temporal delimitation, (4) Decide sample support and type of sample, (5) Choose approach, and (6) Decide number of samples and their positions. The three different approaches described are: (i) a probability-based approach, (ii) a judgemental approach, and (iii) a search-based approach. Both the probability-based approach and the search-based approach allows for demands on the level of uncertainty in the results to calculate the number of samples.

News regarding the probability-based approach is the practical methods to calculate the number of samples needed under certain demands on the uncertainty in data, for both normally and lognormally distributed data. Four diagrams are introduced in the report which can be used to estimate the number of samples needed given a certain level of acceptable uncertainty. They can also be used to estimate the level of uncertainty in data that can be expected given a certain number of data, i.e. given a certain budget. For a judgemental approach the hypothesis of the contamination situation and the idea of how data should be interpreted depending on the results are emphasised. Regarding the search-based approach, it is emphasised that complementing methods should be used in order for the number of samples to be reasonable.

The methodology should not be seen as a simple step-wise chain successively worked through, but rather an iterative mode of working where one moves back and forth and reviews the assumptions and estimations made.

The methodology is designed for traditional projects, i.e. projects where the sampling campaign is planned in the office, a field campaign is performed and the analyst returns to the office where data is analysed. This is still the most common way of working and the aim of this work is to create a structure in order to plan these types of traditional projects, where the sampling objective is clearly defined. The methodology can be a basis for the consultancy to communicate demands on uncertainties with the problem-owner and the authorities.

A number of alternative strategies are described: data worth analysis where costs and benefits are included, some dynamic strategies suitable for a more dynamic working approach (TRIAD), and lines of evidence.

In the examples, the application of the methodology in the main report is described rather short for five different problems. The examples described are: (A) delimitation of contamination from a diesel tank above ground, (B) delimitation and estimation of amount of contamination at a former wood preserving facility, (C) estimation of the mean concentration at a former storage area for the treated wood, (D) search for buried preserving chemicals, and (E) estimation of the mean concentration in fillings. The collection of examples can be used as a support to apply the methodology.

1 Inledning

1.1 Bakgrund

För en långsiktigt hållbar hantering av förorenade områden är det viktigt att efterbehandlingsåtgärder sätts in där de verkligen behövs. Bedömningen av var och i vilken utsträckning sanering är nödvändig baseras till stor del på resultaten av provtagningar. Eftersom en provtagning innehåller flera typer av osäkerheter kan felaktiga bedömningar göras om dessa osäkerheter inte hanteras korrekt. Om föroreningar felaktigt lämnas utan åtgärd kan detta leda till långsiktiga risker för människor och miljön, eller till framtida oförutsedda kostnader när den kvarvarande föroreningen upptäcks. På motsvarande sätt leder efterbehandling av ”ren” jord till onödiga kostnader, vilket innebär minskade resurser till andra objekt som kan behöva efterbehandlas. En viktig faktor för att uppnå en långsiktigt hållbar hantering av förorenade områden är att undersökningar utförs så att insamlad data ger en grund för att kunna fatta beslut med acceptabel osäkerhet.

Inom ramen för Naturvårdsverkets kunskapsprogram Hållbar Sanering genomfördes projektet ”Provtagningsstrategier för förorenade områden” (Engelke & Norrman, 2009). Projektet bestod av följande delar:

- inventering av befintliga svenska krav och rekommendationer för provtagningsstrategier
- utvärdering av svenska och danska undersökningar
- internationell inventering av koncept och strategier inklusive databaserade verktyg för jord, grundvatten och porgas.

I projektet konstaterades att det finns flera olika nationella och internationella vägledningar för hur en provtagning bör utföras. Däremot saknas en svensk vägledning vad gäller att effektivisera provtagningar avseende strategi, riktlinjer för hur stora osäkerheter som kan accepteras och hur val av strategi kan kopplas till exempelvis en riskbedömning. Utländska vägledningar och verktyg för utrednings- och provtagningsstrategier kan utnyttjas som underlag men behöver anpassas efter svenska förhållanden.

1.2 Syfte

Syftet med det projekt som redovisas i föreliggande rapport har varit att utveckla och beskriva en arbetsmetodik som kan användas för planering av provtagning av förorenad jord. Den föreslagna arbetsmetodiken utgår från att de frågeställningar man vill besvara med provtagningen styr hur provtagningen bör utformas. Ett annat viktigt syfte med projektet har varit att lyfta fram statistiska metoder som kan användas i planeringen för att hantera de stora osäkerheter som finns vid provtagning av förorenad jord.

1.3 Rapportens upplägg och avgränsning

Rapporten är indelad i tre delar:

- Huvudrapporten
- Exempelsamlingen i Bilaga A
- Teoretiska beskrivningar i Bilaga C – F

Huvudrapporten är inriktad på att kortfattat beskriva en metodik för att planera provtagningar. Den har en kort introduktion om de olika komponenterna i ett efterbehandlingsprojekt (kapitel 2), och går därefter in på själva metodikbeskrivningen (kapitel 3). Metodikbeskrivningen riktar sig främst mot relativt traditionella projekt, dvs. projekt som inledningsvis planerar en provtagning, sedan etablerar sig i fält och genomför provtagningen och därefter analyserar data. Kapitel 4 går igenom ett antal alternativa strategier, dels inriktad mot datavärdesanalys där man väger kostnader och nytta, dels ett mer dynamiskt arbetssätt (enligt TRIAD) med korta beskrivningar till de dynamiska strategier som är lämpliga. Avslutningsvis i kapitel 4 nämns beviskedjor kort. Kapitel 5 ger en kort diskussion och några rekommendationer, dels baserat på huvudrapporten, men även baserat på exemplen i exempelsamlingen.

Exempelsamlingen i Bilaga A har först en introduktion till området som exemplifieras där de olika delområdena beskrivs med avseende på föroreningshistorik etc. I efterföljande avsnitt ges exempel på tillämpning av metodiken som beskrivits i huvudrapporten för fem delområden. Exempelgenomgångarna är relativt kortfattade. Exempelen som beskrivs är: (A) avgränsning av förorening från en dieselcistern ovan mark, (B) avgränsning och uppskattning av mängd vid en impregneringsanläggning, (C) uppskattning av medelhalt i en gammal upplagsyta, (D) sökning efter nedgrävt impregneringssalt, och (E) uppskattning av medelhalt i ett fyllnadsområde.

För att inte göra texten för lång och teknisk har en del teoretiska beskrivningar och härledningar lagts i Bilagorna C – F. En bilaga är på engelska och beskriver en metod för att skatta antalet prover för lognormalfördelad data givet vissa krav på osäkerheten. Den är skriven av statistikern Landon Sego från Pacific Northwest National Laboratory (PNNL).

Ambitionen har varit att skriva rapporten så att innehållet är tillgängligt och förståeligt för personer med grundläggande statistiska kunskaper och bred allmänkompetens inom arbete med förorenade områden, t.ex. handläggare på länsstyrelser och kommuner, konsulter, efterbehandlingsföretag och exploatörer. Målsättningen har också varit att använda referenser som finns tillgängliga på nätet i så stor utsträckning som möjligt, med bifogade länkar för att läsaren lätt skall kunna fördjupa sig i ytterligare material.

Programvara som har använts är framförallt Excel eller gratisprogramvaran Visual Sample Plan (VSP, 2008). VSP är en användarvänlig programvara som är utvecklad av PNNL för US EPA och har betydligt fler funktioner än de som exemplifieras här.

Rapporten behandlar endast provtagning av jord *in-situ*¹, men grundprinciperna i den arbetsmetodik som presenteras i rapporten gäller även för andra medier.

Den vägledning Naturvårdsverket har gett ut tidigare angående kvalitets-säkring vid miljötekniska undersökningar "Rätt datakvalitet" (NV, 1996), griper över ett bredare område än föreliggande rapport. Här har istället fokuserats på att beskriva den stegvisa arbetsmetodiken för att planera provtagning av jord samt att lyfta fram några i sammanhanget användbara statistiska metoder.

1.4 Definitioner och begrepp

I Bilaga B finns en ordlista med begreppsförklaringar, men här görs även ett förtydligande angående definitionen av vad som är förorenat. I Kvalitetsmanualen utgåva 4 (NV, 2008) definieras ett förorenat område som synonymt med efterbehandlingsobjekt: ett område, deponi, mark, grundvatten eller sediment som är förorenat och vars halter påtagligt överskrider lokal/ regional bakgrundshalt.

Utgångspunkten i denna rapport har varit att ett område är förorenat om den representativa halten² överskrider ett riktvärde. Riktvärdet kan i det här fallet vara både generella riktvärdet, ett platsspecifikt riktvärde eller ett annat uppställt åtgärds mål. Det kan givetvis också vara bakgrundhalten. Motivet för att tillämpa denna definition har varit att förenkla diskussionerna något.

Amerikanska naturvårdsverket (US EPA) använder begreppet AL – *Action Level*, för att beteckna att det är vid den föroreningshalten som åtgärder av något slag bör vidtas, t.ex. i form av efterbehandling eller vidare utredning. Det bör dock tilläggas att bara för att den representativa halten inom ett område överskrider ett riktvärde så är det inte säkert att en aktiv efterbehandlingsinsats krävs. I Sverige avgörs detta i en s.k. riskvärderingsprocess.

¹ Med *in-situ* avses här provtagning av jord direkt på plats, alltså inte jord från upplag eller deponier, sorterat eller siktat material.

² Med "representativ halt" avses en halt som ska representera ett mått på föroreningsgraden i ett område. En sådan halt kan t.ex. vara medelhalten, den 95%-iga övre konfidensgränsen för medelhalten eller en percentil i den statistiska fördelningen av föroreningshalter i området. I rapporten används framförallt den 95%-iga övre konfidensgränsen för medelhalten som representativ halt.

2 Genomförande av efterbehandlingsprojekt

Utformandet av strategier för utredning och provtagning av förorenade områden är viktigt för att kunna genomföra kostnadseffektiva undersökningar. Planering och genomförandet av undersökningar av förorenade områden kan delas in i en hierarkisk struktur, se Figur 2-1. Denna hierarkiska struktur innebär att kvaliteten i genomförandet av en viss hierarkisk nivå avgörs av kvaliteten i de övre (ofta tidigare) nivåerna. Systemet innehåller även stora mått av dynamik som innebär att genomförandet/kvaliteten på olika nivåer påverkar varandra. Ett bra genomförande av ett projekt förutsätter därför att arbetet sker med en samlad helhetssyn.

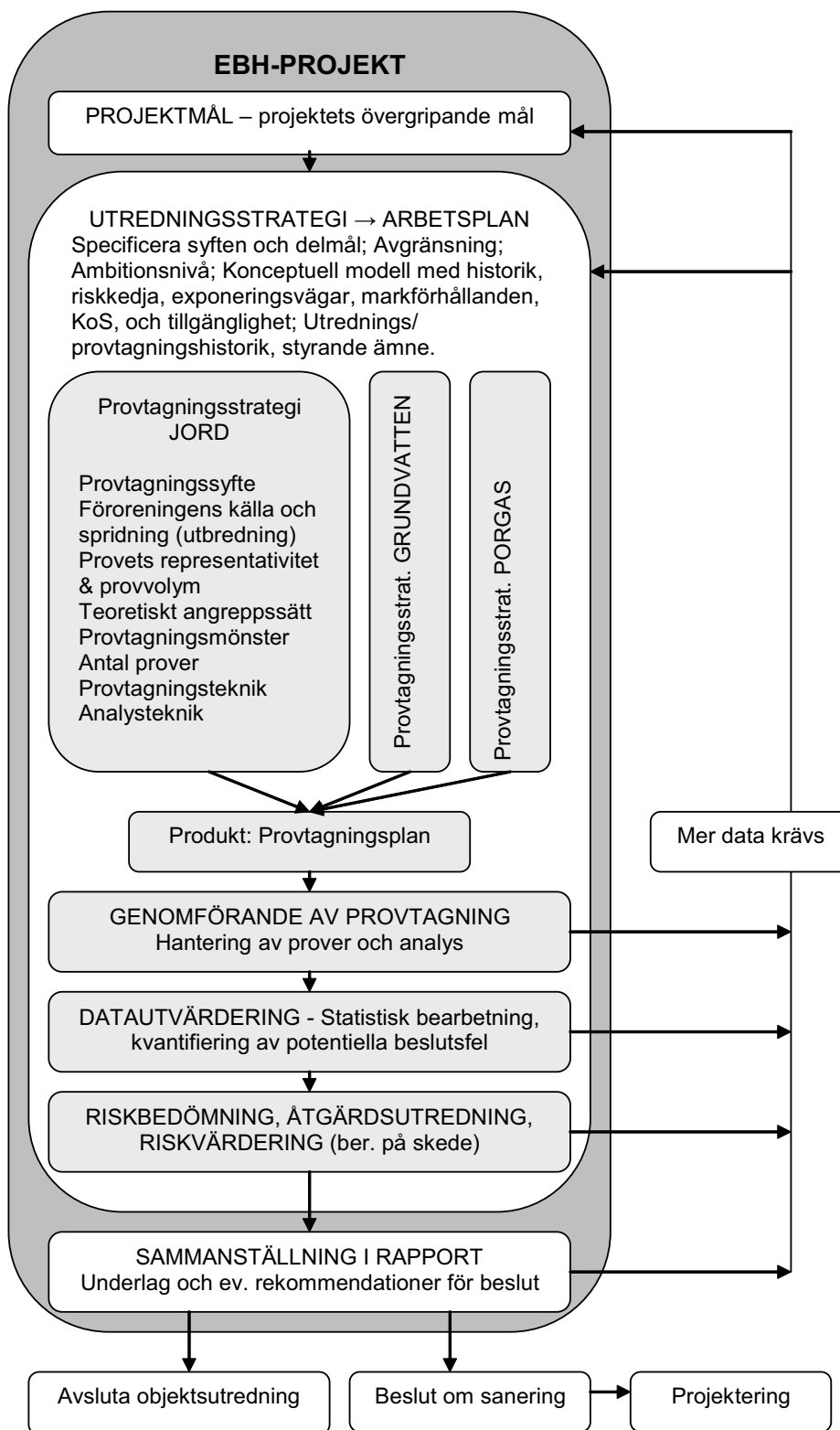
Möjligheten att utforma en provtagningsstrategi kan dock påverkas av upphandlingsformen. Till exempel kan lagen om offentlig upphandling (LOU) ibland göra det svårt att ta fram en optimal provtagningsstrategi för ett område då en iterativ process kan vara problematisk att handla upp. Provtagningsstrategin är ofta fördefinierad i förfrågningsunderlaget för att enklare kunna jämföra olika anbud, men förfarandet riskerar ibland att låsa utförandet och kan därmed vara ett hinder för en iterativ process.

2.1 Projektmål

Projektets övergripande mål bör vara klarlagt innan en utredning påbörjas. Vanliga projektmål är till exempel att kunna avgöra om ett område utgör en risk för miljö eller hälsa, om det behöver saneras, var saneringen skall ske, men även hur den bör ske.

2.2 Utredningsstrategi

När ett projekt påbörjats bör en utredningsstrategi utformas. Ofta görs denna av eller i samråd med beställaren, eller så är den given utifrån kvalitetsmanualen. Vi har här definierat utredningsstrategi att innefatta hela utredningen, dvs. från planering av provtagning och insamling av data till själva utvärderingen av den insamlade informationen (se Figur 2-1). Utredningsstrategin beskrivs i en arbetsplan. Val av utredningsstrategi är i stor grad beroende på vilken osäkerhet som kan accepteras i t ex riskbedömningar och saneringsomfattning. Att ta fram en utredningsstrategi är för större projekt ett förhållandevis omfattande arbete som kräver lång erfarenhet. All relevant tillgänglig information som t ex. verksamhetshistorik, tidigare undersökningar, arkivmaterial, ritningar, flygbilder och geologisk information osv., sammanställs för att skapa en första konceptuell modell för området. Syftet med undersökningen bör om möjligt specificeras ytterligare och gärna även uttryckas kvantifierbara termer



Figur 2-1. Hierarkisk struktur för planering och genomförande av EBH-projekt. KoS står för Känslighet och Skyddsvärde.

Vilka medier som skall provtas för att bäst besvara frågeställningarna bör noggrant utvärderas innan en utredningsstrategi föreslås. För att slutligen kunna bedöma t ex risker eller saneringsbehov krävs nästan alltid provtagning från flera olika medier (jord, grundvatten, sediment, ytvatten, inomhusluft, byggnader osv.) som utvärderas tillsammans med information kring verksamhets-historik, fysiska förhållanden osv.

2.2.1 Konceptuell modell

Genom en konceptuell modell används tillgänglig information för att beskriva objektet som skall undersökas. Triad (2008) använder begreppet *Conceptual Site Model* (CSM) och skriver att en konceptuell modell använder historisk och nutida information för att bedöma vad gäller föroreningshalter:

- var föroreningar finns,
- hur mycket som finns, och
- hur halterna kan variera och vilken grad av rumslig korrelation som kan vara närvarande.

Dessutom bör följande bedömas vad gäller riskbedömning:

- vad sker med föroreningen vad gäller nedbrytning och spridning,
- vem som kan exponeras för föroreningen eller skadliga nedbrytningsprodukter, samt
- vad kan göras för att hantera risker genom att förminska exponering.

En noggrann och bra konceptuell modell urskiljer och avgränsar olika föroreningspopulationer, för vilka beslut rörande risk och efterbehandling skiljer sig åt inom ett och samma förorenade område. Ett exempel på två olika föroreningspopulationer inom t.ex. en industritomt kan vara dels fyllningsmassorna och dels en nedgrävd oljecistern. Dessa två olika typer av föroreningspopulationer bör provtas och efterbehandlas på olika sätt.

Modellen bör uppdateras kontinuerligt med resultat som bekräftar eller förkastar de hypoteser och bedömningar som gjorts i modellen. Tanken med Triad (2008) är att göra denna uppdatering direkt i fält med hjälp av screeninganalyser.

Listan ovan är omfattande och i mindre projekt är det ovanligt att en fullständig och kvantifierbar konceptuell modell tas fram. Den metodik för provtagningsstrategi som föreslås i denna rapport försöker dock översätta den kvalitativa informationen i den konceptuella modellen i kvantitativa termer. Även kvantitativ information bör därför arbetas in i den konceptuella modellen. Däremot är det viktigt att vara införstådd med att den konceptuella modellen är en hypotes som kan förkastas och revideras under projektets gång.

2.2.2 Provtagningsstrategi

Då projektmål, syften och konceptuell modell är specificerade kan arbetet med en effektiv provtagningsstrategi påbörjas. Provtagningsstrategin innefattar här

alltså själva planeringen och upplägget av provtagningen. Den färdiga produkten av detta är provtagningsplanen, se Figur 2-1 och avsnittet nedan. För jord måste hänsyn tas till flera olika faktorer, t ex. hypoteser om föroreningskällor och eventuell spridning, typ av förorening, representativet, önskad säkerhet, provtagningsmönster, provtagningsmetodik, antal prover och ekonomiska ramar. För att kunna genomföra en effektiv provtagning måste kunskap finnas om varför provtagning skall ske, hur och var den skall genomföras, vilka beslut som skall fattas baserat på resultaten, vad resultaten skall användas till samt hur resultaten sedan bör utvärderas. Den här rapporten ger förslag på en arbetsgång för att ta fram en provtagningsstrategi för förorenad jord.

2.2.3 Provtagningsplan

Provtagningsstrategin som utformas ska alltid beskrivas och dokumenteras i en provtagningsplan. Provtagningsplanen bör innehålla all relevant information som gör att provtagningen i princip ska kunna upprepas vid ett senare tillfälle om så önskas. SGFs Fälthandbok för miljötekniska markundersökningar (SGF, 2004) listar kortfattat vad som bör ingå i en provtagningsplan. Det primära syftet är att provtagningsstrategin ska vara så pass väl beskriven i provtagningsplanen att data senare ska kunna utvärderas på ett korrekt sätt. Detta innebär bland annat att även syftet med provtagningen måste beskrivas och vara tydligt. Utan dokumentation riskerar en provtagning att förlora en hel del av sitt värde eftersom bakgrunden till när/var/hur/varför proverna togs saknas. Om denna dokumentation saknas blir det även svårt att i efterhand bedöma hur representativa proverna är och om data kan återanvändas för andra syften i senare skeden av ett efterbehandlingsprojekt.

2.2.4 Genomförande av provtagning

Provtagningen ska genomföras i enlighet med den provtagningsplan som upprättats och där provtagningsstrategin finns definierad. Alla moment, t.ex. fältprovtagning, provhantering och provbearbetning, bör utföras med rätt kvalitet. Det finns flera rapporter och vägledningar som beskriver hur fältprovtagning och provhantering bör utföras, se t.ex. SGF (2004).

Genomförd provtagning bör dokumenteras i en fältrapport tillsammans med analysresultat, jordlagerföljder, mätinstrument, intryck från provtagning, osv. Fältrapporten kan vara en särskild rapport eller en del av utredningens huvudrapport.

2.2.5 Datautvärdering

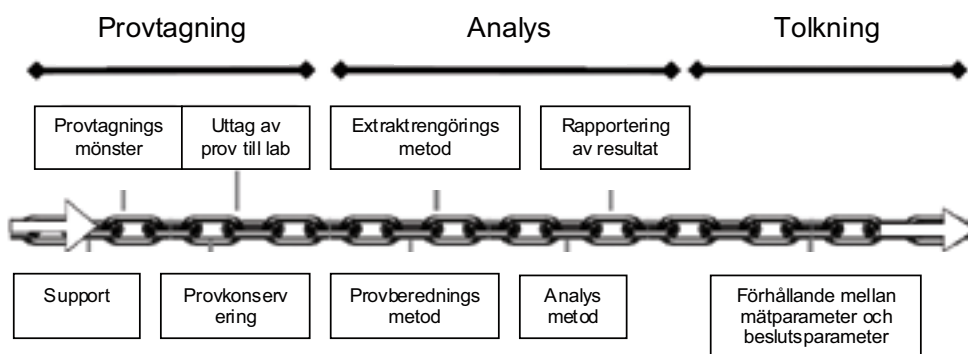
Data ska utvärderas med hänsyn till provtagningens syfte. Det finns således en tydlig koppling mellan provtagning och datautvärdering. Ofta blandas data från olika provtagningar eller från olika analysmetoder och utvärderas tillsammans. Detta kan leda till problem om data kommer från provtagningar med olika syften eller analysmetoder, t.ex. då slumpmässig/systematiska data blandas med data från riktade provtagningar eller då fält- och laboratorieanalyser blandas. Det är därför viktigt att man i datautvärderingen försöker ta hänsyn till sådana skillnader, eller åtminstone försöker beskriva skillnaderna och diskutera vilka konsekvenser dessa kan ge i utvärderingen.

I datautvärderingen kan ingå olika statistiska bearbetningar. Det finns en mängd olika metoder som kan användas, vilka dock faller utanför den här rapportens omfattning. En annan rapport inom kunskapsprogrammet Hållbar Sanering behandlar olika metoder för datautvärdering, Ramverk för statistisk utvärdering av miljötekniska undersökningar i jord (Norrman et al., 2009).

2.2.6 Datakvalitetskedjan i efterbehandlingsprojekt

Datakvalitet brukar definieras som rätt data till rätt kvalitet, dvs. inte nödvändigtvis så mycket data som möjligt med så liten osäkerhet som möjligt, utan helst precis den kvalitet som krävs för att uppnå det mål man har med ett visst projekt (NV, 1996; US EPA, 2006b).

Hög kvalitet på utförandet i en viss nivå kan spolieras av låg kvalitet i en annan nivå, vilket NARPM (2005) åskådliggör med hjälp av en datakvalitetskedja, se Figur 2-2. Varje länk i kedjan utgör en variabel som bidrar till den övergripande kvaliteten i analysresultatet och varje länk i kedjan måste vara intakt för att data skall vara av tillräcklig kvalitet för att användas i beslutsfattande.



Figur 2-2. Datakvalitetskedjan, NARPM (2005).

Osäkerheten i den svagaste länken blir styrande för den totala osäkerheten och det är viktigt att ha en helhetsbild för att förstå var den svagaste länken är för att försöka reducera osäkerheten på rätt ställe. Ett exempel är att nyttan av precisa laboratorieanalyser helt kan försvinna om provtagningen utförs på ett bristfälligt sätt, se exempelrutan nedan. Osäkerheten i laboratorieosäkerheter är generellt sett mindre betydelsefull än provtagningsosäkerheterna.

Osäkerhetsmatematik tydliggör effekten av den svagaste länken i datakvalitetskedjan!

Osäkerheter adderas enligt: $a^2 = b^2 + c^2$

Ett exempel illustrerar hur den totala osäkerheten (TO) beräknas genom att addera den analytiska osäkerheten (AO) och provtagningsosäkerheten (PO), dvs. addition av varianser: $TO^2 = AO^2 + PO^2$

AO är 10 mg/kg och PO är 80 mg/kg → TO = 81 mg/kg

Om man försöker förbättra sin laboratorieanalys för att minska den analytiska osäkerheten till hälften, ger detta följande:

AO är 5 mg/kg och PO är 80 mg/kg → TO = 80 mg/kg

Detta är ingen vidare förbättring. Om man istället försöker minska sin provtagningsosäkerhet med hälften (t.ex. genom att ta fler prover) så får det följande effekt:

AO är 10 mg/kg och PO är 40 mg/kg → TO = 41 mg/kg

Att minska provtagningsosäkerheten till hälften påverkar i detta exempel den totala osäkerheten i mycket högre grad. I själva verket kan man ha en större analytisk osäkerhet utan att detta påverkar resultatet nämnvärt:

AO är 20 mg/kg och PO är 40 mg/kg → TO = 45 mg/kg

Alltså: lägg krutet på den svagaste länken i kedjan!

I ovanstående exempel innebär detta att det är mer värdefullt med ett stort antal mindre noggranna mätningar än ett litet antal mycket noggranna mätningar. Exemplet är hämtat från US EPAs session "Environmental Sampling Part 1" på ConSoil 2008 i Milano.

Tumregel: Man kan bortse från alla osäkerheter som är mindre än en fjärdedel av den största osäkerheten (i och med att man adderar varianser).

2.3 Rapportering

Rapporteringen bör vara överskådlig och fullständig. Bedömningar, hypoteser, vald metodik, strategier, resultat och beräkningar samt källor till osäkerhet bör redovisas i tillräcklig omfattning för att en utomstående med ämneskunskap skall kunna sätta sig in i projektet och granska samtliga avgörande moment. Det är inte heller ovanligt att ett projekt pågår under lång tid och i flera etapper där projektet kan ha flera olika utförare. Detta skärper kraven på en fullgod rapportering ytterligare för att inte kunskap skall försvinna mellan olika etapper och utförare. Detta är särskilt viktigt när data återanvänds för olika syften i flera steg.

NV (1994) ger en generell mall för rapportering med följande huvudrubriker: sammanfattning, uppdrag och syfte, områdesbeskrivning, mark-, vatten- och sedimentförhållanden, historisk redogörelse, utförda undersökningar, föroreningarnas omfattning, utvärdering/riskbedömning, rekommendationer, referenser samt bilagor och figurer.

3 Metodik för att utforma en provtagningsstrategi

I detta kapitel föreslås en metodik för att utforma provtagningsstrategier som kan användas vid miljötekniska markundersökningar. Metodiken avser i första hand provtagning av jord men principerna är i många fall generella och kan tillämpas på andra medier. För vissa medier kan även andra aspekter tillkomma, t.ex. mobiliteten av olika föroreningar i grundvatten och porgas. En förutsättning för att en bra provtagningsstrategi ska kunna tas fram är att det finns en konceptuell modell för området. En konceptuell modell är en beskrivning av förhållandena på området, både kvalitativt och kvantitativt, med avseende på föroreningssituation, geologi, spridningsförhållanden, exponeringsförhållanden osv., se tidigare avsnitt 2.2.1.

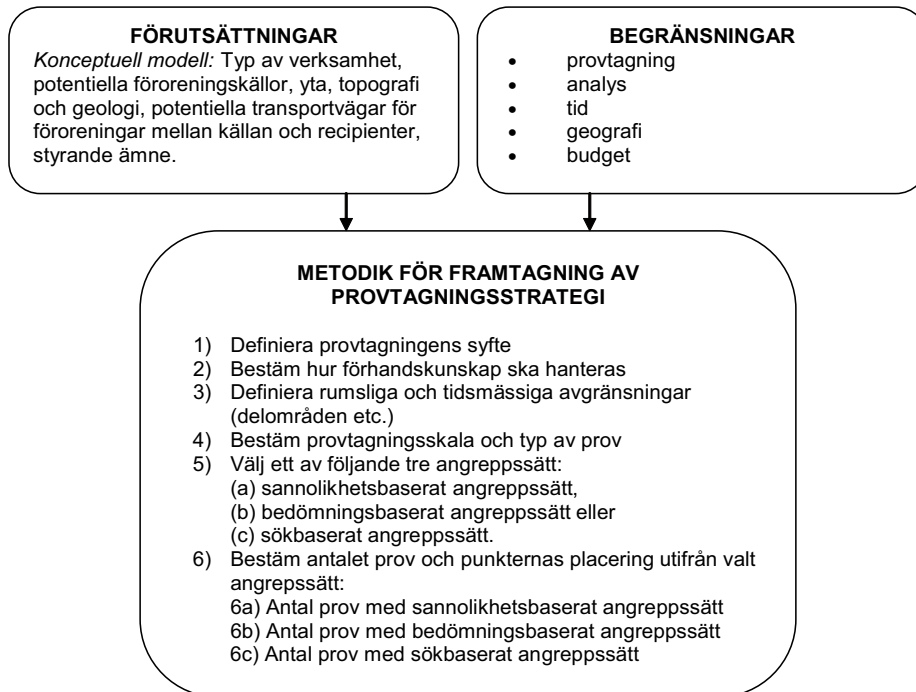
Begränsningar i att ta fram en optimal provtagningsstrategi kan finnas i form av provtagnings- och analysteknik, tid, geografi och budget, samt inte minst LOU vilket alla är faktorer som måste tas hänsyn till under planeringen. En avvägning av hur mycket resurser finns att tillgå för en provtagning bör göras. Avvägningen är beroende av en mängd olika faktorer, t.ex. i vilken fas projektet är (förstudie eller huvudstudie), hur stora riskerna bedöms kunna vara och hur begränsade de ekonomiska ramarna är. Finns det en strategi för att hantera osäkerheter och en definierad osäkerhet som inte får överskridas? Hur stor osäkerhet som är acceptabel är beroende av hur stora konsekvenserna blir av ett möjligt felaktigt beslut. Ambitionsnivån för provtagningen bör stå i relation till konsekvenserna av att fatta fel beslut. Om de negativa konsekvenserna bedöms vara stora, bör ambitionsnivån för provtagningen vara hög.

Metodiken som presenteras för att utforma en provtagningsstrategi innehåller följande sex generella moment (se även Figur 3-1):

- 1) Definiera provtagningens syfte
- 2) Bestäm hur förhandskunskap ska hanteras
- 3) Definiera rumsliga och tidsmässiga avgränsningar (delområden etc.)
- 4) Bestäm provtagningskala och typ av prov
- 5) Välj ett av följande tre angreppssätt:
 - (a) sannolikhetsbaserat angreppssätt,
 - (b) bedömningsbaserat angreppssätt, eller
 - (c) sökbaserat angreppssätt.
- 6) Bestäm antal prov och deras placering utifrån valt angreppssätt:
 - (a) antal prov med sannolikhetsbaserat angreppssätt,
 - (b) antal prov med bedömningsbaserat angreppssätt, eller
 - (c) antal prov med sökbaserat angreppssätt.

De sex momenten beskrivs utförligare i följande avsnitt. De olika stegen redovisas i den ordning som bedöms vara den mest logiska. Metodiken ska dock inte betraktas som en enkel kedja av steg som man successivt arbetar sig igenom från början till slut för att sedan vara färdig. Tvärtom rekommenderas

ett iterativt arbetssätt där man går tillbaka en eller flera gånger och eventuellt reviderar de antaganden och skattningar som gjorts.



Figur 3-1. Översiktsbild över den föreslagna metodiken för framtagning av provtagningsstrategi.

3.1 Steg 1. Definiera provtagningens syfte

Tänk på...

- Syftet med en provtagning ska alltid definieras, även om proverna består av enstaka riktade prover.
- Ett prov utan syfte är meningslöst och kan slopas.
- Provtagning kan ha flera olika syften. Ange i så fall samtliga.
- Dokumentera syftet så att man i framtiden kan avgöra vad data representerar.

Ett av de allra viktigaste stegen då en provtagningsstrategi utformas är att definiera provtagningens syfte. Detta är inte alltid enkelt eftersom en undersökning kan ha flera syften som dessutom kan vara otydligt uttalade. Syftet med provtagningen måste definieras mycket tydligt för att en lämplig strategi ska kunna väljas. Som en hjälp att definiera syftet kan följande frågor ställas (Domburg et al., 1994):

- Vad finns i marken?
- Hur mycket finns det?
- Var finns det?
- Sker förändring över tiden och i vilken grad?

Dessa frågor kan omformuleras till några typiska syften för att karaktärisera föroreningsförekomst i jord³:

- Vad?
 - Belägga förekomst av förorening
 - Identifiera typ av förorening
- Hur mycket?
 - Ta fram representativ statistik för föroreningssituationen
 - Ta fram representativ statistik för bakgrundshalter
- Var?
 - Belägga misstänkt *hotspot*
 - Hitta okänd *hotspot*
 - Avgränsa föroreningsutbredning, horisontellt och vertikalt
 - Avgränsa rena områden
- Förändring?
 - Avgöra om föroreningsmängd eller halt förändras över tiden

Provtagningens syfte bör vara tillräckligt tydligt definierat för att genom utvärdering av mätdata möjliggöra en bedömning om hur väl syftet uppfyllts. Vaga och oprecisa syften ska undvikas. Exempel på ett vagt syfte som inte bör användas är t.ex.: ”Syftet med provtagningen är att undersöka föroreningssituationen i området”. Ett sådant syfte ger ingen eller liten vägledning för provtagningen. De syften som angetts i listan ovan är precisa syften, vilka sedan kan brytas ned i delsyften. Exempelvis kan syftet att ”ta fram representativ statistik av föroreningen i ett egenskapsområde (delområde)” brytas ned till syftena att beräkna föroreningsmängden samt volymen förorenad jord i detta område.

3.2 Steg 2. Bestäm hur förhandskunskap ska hanteras

Tänk på...

- Om det inte redan är gjort är det nu hög tid att ta fram en konceptuell modell.
- Beakta den förhandskunskap du har om området och den verksamhet som förekommit.
- Beakta vilka typer av föroreningar som du förväntar dig och hur dessa kan ha spridits i markprofilen.
- Bestäm hur din förhandskunskap om området ska hanteras vid provtagningen:
 - Beakta förhandskunskapen subjektivt/kvalitativt
 - Bortse från förhandskunskap
 - Beakta förhandskunskapen matematiskt/statistiskt/kvantitativt

³ Man kan tänka sig en mängd andra frågeställningar kopplade till förorenade områden, t.ex. hur föroreningen sprids, om föroreningen rör sig som fri fas, vilka recipienter som spridning sker till etc., men den här rapporten avgränsas till frågeställningar som hör samman med att beskriva den förorenade jordvolymen med avseende på halter, utbredning, variabilitet etc..

Förhandskunskap är sådan information som man redan har innan provtagningen utförs. Denna information kan delas in i:

- specifik kunskap om området och
- generell kunskap om de föreningar som hanterats inom undersökningsobjektet.

Dessa två typer av information diskuteras nedan. Därefter diskuteras olika sätt att hantera förhandskunskapen när en provtagning planeras.

3.2.1 Kunskap om området

Den specifika kunskapen om områdets föroreningssituation kan t.ex. utgöras av historiska uppgifter om verksamheten på platsen, synliga spår på markytan, geologiska förhållanden som påverkar föroreningens uppträdande (t.ex. jordlagerföljder, fyllnadsmassor), byggnader, ledningsgravar etc. Sådan information kan ge en förväntad bild av föroreningssituationen, både beträffande vilka föroreningsämnen som kan finnas i jorden, hur troligt det är att de förekommer, typiska föroreningshalter, rumslig variation, föroreningsmängder, spridningsvägar m.m.

Den specifika kunskapen om området ska användas för att ta fram en konceptuell modell för platsen, i kombination med kunskap om de föroreningstyper som förekommer (se nedan). Historiken är en mycket viktig del i uppbyggnaden av den konceptuella modellen och således även i framtagandet av en provtagningsstrategi. Verksamhet eller ämnen som förbises i den historiska informationen om området utgör ett vanligt fel i miljötekniska utredningar. Vad en konceptuell modell bör omfatta beskrivs kort i avsnitt 2.2.1.

3.2.2 Kunskap om aktuella föroreningar

De ämnen som kan finnas i ett förorenat område har olika fysiska egenskaper. Det innebär att de kommer att uppträda på väsentligt skilda sätt i marken och även innebära olika exponeringsrisker. Man kan grovt dela in olika föroreningar i fyra grupper:

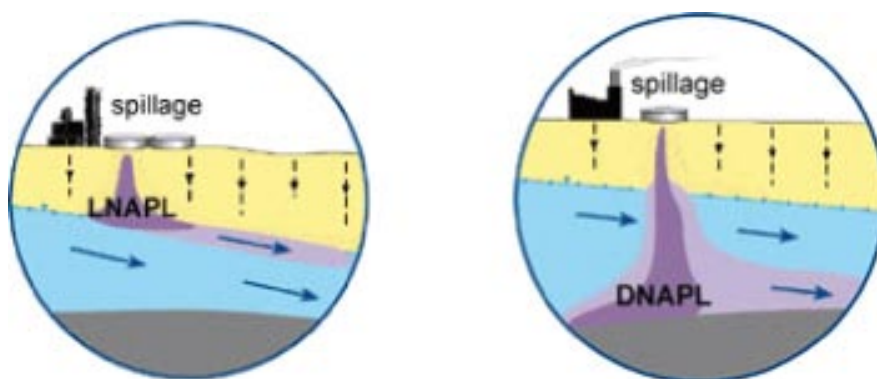
- Adsorberande ämnen
- Flytande icke vattenlösliga vätskor (eng. *LNAPL, light non-aqueous phase liquid*)
- Sjunkande icke vattenlösliga vätskor (eng. *DNAPL, dense non-aqueous phase liquid*)
- Vattenlösliga ämnen

Adsorberande ämnen är exempelvis metaller och tunga organiska ämnen som dioxiner som adsorberar till jordpartiklar. Flytande vätskor är sådan som har en lägre densitet än vatten. Ett typiskt exempel på en flytande vätska är diesel eller bensin, som kan spridas och röra sig som fri fas ovanpå en vattenyta. Sjunkande vätskor är sådana som har högre densitet än vatten och därför sjunker ner genom markprofilen till dess tätande jordlager eller sprickfattigt berg påträffas. Ett typiskt exempel på en sjunkande vätska är trikloretylen.

Vattenlösliga ämnen kan t.ex. vara vissa typer av pesticider eller kemikalier som svavelsyra, salter m.m.

Det är välkänt att ovanstående ämnesgrupper sprids på helt olika sätt i en markprofil. Denna generella kunskap måste man beakta när man lägger upp sin provtagningsstrategi. Provtagning av förorenad mark är ett tredimensionellt (3D) problem, vilket blir tydligast för vätskor som kan flöda såväl horisontellt som vertikalt. Ett misstag vid provtagning av förorenad mark är att inte beakta den tredje dimensionen (djupet) i tillräckligt hög grad, vilket gör att en förorening som spridits på djupet riskerar att missas.

Då man väljer angreppssätt för provtagningen måste ämnenas olika spridningsegenskaper beaktas. En sannolikhetsbaserad provtagning (avsnitt 3.6 i en stor jordvolym (tre dimensioner) är exempelvis inte särskilt meningsfullt om föroreningen bara förväntas längst ner i denna volym eller i området runt grundvattenytan (flytande vätska). Då kan ett tvådimensionellt bedömningsbaserat angreppssätt vara mer relevant, då provpunkterna riktas dit föroreningen förväntas finnas.



Figur 3-2. En flytande vätska (LNAPL) kan påträffas ovan grundvattenytan, medan en sjunkande vätska (DNAPL) har en högre densitet än vatten och sjunker genom markprofilen, UNEP (2006)

3.2.3 Hantering av förhandskunskapen

Förhandskunskapen har ofta ett stort värde, men den kan innehålla fel eller vara missvisande beroende på hur kunskapen har erhållits. Därför är det inte alltid självklart på vilket sätt förhandskunskapen ska värderas och hanteras. Inför en provtagning finns det olika möjligheter att använda förhandskunskapen i provtagningsstrategin.

Ett sätt att ta hänsyn till förhandskunskap är att dela in det förorenade området i olika egenskapsområden (delområden) med avseende på föroreningshistorik, typ av förorening och annat. Detta bör *alltid* göras och det behandlas separat i Steg 3: Definiera rumsliga och tidsmässiga avgränsningar (avsnitt 3.3). Resten av diskussionen i detta stycke förutsätter att man alltid utgår från olika egenskapsområden.

Ett inte helt ovanligt sätt är att basera provtagningen helt och hållet på förhandskunskapen (bedömningsbaserat angreppssätt, se avsnitt 3.5.2).

Fördelen är att all befintlig kunskap utnyttjas, oavsett vilken typ av information det rör sig om. Nackdelen är att det blir svårt att bedöma osäkerheten i undersökningen eftersom provtagningen till stora delar blir subjektiv och beroende av bedömningar. En annan nackdel är att förhandskunskapen kan vara felaktig (t.ex. på grund av brister i den historiska inventeringen), vilket kan leda till att felaktiga slutsatser görs med ledning av provtagningsresultaten.

Det är även i vissa fall motiverat att bortse från eventuell förhandskunskap och välja ett renodlat statistiskt angreppssätt. Fördelen med detta är att representativ statistik, antal prov och graden av osäkerhet kan beräknas statistiskt. Nackdelen är att ingen hänsyn tas till sådant som man faktiskt känner till, vilket kan leda till en ineffektiv och kostsam provtagning och i värsta fall att helt fel slutsatser dras.

En kombination är att uttrycka sin förhandskunskap matematiskt/statistiskt så att den kunskapen kan hanteras med ett statistiskt angreppssätt. Ett exempel är t.ex. att göra en skattning av sannolikheten att det skall finnas en *hotspot* i ett område och väga in den kunskapen när provtagningen planeras. Fördelen med att uttrycka sin förhandskunskap matematiskt/statistiskt är att det blir möjligt att väga in sina förhandskunskaper för att ta fram representativ statistik och bedöma graden av säkerhet/osäkerhet. Nackdelen är att det kan vara svårt att uttrycka förhandskunskapen statistiskt, speciellt om man inte är van att göra den typen av bedömningar/skattningar. Dessutom krävs ofta mer komplicerade beräkningar än om ingen hänsyn tas till förhandskunskapen.

I praktiken är det vanligt att förhandskunskapen hanteras på olika sätt för olika delområden och olika frågeställningar.

3.3 Steg 3. Definiera rumsliga och tidsmässiga avgränsningar

Tänk på...

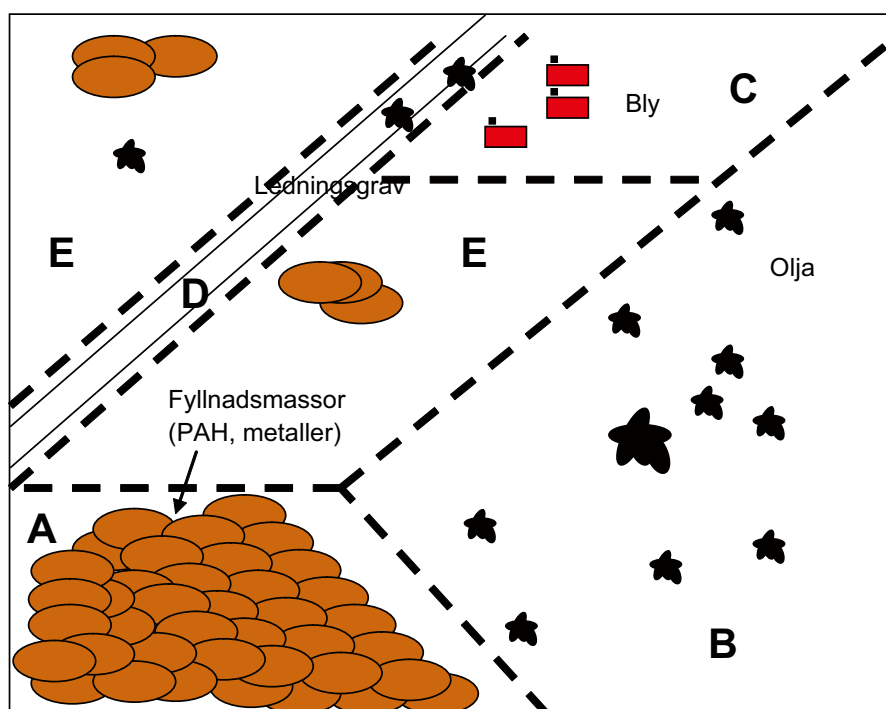
- Om det inte redan är gjort är det nu hög tid att ta fram en konceptuell modell!
- Dela in området i egenskapsområden (delområden) som är förhållandevis homogena.
- Dela in jordprofilen vertikalt i lager som stämmer överens med den konceptuella modellen och geologiska lager.
- Bestäm vilket/vilka kemiska och fysikaliska parametrar som ska analyseras.

För att resultatet från en provtagning ska bli tolkningsbart måste man redan när provtagningen planeras definiera vilken frågeställning proverna skall försöka besvara. En viktig sak att definiera är vilken eller vilka kemiska och fysikaliska parametrar som ska analyseras. Detta baseras på vilka föroreningar som förväntas på platsen och syftet med provtagningen. Förutom detta ska rumsliga och tidsmässiga avgränsningar göras.

Ett förorenat område är i regel heterogent ur föroreningssynpunkt, med delområden som kan vara opåverkade av föroreningar medan andra delar

är kraftigt förorenade (se Figur 3-3). I sådana fall är det orimligt att ta prov som skall vara representativa för hela området. Dessutom kan koncentrationerna av vissa ämnen förändras över tiden (nedbrytning, omvandling, utlakning, sanering), vilket gör att en tidsmässig avgränsning kan behöva göras. Provtagningen bör alltså avgränsas i både tid⁴ och rum⁵.

Den rumsliga (spatiala) avgränsningen kan göras genom att dela in området i delområden eller delvolym. Indelningen görs så att varje del blir någorlunda homogent. Uppdelningen kan göras av markytan och/eller mot djupet, principen är densamma. Det allra bästa är om avgränsningen görs så att varje delområde/delvolym blir någorlunda statistiskt homogent, dvs. att medelvärde, varians m.m. inte varierar alltför kraftigt inom området/volymen. Detta är svårt att göra innan föroreningsituationen och de geologiska förhållandena är kända och istället bör därför avgränsningen baseras på den konceptuella modellen. I takt med att området provtas och resultaten utvärderas bör även den konceptuella modellen uppdateras och vidareutvecklas (jfr det dynamiska arbetssätt som beskrivs i avsnitt 4.2).



Figur 3-3. Schematisk bild av ett förorenat område. Föroreningen är heterogent fördelad över området. En indelning i 5 delområden (A-E) har gjorts, baserat på tidigare verksamheter och övriga förhållanden, där varje delområde då blir mer homogent.

⁴ I jord betraktas ofta förändringen över tid beroende på naturliga processer som mycket liten, varför man sällan tar hänsyn detta utom t.ex. då man vill undersöka skillnaden före och efter sanering. Tiden kan dock spela stor roll för flyktiga ämnen eller ämnens biotillgänglighet och farlighet samt vid helt nya spill.

⁵ Som redan framgått är provtagning av jord ett 3-dimensionellt provtagningsproblem. Den tredje dimensionen (djupet) måste beaktas i lika hög grad som de horisontella dimensionerna. Om detta inte görs kan grova fel begås, se avsnitt 3.2.2.

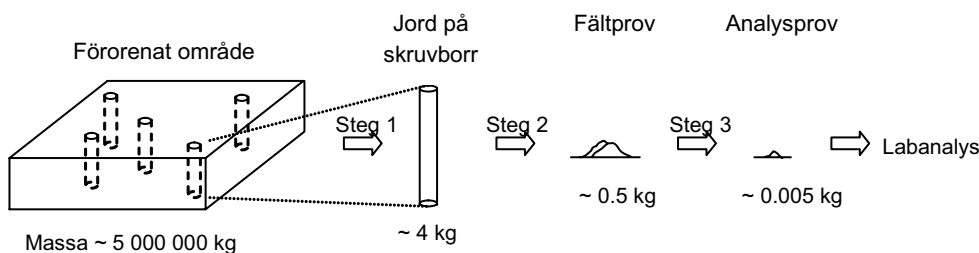
Avgränsningen av delområdena kan baseras på:

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

Avgränsningen måste göras objektspecifikt. Ju bättre avgränsning som görs, desto lättare blir det att utvärdera data. Den mest detaljerade indelningen kan göras inför en sanering genom att definiera selektiva efterbehandlingsenheter, se NV (1997).

Efter avgränsningen utgör varje delområde en s.k. målpopulation som ska karaktäriseras. Målpopulationen är den totala mängd av koncentrationvärden i delområdet som vi önskar kunna säga något om (i princip en oändlig mängd). Eftersom målpopulationen inte är känd måste vi ta prover för att karaktärisera den. Samlingen av prov som tas från målpopulationen utgör stickprovet⁶. Om proverna tas på ett genomtänkt sätt och är tillräckligt många kommer stickprovet att likna målpopulationen och man kan på basis av stickprovet uttala sig om egenskaperna hos målpopulationen. Med hjälp av statistiska metoder kan man också säga något om hur stor avvikelse det kan vara mellan målpopulationen och stickprovets egenskaper, dvs. osäkerheten i egenskaperna hos målpopulationen kan bedömas.

Figur 3-4 visar att provtagningskedjan består av flera steg. Rumsliga avgränsningar måste göras i varje steg men är viktigast vid provtagning i fält.



Figur 3-4. Exempel på provtagningskedja bestående av tre provtagningssteg, två i fält och ett i laboratorium. Målpopulationen har avgränsats horisontellt och vertikalt. Ett stickprov bestående av fem enskilda prov samlas in (Back, 2003).

⁶ Stickprov är med statistisk terminologi den mängd prover som tas för att karaktärisera en jordvolym, inte varje enskilt prov.

3.4 Steg 4. Bestäm provtagningsskalan

Tänk på...

- Provtagningsskalan är den volym som provet representerar, Provtagningsskalan är INTE detsamma som volymen på det prov som skickas till laboratorium (provvolymen). Provtagningsskalan är oftast större, ibland mycket större, än provvolymen.
- Provtagningsskalan ska alltid anges i rapporter som skrivs, så att data kan värderas i efterhand.
- Om akuttoxiska risker ska undersökas bör provtagningsskalan vara liten.
- Om långtidsrisker ska undersökas kan provtagningsskalan vara stor.

Provtagningsskalan är den volym ett prov⁷ representerar (*eng. support*; se Starks, 1986), samt volymens form och orientering i rummet. Denna skala beror på provtagningsstekniken men även på hur provtagningen läggs upp. En lämplig provtagningskala måste väljas med hänsyn till undersökningens syfte, bl.a. vilka krav som ställs på data i riskbedömningen. Om man exempelvis är ute efter att bedöma akutrisken för småbarn (intag av arsenikförorenad jord) kan en skala i storleksordningen cm-dm vara lämplig. Man vill i detta fall veta hur föroreningskoncentrationerna varierar i denna lilla skala. Om däremot spridning till grundvatten är problemet man önskar utreda så har haltvariationer i denna lilla skala ingen betydelse. Då kan istället skalan 10 m eller större vara intressant, dock helt beroende på föroreningens egenskaper.

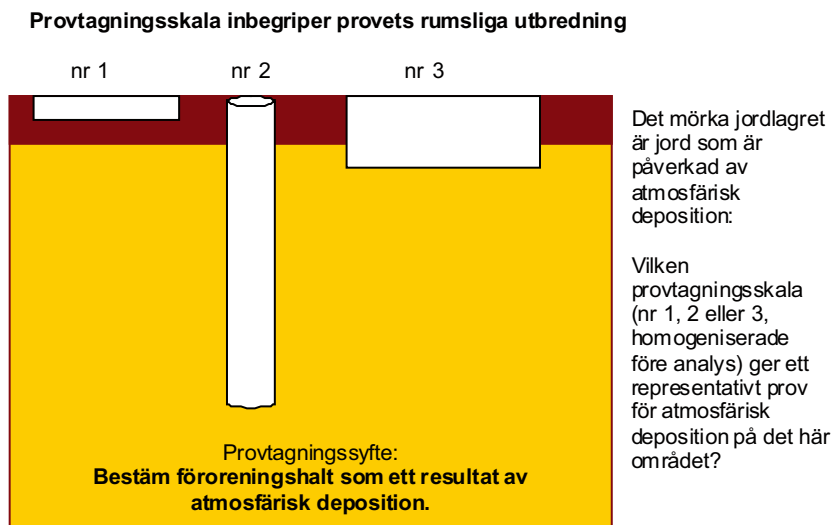
Provtagningsskalan kan ökas på två sätt:

- ta större provvolym jord eller
- ta flera enskilda prov inom en definierad yta/volym och slå samman dessa till samlingsprov som det görs en enda analys på.

Figur 3-5 ger exempel på olika provtagningsskalor samt även betydelsen av provvolymens form och orientering i rummet (*eng support*). Beroende på syftet med provtagningen är de olika lämpliga. Om syftet är att ta ett representativt prov för atmosfärisk deposition så är provtagningskala nr 1 mest lämplig i figuren nedan.

Prov tagna med skruvborr representerar ofta en betydligt mindre skala än jordprover från provgropar.

⁷ Resonemanget i avsnitt 3.4 baseras på provtagning i jord.



Figur 3-5. Exempel på olika provtagningsskalor (support) vid provtagning av jord. Mest lämplig skala ges av provtagningssyftet. Anpassad efter NARPM 2005.

3.5 Steg 5. Välj angreppssätt

Tänk på...

- Välj ett sannolikhetsbaserat angreppssätt baserat på indelning i egenskapsområden istället för ett bedömningsbaserat, om det är praktiskt och ekonomiskt rimligt eftersom det möjliggör en redovisning av osäkerheter.
- Om ett bedömningsbaserat angreppssätt ändå väljs bör det så långt möjligt likna ett sannolikhetsbaserat. Detta innebär bl.a. att man inte bör rikta alla prov mot enbart misstänkt förorenade delar, utan se till att provpunkterna får en viss spridning mot områden som inte bedöms vara förorenade.
- Vid sökning efter *hotspots* bör man väga in förhandskunskapen. Om möjligt bör man göra detta på ett sätt som gör det möjligt att beräkna osäkerheten, dvs. med ett sök-baserat angreppssätt.

Tre alternativa och grundläggande typer av angreppssätt för att lägga upp en provtagningsstrategi listas nedan och beskrivs i de följande avsnitten:

- 1) sannolikhetsbaserade angreppssätt,
- 2) bedömningsbaserade angreppssätt och
- 3) sökbaserade angreppssätt.

3.5.1 Sannolikhetsbaserat angreppssätt

Ett sannolikhetsbaserat angreppssätt bör väljas när man vill ta fram representativ statistik över ett område eller ett delområde. Sådan statistik ligger till grund för beräkning av representativa halter som sedan kan jämföras med riktvärden i en riskbedömning. Representativ statistik kan även användas för att beräkna föroreningsmängder och mängd förorenad jord.

Grunden för ett sannolikhetsbaserat angreppssätt är att varje liten delvolym jord har en känd sannolikhet att bli vald vid provtagningen⁸. I provtagningslitteraturen beskrivs en mängd olika sätt hur sannolikhetsbaserad provtagning kan utföras, dvs. hur proverna väljs i tid och rum, se avsnitt 3.6. Om ett sannolikhetsbaserat angreppssätt väljs blir det möjligt att beräkna antalet prover som krävs för att nå en viss säkerhet i resultatet. Det blir även möjligt att använda resultatet för att ta fram representativ statistik utan att riskera systematiska provtagningsfel för området eller delområdet. Därmed blir det också möjligt att beräkna osäkerheten i resultat, exempelvis osäkerheter i mängdberäkningar. Med ett sannolikhetsbaserat angreppssätt blir det exempelvis möjligt att visa att riktvärdet riskerar att överskridas med en viss säkerhet eller att uttrycka osäkerheten i en volym- eller kostnadsberäkning med relevanta intervall.

De tre vanligaste sannolikhetsbaserade provtagningsmönstren är (se även Bilaga C):

- slumpmässig provtagning,
- systematisk provtagning, eller
- systematisk slumpmässig provtagning.

3.5.2 Bedömningsbaserat angreppssätt

Med ett bedömningsbaserat angreppssätt baseras antalet prover och provpunkternas lägen på mer eller mindre subjektiva bedömningar som bygger på den förhandskunskap som finns om föroreningsituationen. Angreppssättet kan också väljas om det inte är praktiskt möjligt (tidsmässigt och ekonomiskt) att ta representativa prover från målpopulationen på ett sannolikhetsbaserat sätt. Begränsad budget är i sig inte ett särskilt bra skäl för att välja ett bedömningsbaserat angreppssätt. Om kostnaderna för ett sannolikhetsbaserat angreppssätt ter sig orimligt höga kan det vara lämpligare att sänka kravet på säkerhet i resultatet (se avsnitt 3.6.3) men behålla det sannolikhetsbaserade angreppssättet, vilket gör att kostnaderna kan minskas.

Vid ett bedömningsbaserat angreppssätt kan det vara svårt eller omöjligt att avgöra hur stor osäkerhet som är förknippad med resultatet. Hur stor osäkerheten blir beror till stor del på kvaliteten på bakgrundsinformation, exempelvis platsens historik och lägen för förorenande verksamheter inom området.

Det finns minst två typer av bedömningsbaserad provtagning med olika syften:

- riktad provtagning och
- radiell provtagning.

⁸ | sin enklaste utformning innebär sannolikhetsbaserad provtagning att varje delvolym jord dessutom har samma sannolikhet att bli vald.

Syftet med en riktad provtagning är typiskt att belägga förekomst av förorening, medan en radiell provtagning syftar till att avgränsa en förorening (se t.ex. NV, 1994).

3.5.3 Sökbaserat angreppssätt

Med ett sökbaserat angreppssätt är syftet att försöka hitta föroreningskällor eller andra objekt med okänd lokalisering, exempelvis *hotspots* eller föroreningsplymer i grundvatten. Metodiken baseras på sök teori och innebär att ett provtagningsmönster väljs så att sannolikheten att hitta objektet kan beräknas. Hur framgångsrikt angreppssättet är beror till stor del på hur tillförlitlig förhandskunskapen är. Utan förhandskunskap blir metodiken ineffektiv. För att öka effektiviteten kan därför geofysiska undersökningsmetoder eller andra screeningmetoder vara användbara.

Med ett sökbaserat angreppssätt hanteras förhandskunskapen så att det blir möjligt att kvantitativt beräkna hur framgångsrik strategin blir, till skillnad från ett bedömningsbaserat angreppssätt. Det sökbaserade angreppssättet fungerar även i fall där det inte finns någon förhandskunskap.

De vanligaste provtagningsmönstren som används för att söka efter *hotspots* är systematiska, även om slumpmässiga mönster också förekommer.

3.5.4 Avvägningar vid val av angreppssätt

Valet av angreppssätt bör baseras på syftet med provtagningen. Om ett bedömningsbaserat angreppssätt väljs så bör det efterlikna ett sannolikhetsbaserat angreppssätt i så stor utsträckning som möjligt så att data senare kan återanvändas i projektet. Detta innebär bl.a. att man inte bör rikta in alla prov mot vissa mindre delar utan istället se till att provpunkterna får en viss spridning, i likhet med en sannolikhetsbaserad provtagning. Det är bara med ett sannolikhetsbaserat angreppssätt som man kan göra en statistiskt korrekt beskrivning av undersökningsområdet, dvs. målpopulationen.

I efterbehandlingsprojekt är det bedömningsbaserade angreppssättet lämpligast i tidiga projektskeden, t.ex. i MIFO fas 2, då syftet är att bekräfta eller förkasta att förorening förekommer inom undersökningsområdet. Detta innebär att förhandsinformationen då får mycket stor betydelse.

Ett sannolikhetsbaserat angreppssätt är lämpligast att tillämpa i senare projektskeden, exempelvis för jämförelse med olika riktvärden, riskbedömningar, volymberäkningar eller andra frågeställningar som uppkommer i huvudstudier eller inför en sanering (NV, 1997).

Ett sökbaserat angreppssätt är lämpligast när objekt eller *hotspots* måste lokaliseras och förhandskunskapen är svag, t.ex. vid sökning efter nedgrävda tunnor eller mindre områden med mycket höga föroreningshalter.

3.6 Steg 6a: Antal prov med sannolikhetsbaserat angreppssätt

Tänk på...

- Gör en bedömning av vilken typ av statistisk fördelning du förväntar dig hos data. Koncentrationsdata från förorenade områden är vanligen inte normalfördelade, även om sådana fall också kan finnas, utan istället skevt fördelade (t ex lognormal- eller gammafördelade).
- Det krävs fler prover från områden med heterogen föroreningssituation än från mer homogena områden, givet att man vill nå samma grad av säkerhet i resultatet.
- Det kan finnas flera alternativa metoder att bestämma antalet prover, t.ex. metoder som bygger på konfidensintervall eller hypotestest. För att undvika felbedömningar bör man då välja en metod där man förstår teorin.
- Med hjälp av graferna i detta avsnitt kan man snabbt bedöma antalet prover som krävs för olika situationer, utan komplicerade beräkningar.
- Om antalet prover som krävs blir mycket stort måste man antingen utöka provtagningsbudgeten eller sänka kravet på säkerhet hos medelvärdet.
- Om antalet prover som ska tas är givet på förhand kan man istället använda graferna i detta avsnitt för att bedöma vilken säkerhet man kan förvänta sig.
- De statistiska beräkningarna av medelhalt och säkerheten i bestämning av medelhalt är inte kopplat till ett områdes storlek. Det finns dock en koppling mellan storleken på ett område och de osäkerheter man är beredd att acceptera, dvs. de kostnader som en felbedömning är förknippad med.

Då provtagningsstrategin ska utformas och ett sannolikhetsbaserat angreppssätt används bör följande moment genomföras:

- Välj statistisk parameter
- Gör en uppskattning av variabiliteten
- Bestäm önskad säkerhet i resultat
- Välj provtagningsmönster
- Bestäm antalet prov

Metodiken innebär att man på förhand försöker räkna ut hur många prover som krävs för att uppnå en önskad säkerhet. Graden av säkerhet måste man själv ange och denna säkerhet ska avse den statistiska parameter som man vill bestämma, vanligen medelvärdet. Ett exempel kan vara om man med minst 95% säkerhet vill veta om medelvärdet i undersökningsområdet underskrider riktvärdet, eller om det räcker med 90%, eller kanske 80% säkerhet. Att uppnå 99% säkerhet kräver ofta ett mycket stort antal prover. Vilken säkerhet man vill uppnå måste avgöras från fall till fall eftersom den är kopplad till de konsekvenser som kan uppstå om man på basis av sin provtagning fattar fel beslut.

3.6.1 Välj statistisk parameter

För att kunna bestämma antalet prover måste man välja vilken statistisk parameter som ska bestämmas, t.ex. medel, max eller medianvärdet. Valet bör

kopplas till provtagningens syfte. Ofta är medelvärdet eller en viss percentil av föroreningspopulationen intressant att bestämma för att användas vid jämförelse mot riktvärden, men det är även vanligt att undersöka andra statistiska parametrar som t.ex. en proportion av en population. En statistiska parameter som väljs i form av en halt kan betecknas som ”representativ halt” om halten på ett bra sätt representerar föroreningssituationen i området.

Det finns inga egentliga riktlinjer för vilken representativ halt som skall jämföras med riktvärden från Naturvårdsverket vid tidpunkten för denna rapport, men arbete pågår med att ta fram sådana rekommendationer. US EPA arbetar ofta med den 95%-iga övre konfidensgränsen för medelhalten, UCL95⁹ (US EPA, 2008). Även Naturvårdsverket (NV, 1997) har använt sig av denna statistiska parameter men med beteckningen UCLM¹⁰.

Vanliga beskrivande statistiska parametrar för ett stickprov

Stickprovet innehåller n stycken enskilda prover, x_1, x_2, \dots, x_n .

$$\text{Aritmetiskt medelvärde, } m: m = \bar{x} = \frac{1}{n} \sum_i x_i$$

$$\text{Varians, } Var: Var = s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

$$\text{Standardavvikelse, } s: s = \sqrt{s^2}$$

$$\text{Variationskoefficient, } CV: CV = \frac{s}{m}$$

3.6.2 Gör en uppskattning av variabiliteten

Antalet prover som krävs för en provtagning är bland annat beroende på hur stor variationen är inom undersökningsområdet. Därför måste en bedömning av variabiliteten göras, även i de fall inga tidigare provtagningar gjorts. Om provdata finns sedan tidigare kan dessa användas för att uppskatta variabiliteten. I annat fall måste bedömningen baseras på erfarenheter från liknande områden eller rimlighetsbedömningar.

Variabiliteten kan uttryckas som varians (V), standardavvikelse (s) eller variationskoefficient (CV). Standardavvikelsen är kvadratroten ur variansen och variationskoefficienten CV är standardavvikelsen dividerat med medelvärdet.

Om data förväntas vara skevt fördelade¹¹, t.ex. lognormalfördelade, är det lämpligt att skatta variabiliteten som CV . Orsaken är att det kan vara

⁹ UCL – Upper Confidence Limit

¹⁰ UCLM – Upper Confidence Limit of the Mean.

¹¹ Koncentrationsdata från förorenad jord brukar generellt vara skevt fördelade, dvs. det finns relativt många mätningar med låga värden och betydligt färre som har höga. En fördelningskurva (t.ex. histogram) för en sådan datamängd ser då ”skev” ut genom att den har en lång svans åt höger. Lognormalfördelningen är en mer eller mindre skev fördelning beroende på dess parametrar, men det finns även andra fördelningsmodeller som kan beskriva en skev datamängd.

svårt att ge rimliga värden på standardavvikelsen när data är skevt fördelade men detta är något lättare om man använder CV. Man kan erfarenhetsmässigt göra bedömningar av hur stor CV brukar vara vid olika förorenings-situationer. Mätdata från liknande föroreningsobjekt är då värdefulla. Tabell 3-1 nedan kan vara till hjälp om data från liknande objekt saknas.

Variationskoefficient, CV:

Variationskoefficienten (*eng. Coefficient of Variation, CV*) definieras som standardavvikelsen dividerat med medelvärdet. CV är ett mått på variationen i data och är dimensionslös, dvs. CV saknar enhet.

Variabiliteten beror på flera faktorer, bl.a. hur föroeningen spridits. Den kan variera mellan olika föroeningar inom samma område. Organiska föroeningar uppvisar ofta, men inte alltid, större variabilitet än metaller. Variabiliteten beror även på provtagnings-skalan. Om provtagnings-skalan ökas, t.ex. genom samlingsprov, kommer CV att minska.

Frågor att ställa sig:

- Förväntas datamängden från det förorenade området att vara relativt homogen? Då är antagligen CV lägre än eller omkring 1.
- Förväntas CV vara avsevärt högre än 1, dvs. en heterogen datamängd? Kommer data från olika målpopulationer? Då bör en annorlunda indelning i delområden övervägas.

Tabell 3-1. Variationskoefficienten CV och förslag på hur den kan tolkas för förorenad jord. Tabellen kan användas som hjälp vid skattning av variabilitet vid planering av en provtagning.

CV	Kommentar
< 0,5	Mycket liten variation i data, homogen datamängd. Data är troligen normalfördelade.
0,5 – 1	Måttlig variation i data, relativt homogen datamängd. Data kan troligen betraktas som normalfördelade.
1 – 1,5	Relativt stor variation i data. Data följer en skev fördelning, t.ex. lognormalfördelning. Dataspannet är några tiopotenser.
1,5 – 2	Stor variation i data, heterogen datamängd. Data följer en skev fördelning, t.ex. lognormalfördelning. Dataspannet är några tiopotenser.
2 – 3	Mycket stor variation i data, mycket heterogen datamängd. Data följer en mycket skev fördelning, t.ex. lognormalfördelning. Dataspannet är åtskilliga tiopotenser. Kontrollera om en annan indelning i mer homogena delområden kan göras.
> 3	Extremt stor variation i data, extremt heterogen datamängd. Dataspannet är åtskilliga tiopotenser. Gör en annan indelning i mer homogena delområden.

Om data förväntas vara någorlunda normalfördelade kan variabiliteten skattas och uttryckas som en standardavvikelse. Detta görs bäst med hjälp av tidigare erfarenheter och mätdata från liknande objekt. Ett annat alternativ är att göra en rimlighetsbedömning om vilken max- och min-halt som förväntas i området vid provtagningen. Därefter kan standardavvikelsen s skattas enligt:

$$s = \frac{\max - \min}{4} \quad (\text{Ekv. 3-1})$$

Ekvationen gäller om spannet max-min omfattar ca 95% av alla data. Om spannet istället omfattar ca 99% av data kan 4:an ersättas med en 6:a i ekvationen. Om programvaran VSP (2008) används för att beräkna antalet prover bör man se upp med att använda default-värden för standardavvikelsen i dialogrutorna, då dessa är mycket låga.

3.6.3 Bestäm önskad säkerhet i resultat

Den önskade säkerheten måste anges. För att kunna göra detta måste man bestämma vilka föroreningar som är mest kritiska att analysera. Vanligen är det ett fåtal ämnen som förväntas förekomma i högst halter relativt riktvärdet. Graden av säkerhet formuleras på olika sätt beroende på vilken metod man vill använda för att beräkna antalet prov. Två huvudtyper kan användas:

- Metoder som bygger på konfidensintervall
- Metoder som bygger på hypotestestning

Vid praktiska tillämpningar bör man välja metod där man förstår bakomliggande teori. Det kan även vara lämpligt att använda sig av båda metoderna för att se om slutsatserna blir ungefär desamma

METODER SOM BYGGER PÅ KONFIDENSINTERVALL

Med konfidensintervall avses en statistisk term som anger graden av osäkerhet för ett försök eller mätvärde. 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.¹²

För att bestämma önskad säkerhet i samband med undersökningar av förorenade områden med hjälp av metoder som bygger på konfidensintervall, kan graden av önskad säkerhet (d) anges som skillnaden mellan den övre konfidensgränsen (UCL) och den verkliga medelhalten (μ):

$$d = UCL - \mu \quad (\text{Ekv. 3-2})$$

Talet d anger hur högt över den verkliga medelhalten som det uppmätta medelvärdet (motsvarande den övre konfidensgränsen, UCL) kan tillåtas hamna, vilket kan jämföras med en form av säkerhetsmarginal. Om exempelvis d sätts till 50 mg/kg så innebär det att det uppmätta medelvärdet tillåts bli upp till 50 mg/kg högre än den verkliga medelhalten i området¹³. I detta fall är den

¹² Konfidensgraden anger med vilken sannolikhet det verkliga medelvärdet ligger inom det beräknade konfidensintervallet. Om det är ett tvåsidigt konfidensintervall så är sannolikheten 0,025 att det verkliga värdet ligger under 27 och 0,025 att det verkliga värdet ligger över 33 (enligt exemplet i texten). Om det anges som ett ensidigt övre konfidensintervall är sannolikheten 0,05 att det verkliga värdet ligger över 33.

¹³ Med det uppmätta medelvärdet avses en väntevärdesriktig skattning av den verkliga medelhalten, ofta angiven med en säkerhetsmarginal uttryckt som UCL_{95} . Den verkliga medelhalten är just den verkliga, men okända, medelhalten i området, dvs. den man försöker mäta. För normalfördelade data är det uppmätta medelvärdet (utan säkerhetsmarginal) lika med det aritmetiska medelvärdet hos stickprovet, men för lognormalfördelade data kan det uppmätta medelvärdet skilja sig från det aritmetiska medelvärdet.

önskade säkerheten d en haltskillnad och det fungerar ofta att ange ett rimligt värde på d om data är normalfördelade. Detta kan dock vara svårt om halterna varierar kraftigt, vilket är fallet när data följer en skev fördelning som t.ex. en lognormalfördelning, vilket är vanligt i förorenade områden. Ett alternativ är då att ange önskad säkerhet (D) som en kvot mellan övre konfidsgräns och verklig medelhalt:

$$D = \frac{UCL}{\mu} \quad (\text{Ekv. 3-3})$$

Kvoten D talar om hur många gånger över den verkliga medelhalten som det uppmätta medelvärdet hamnar. Om exempelvis D sätts till 1,7 så motsvarar det att det uppmätta medelvärdet kan bli upp till 70 % högre än den verkliga medelhalten i området. Om exempelvis medelhalten är 100 mg/kg så accepterar man därmed en övre konfidsgräns (UCL) på 170 mg/kg.

När d eller D skattas behöver man ange vilken konfidsgrad som avses. Det normala är att man använder sig av ett ensidigt 95% konfidsintervall, dvs. en konfidsgrad på 95%. Detta innebär att med 95% sannolikhet kommer den verkliga medelhalten att vara mindre än UCL (ensidigt konfidsintervall). Även andra konfidsgrader kan användas om man så önskar men 95% är en relativt väletablerad nivå som används i många sammanhang. Det bör dock påpekas att 95% inte alltid är en lämplig konfidsnivå. För att beräkna UCL för en viss konfidsgrad finns flera olika programvaror som kan användas. ProUCL är ett användarvänligt och gratis datorprogram som tagits fram av US EPA (2008) specifikt för sådana beräkningar. Även Excel kan användas.

METODER SOM BYGGER PÅ HYPOTESTESTNING

Metoder som bygger på hypotestestning har en något annan utgångspunkt och bygger på teorin om statistisk hypotestestning. I statistisk hypotestestning relateras osäkerheterna till potentiella beslutsfel, dvs. att osäkerheter finns innebär att beslutsfattaren riskerar att fatta ett felaktigt beslut.

En typisk situation där hypotestest kan användas är vid frågeställningar om medelhalten i området överskrider ett bestämt riktvärde. Det potentiella beslutsfelet är då att ett rent område felaktigt klassas som förorenat (benämnt som Typ II-fel eller beta (β)-fel), eller tvärtom, det vill säga att ett förorenat område felaktigt klassas som rent (benämnt som Typ I-fel eller alfa (α)-fel). När statistisk hypotestestning används för att bestämma antalet prover behöver man bestämma sig för hur stora beslutsfel som kan accepteras, både vad gäller α - och β -fel. Vanliga nivåer på risken för α - och β -fel är 0,05 respektive 0,20 (Grandin, 2006). Detta innebär att man accepterar att klassa ett förorenat område som rent i 1 fall av 20 och att i 1 fall av 5 så accepteras ett rent område att klassas som förorenat. Hur stora felrisker som kan accepteras samt vilken effektstorlek som kan krävas beror både på provtagningsbudgeten och på hur stora konsekvenserna blir av att fatta fel beslut.

Hänsyn måste även tas till den så kallade effektstorleken som kan beskrivas som den förändring eller skillnad man vill kunna upptäcka vid en provtagning. 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.

Principerna för hypotestest och hur ett sådant kan formuleras beskrivs utförligt i Bilaga D. Där beskrivs även effektstorlek och hur US EPA brukar illustrera *gray region* i ett exempel med en s.k. *Decision Performance Curve*.

Effektstorlek, minsta detekterbara skillnad eller *gray region*

Effektstorleken kan beskrivas som den förändring eller skillnad man vill kunna upptäcka vid en provtagning. 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.

För att bestämma hur stor effektstorlek man vill kunna detektera så behöver man bestämma hur mycket lägre än riktvärdet som den verkliga medelhalten får vara och att man ändå kan acceptera att området klassas som förorenat.

Riktvärdet för bly är 300 mg/kg för mindre känslig markanvändning. Om acceptabel effektstorlek är 30% så innebär det att den minsta detekterbara skillnaden, eller *gray region*, är 90 mg/kg (betecknas som Δ). Om områdets verkliga (men okända) medelhalt ligger mellan 210 mg/kg och 300 mg/kg så anser man det i detta fall acceptabelt att området klassas som förorenat trots att det egentligen borde klassas som rent.

Om man väljer en motsvarande acceptabel effektstorlek på 10% så blir $\Delta=30$ mg/kg och *gray region* är då intervallet mellan 270 mg/kg och 300 mg/kg.

3.6.4 Välj provtagningsmönster

Det finns flera vanliga provtagningsmönster som kan användas vid sannolikhetsbaserad provtagning. De viktigaste är slumpmässig provtagning, systematisk provtagning samt systematisk slumpmässig provtagning. Fördelar och nackdelar med de olika metoderna beskrivs närmare i Bilaga C. Det viktigaste vid valet av provpunkternas lägen är dessa är slumpmässiga och inte styrda till områden med en viss halt. Det skall alltså inte finnas någon koppling mellan provpunktsplacering och föroreningshalt, annars blir resultaten inte representativa. Även systematiska provtagningsmönster kan betraktas som

slumpmässiga, under förutsättning att den första provpunkten samt mönstrets orientering väljs slumpmässigt.

Ett enkelt sätt att slumpa ut provpunkter är att använda sig av slumpfunktionen i kalkylprogram som t.ex. Excel. Koordinater slumpas fram och punkter som hamnar utanför området eller på svåråtkomliga ställen, t.ex. under en byggnad, kan slumpas ut på nytt. Det finns också funktioner i programvaror som Visual Sample Plan (VSP, 2008) och Spatial Analysis and Decision Assistance (SADA, 2008) för att slumpa ut provpunkter eller att slumpa ut ett systematiskt provtagningsmönster inom ett definierat område.

3.6.5 Bestäm antalet prover

För att bestämma hur många prov som behövs för att uppnå önskad säkerhet måste hänsyn tas till vilken statistisk fördelning som mätdata kan förväntas ha (förutom vilken variabilitet data kan förväntas ha, se avsnitt 3.6.2). Olika metoder finns beroende på om vilken fördelning som är rimlig. Vid jordprovtagning av förorenade områden är sällan föroreningshalterna normalfördelade utan snarare brukar halterna vara skevt fördelade t.ex. lognormalfördelade.

De metoder som presenteras är, för både normalfördelade och lognormalfördelade, data:

- Metod baserad på konfidensintervall
- Metod baserad på hypotestestning
- Metod baserad på hypotestestning och sekventiell provtagning

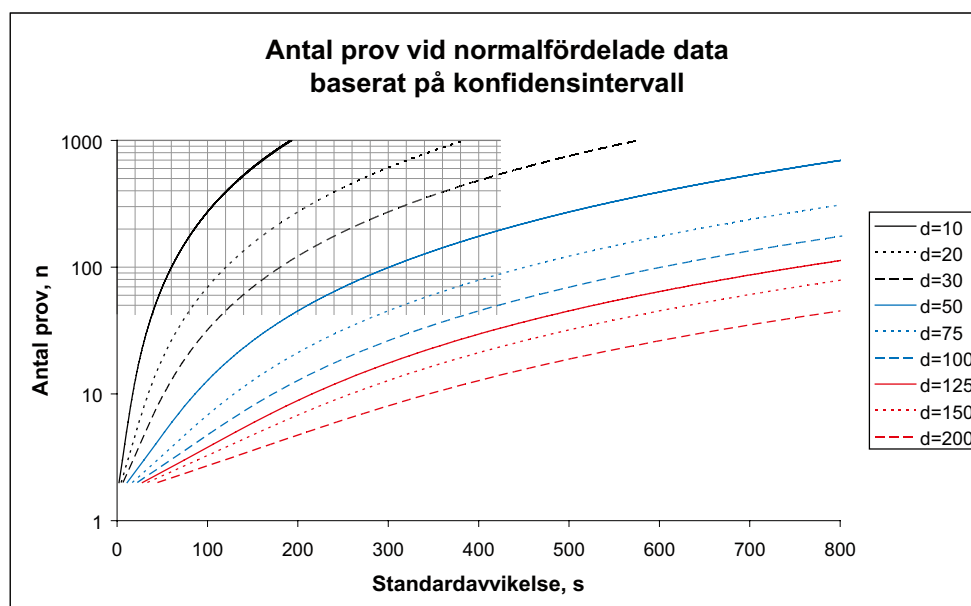
Normalfördelade data: metod baserad på konfidensintervall

Denna metod är den som oftast beskrivs i provtagningslitteratur. Antalet nödvändiga prov (n) beräknas med följande uttryck:

$$n = \left(\frac{t_{n-1, \alpha} \cdot s}{d} \right)^2 \quad (\text{Ekv. 3-4})$$

Talet $t_{n-1, \alpha}$ är ett värde från Student's t-fördelning med $n-1$ frihetsgrader och signifikansnivån α . Sådana t-värden finns tabellerade i alla vanliga statistikböcker och finns även som inbyggd funktion i kalkylprogrammet Excel. Signifikansnivån α anger sannolikheten att skillnaden mellan uppmätt medelhalt och verklig medelhalt faktiskt är större än d . Ett vanligt värde på α är 0,05 (ensidigt 95% konfidensintervall). Beteckningen s står för standardavvikelsen. Ekvation 3-4 måste lösas genom iteration eftersom t-värdet är beroende av antalet prov. I Figur 3-6 illustreras ekvation 3-4 grafiskt. Denna graf kan användas för att snabbt bestämma antalet prover vid ett övre konfidensintervall på 95% utan att beräkningar behöver göras.

Figur 3-6 kan användas även för koncentrationsvärden som inte täcks in av skalorna i figuren. Antag t.ex. att $s = 900$ mg/kg och $d = 300$ mg/kg. Då kan man helt enkelt ta bort en nolla i båda dessa värden. Grafen visar då att antalet prover som krävs är ca 25.



Figur 3-6. Graf för bestämning av antal prov med metod baserad på ett 95% ensidigt konfidensintervall för medelvärdet. Skillnaden mellan övre konfidensgräns (UCL95) och verklig medelhalt betecknas med d. Variablerna s (standardavvikelsen) och d måste ha samma enhet.

NORMALFÖRDELADE DATA: METOD BASERAD PÅ HYPOTESTEST

Den här metoden används av US EPA och finns även implementerad i VSP (2008) och SADA (2008). US EPA (2006b) ger en ekvation för antalet prover (n) baserad på den standardiserade normalfördelningen (z):

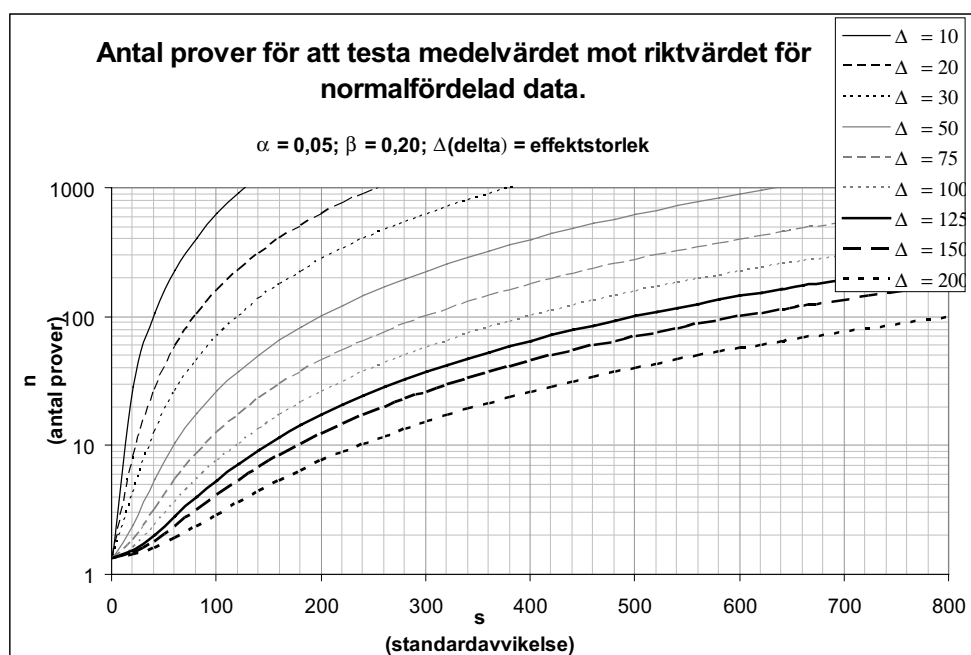
$$n = \frac{s^2(z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2} + \frac{z_{1-\alpha}^2}{2} \quad (\text{Ekv. 3-5})$$

Talet $z_{1-\alpha}$ och $z_{1-\beta}$ är värde från den standardiserade normalfördelningen med signifikansnivån α respektive β . Standardavvikelsen (s) skall anges samt effektstorleken (eng *gray region*) (Δ). Detta finns beskrivet utförligare i Bilaga D. Ekvation 3-5 finns även representerad som en graf i Figur 3-7 där en bestämning av antalet prov som behövs kan göras utifrån standardavvikelse och effektstorlek (Δ), förutsatt att data är normalfördelade.

Figur 3-7 kan användas även för koncentrationvärden som inte täcks in av skalorna i figuren. Antag t.ex. att $s = 900$ mg/kg och $\Delta = 300$ mg/kg. Då kan man helt enkelt ta bort en nolla i båda dessa värden. Grafen visar då att antalet prover som krävs är 55-60.

NORMALFÖRDELADE DATA: HYPOTESTEST OCH SEKVENTIELL PROVTAGNING

Sekventiell hypotestestning med Student's t-test, bygger på att data provtas sekventiellt (i omgångar) från en normalfördelning med okänd eller känd standardavvikelse. I Visual Sample Plan (2008) finns båda varianterna inbyggda, men i den här rapporten har vi endast exemplifierat proceduren med okänd standardavvikelse i Exempelsamlingen (Bilaga A), exempel C. Metoden kallas



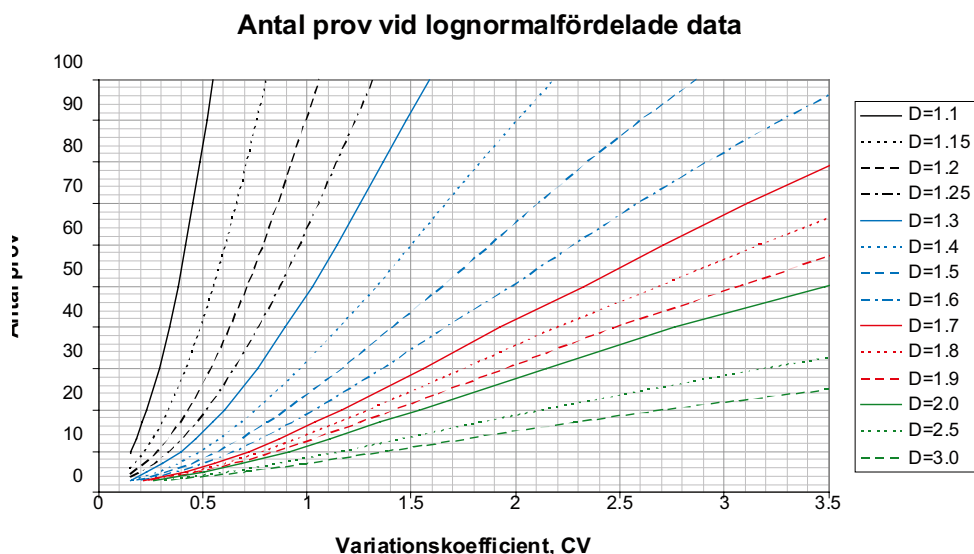
Figur 3-7. Graf för bestämning av antal prov för att bestämma medelvärdet med önskad grad av säkerhet angivet som acceptabla α - och β -fel samt effekttorlek (Δ).

”Barnard’s sequential t-test” och teorin finns beskriven i Appendix A, i Gilbert et al. (2002), som finns tillgänglig på nätet.

Metoden är användbar när man misstänker att den verkliga (men okända) medelhalten i ett område ligger högt över eller långt under riktvärdet. De två föregående metoderna bygger på att man dimensionerar sin provtagning för värsta fallet, dvs. ett fall där den verkliga medelhalten i området ligger nära riktvärdet. *Ju längre ifrån riktvärdet den verkliga medelhalten ligger, desto färre prover behövs för att statistiskt säkerställa att området är förorenat eller rent.* Detta utnyttjas i en sekventiell provtagning. Eftersom man med traditionella metoder skapar flera etableringskostnader om man provtar i flera omgångar är förhandskunskap eller tidigare provtagningar en viktig faktor för att bedöma om det lönar sig med en sekventiell provtagning eller ej.

LOGNORMALFÖRDELDE DATA: METOD BASERAD PÅ KONFIDENSINTERVALL

I Bilaga E härleds en ekvation för att beräkna antalet prover då data är lognormalfördelade, något som är vanligt i förorenade områden. Denna ekvation är förhållandevis komplicerad och därför har en graf tagits fram med hjälp av ekvationen, se Figur 3-8. Grafen kan användas för att snabbt beräkna antalet prov som krävs när data förväntas vara lognormalt fördelade. Innan grafen kan användas måste man bestämma en önskad grad av säkerhet (D) för medelvärdet (se avsnitt 3.6.3) samt en bedömd variabilitet uttryckt som en variationskoefficient (CV), enligt ovan.



Figur 3-8. Graf för bestämning av antal prov för att bestämma medelvärdet med önskad grad av säkerhet för ett 95% ensidigt konfidensintervall (UCL95). D anger hur många gånger högre uppmätt medelvärde får vara än det verkliga medelvärdet. Variationskoefficienten CV är ett mått på föroreningsvariabilitet i området.

Metoden bygger på Lands metod att beräkna konfidensintervall för medelvärdet när data är lognormalfördelade och beskrivs av bl.a. Gilbert (1987). Notera att vid låga CV blir antalet prov som krävs ungefär detsamma som om man hade antagit att data var normalfördelade. Detta beror på att vid låga CV liknar lognormalfördelningen en normalfördelning.

LOGNORMALFÖRDELDE DATA: METOD BASERAD PÅ HYPOTESTEST

Eftersom data från förorenad jord ofta inte är normalfördelad utan oftare följer en lognormalfördelning, utvecklades inom detta projekt en förenklad metod för att skatta antalet prover utifrån krav på α -fel (typ I), β -fel (typ II) och effektstorlek (*gray region*).

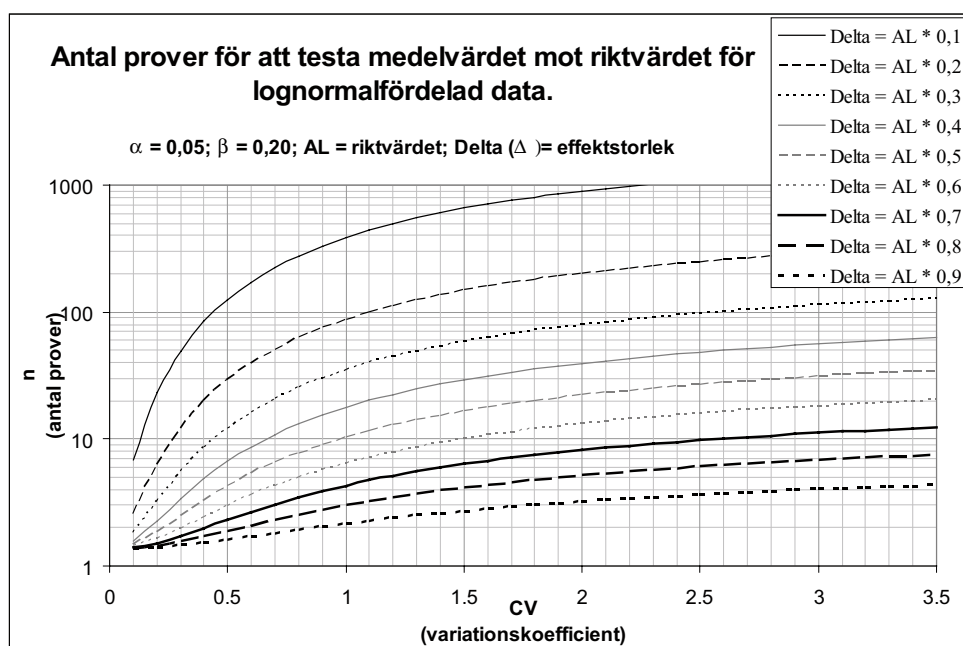
Den förenklade metoden för att beräkna antalet prover utifrån antagande om en lognormalfördelad datamängd har tagits fram av statistikern Landon Sego på Pacific Northwest National Laboratory (PNNL), USA. En utförlig beskrivning av metoden på engelska finns i Bilaga F (avsnitt F.1) tillsammans med en analys av hur insamlad data bör utvärderas (avsnitt F.2 – F.3), en diskussion av metoderna (avsnitt F.4), samt några praktiska rekommendationer (avsnitt F.5). Texten kan upplevas som tung för de med begränsade statistiska förkunskaper.

I korthet baseras metoden på ett antagande om en konstant variationskoefficient (CV) och att parametrarna s (standardavvikelsen) och Δ (effektstorleken) som ingår i ekvation 3-5 beräknas på lognormalskalan enligt ekvationerna F-9 och F-10 i Bilaga F. Ekvation 3-5 används sedan för att beräkna antalet prover som behövs för att uppnå en önskad säkerhet. I Bilaga F rekommenderas *starkt* att en pilotstudie utförs för att få en så bra skattning på standardavvi-

kelsen (s) som möjligt. En annan möjlighet är att skatta variationskoefficienten och använda den för att göra skattningen av standardavvikelsen (s), se avsnitt 3.6.2. Anledningen till att det är viktigt att få en bra skattning av standardavvikelsen är att metoden bygger på antagandet om en konstant variationskoefficient. På lognormalskalan innebär detta ett antagande om att standardavvikelsen är konstant, vilket gör metoden känslig för stora fel i skattningen av denna.

PNNL har även utvecklat gratisprogramvaran Visual Sample Plan (VSP, 2008) och metoden går att använda i VSP, där värdena för standardavvikelse (Estimated Standard Deviation), effektstorlek (Width of Gray Region) samt riktvärde (Action Level) anges på lognormalskalan istället för på normalskalan.

För att slippa beräkningar (som i och för sig inte är besvärliga utan enkelt kan göras i Excel) har ett diagram tagits fram (Figur 3-9) där man kan avläsa antalet prover som krävs givet acceptabla α - och β -fel samt olika effektstorlekar uttryckt som procent av riktvärdet. I Figur 3-9 är α - och β -felen 0,05 respektive 0,20.



Figur 3-9. Graf för bestämning av antal prov för att bestämma medelvärdet med önskad grad av säkerhet angivet som acceptabla α - och β -fel samt effektstorlek (delta, Δ). AL står för engelska Action Level, motsvarande riktvärdet. Effektstorleken (Δ) och riktvärdet (AL) anges i samma enhet, t.ex. mg/kg eller μ g/kg. Variationskoefficienten CV är ett mått på föroreningsvariabilitet i området.

LOGNORMALFÖRDELDE DATA: HYPOTESTEST OCH SEKVENTIELL PROVTAGNING

Den metod som finns beskriven i Bilaga F (se även avsnittet ovan) kan eventuellt användas för att räkna på sekventiell provtagning också, m.h.a. Barnard's sequential t-test som kortfattat beskrivits tidigare. Dock krävs det vidare analys för att verifiera detta (se avsnitt F.4.3) varför inget exempel genomförs här.

Sekventiell provtagning kan vara mycket lönsam även med traditionella metoder när den verkliga medelhalten ligger långt över eller under riktvärdet. Om man misstänker att så är fallet bör man först göra en pilotstudie för att få en första skattning på medelhalten i området. Om det visar sig att den skattade medelhalten ligger långt över eller under riktvärdet kan man, för att minska antalet nödvändiga prover, ha lägre krav på effektstorleken eller β - respektive α -felet.

KOSTNADSASPEKTER

Då de olika beskrivna metoderna tillämpas upptäcker man snart att antalet prov som krävs kan bli mycket stort. Inte sällan skulle en så omfattande provtagning bli orimligt kostsam. I dessa fall bör man överväga om kraven på säkerheten i resultatet ställts alltför högt i förhållande till tillgänglig provtagningsbudget. Genom att sänka kravet på säkerhet kan en rimligare avvägning mot kostnaderna göras, men då ökar även risken för att dra felaktiga slutsatser i och med att osäkerheterna blir större.

3.7 Steg 6b: Antal prov med bedömningsbaserat angreppssätt

Tänk på ...

- Formulera alltid en hypotes baserat på en konceptuell modell och välj provpunkternas lägen utifrån denna. Ta även hänsyn till att data kommer att användas i senare skeden av projektet.
- Bra tumregler saknas för hur många prov som krävs. Kostnader och nytta måste vägas mot varandra från fall till fall.
- Erfarenhet och expertkunskap från andra områden där man både undersökt och åtgärdat bör utnyttjas i planeringen.
- Om möjligt, utgå från det antal prov som skulle krävas vid sannolikhetsbaserad eller sökbaserad provtagning och gör sedan en subjektiv anpassning.
- Tänk på att beräknade medelvärden som baseras på prover riktade mot de mest förorenade delvolymerna blir missvisande (medelvärdet blir högre vilket kan leda till överskattning av riskerna). Detsamma gäller andra statistiska parametrar.

Om ett bedömningsbaserat angreppssätt används bör följande moment ingå:

- Formulera hypotes om föroreningsituationen
- Bestäm typ av prov
- Bestäm antal prov
- Bestäm provtagningspunkternas lägen
- Definiera hur data ska tolkas

3.7.1 Formulera hypotes om föroreningsituationen

Eftersom de bedömningsbaserade angreppssätten baseras på den förhandskunskap som finns är det viktigt att formulera en hypotes om föroreningsituationen. Denna hypotes kan vara enkel men det är ändå viktigt att den uttrycks. Ett exempel på en sådan hypotes kan vara följande: *”Ett cirkelformat område med radien 5 m runt det gamla dopningskaret bedöms vara kraftigt förorenat. Ett motsvarande område med radien 20 m bedöms vara något förorenat.”* Provtagningen kan sedan inriktas mot att verifiera eller avfärda denna hypotes.

En hypotes för ett misstänkt oljeförorenat område kan vara. *”Oljan har trängt ner i marken från den läckande cisternen och sedan spridits horisontellt på relativt tät siltlager samt på grundvattenytan”.*

Givetvis kan hypoteserna vara betydligt mer omfattande än så. Det är bra om de dokumenteras skriftligt. Detta underlättar senare utvärdering och tolkning av data.

3.7.2 Bestäm typ av prov

De vanligaste typerna av prov är enskilda prov samt samlingsprov. Enskilda prov representerar en mindre volym än samlingsprov. Samlingsprov slås samman av ett antal delprov från en större volym.

Vilken typ av prov som är lämpligast beror på syftet med provtagningen och föroreningsens egenskaper. Om exempelvis provresultaten ska användas för bedömning av akuta risker är det bäst att välja mindre provvolym som representerar risksituationen bättre än stora jordvolym. Om riskerna å andra sidan avser långtidsrisker kan samlingsprov över stora jordvolym vara lämpligare och dessutom mer kostnadseffektiva.

3.7.3 Bestäm antal prover

Det finns ingen tydlig metodik för att bestämma lämpligt antal prover då ett bedömningsbaserat angreppssätt används. Detta gäller särskilt när provtagningen gäller avgränsning av en förorenings utbredning. Istället måste antal prov baseras på den hypotes om föroreningsituationen som formulerats, den konceptuella modellen, kostnader samt hur säker man vill vara att inte resultaten ska leda till felaktiga bedömningar om risker och åtgärder.

En möjlig väg för att bestämma antalet prov vid t.ex. riktad provtagning kan vara att utgå från det antal prov som skulle krävas vid sannolikhetsbaserad eller sökbaserad provtagning och sedan göra en subjektiv anpassning av detta. Generellt krävs färre prov när förhandskunskapen är god än när den är svag. Färre prov krävs även när större osäkerhet accepteras, exempelvis i tidiga skeden som MIFO fas 2.

3.7.4 Bestäm provtagningspunkternas lägen

Vid bedömningsbaserad provtagning ska provpunkternas lägen baseras på hypotesen om föroreningsituationen (se ovan) samt den konceptuella modellen. Om syftet är att bekräfta eller förkasta hypotesen så väljer man lämpli-

gen att placera provpunkterna där förhandskunskapen är störst. Om syftet däremot är att avgränsa föroreningen bör man istället placera provpunkterna där osäkerheterna är störst. I tidiga skeden förekommer ofta en kombination av dessa båda syften.

Det finns även några andra aspekter som bör beaktas när provpunkternas lägen bestäms. En sådan är hur data kommer att användas i senare skeden. Många gånger kommer äldre data att blandas med nya data, t.ex. i en huvudstudie, och användas för att statistiskt beskriva området (medelvärden, mängdberäkningar etc.). Därför bör provpunkternas lägen väljas så att provtagningsmönstret blir någorlunda likt ett sannolikhetsbaserat mönster. Detta innebär att det kan vara direkt olämpligt att rikta in samtliga provpunkter mot de delvolymer som har allra högst föroreningshalter. En liknande aspekt bör beaktas om data ska användas för att visualisera föroreningsbilden i form av en interpolerad kartbild. Då bör provpunkter placeras med olika avstånd från varandra, även på mycket korta avstånd. Detta är särskilt viktigt om geostatistiska metoder ska användas, där korrelationsstrukturen¹⁴ undersöks, t.ex. med variogram.

3.7.5 Definiera hur data ska tolkas

Enligt ovan ska provpunkterna placeras med hänsyn till hypotesen som formulerats. För att underlätta kommande utvärdering av analysdata bör man redan före provtagningen definiera hur olika möjliga resultat ska tolkas. Om exempelvis en provpunkt har placerats ut med syfte verifiera hypotesen att området är förorenat så bör man i förväg även tänka igenom hur en låg halt i analysvaret ska tolkas.

3.8 Steg 6c: Antal prov med sökbaserat angreppssätt

Tänk på...

- Det krävs ett stort antal prover för att finna en okänd *hotspot* med hög säkerhet. En enkel tumregel är att dividera sökområdets area med *hotspotens* förmodade area för att få en uppfattning om hur många prover som kan behövas.
- Sökbaserad provtagning är mer effektiv desto mer förhandskunskap som finns. Därför kan geofysiska undersökningar, andra screeningmetoder eller undersökningar i andra medier än jord (t.ex. grundvatten eller porluft) ofta vara bra som ett första steg.

¹⁴ Korrelationsstrukturen beskriver hur föroreningshalter på olika avstånd liknar varandra. Om det föreligger en rumslig korrelation så är typiskt koncentrationerna mer lika i punkter på korta avstånd från varandra än i punkter på stort avstånd.

När ett sökbaserat angreppssätt används för att detektera *hotspots* (eller andra okända objekt) ska följande utföras:

- Ange objektets form och storlek
- Välj modell och provtagningsmönster
- Bestäm önskad säkerhet i resultaten
- Bestäm antalet prov

Metodiken innebär att man på förhand räknar ut hur många prov som krävs för att uppnå en viss önskad sannolikhet att detektera ett objekt av angiven storlek och form. Detta är metoder som bygger på sökteori som ursprungligen utvecklades inom gruvindustrin och bygger på geometriska förutsättningar. Här anges inte ekvationer för att räkna ut antalet prover, istället hänvisas läsaren till en bok (Gilbert, 1987) eller programvaran Visual Sample Plan (2008) som kan laddas ner gratis från nätet

I det sökbaserade angreppssättet kan man välja att ignorera eventuell förhandskunskap om objektet eller att försöka beakta den, se även avsnitt 3.2. Båda metoderna har beskrivits av Gilbert (1987). I Gilberts metod att beakta förhandskunskapen kan man ange hur sannolikt det är att det förekommer en *hotspot* (eller annat objekt) inom området. Denna bedömning kan man sedan uppdatera med informationen som kommer fram vid provtagningen med Bayes' formel.

I Visual Sample Plan (VSP, 2008) kan man specificera att man kan ha felaktigt negativ detektion, dvs. att man inte detekterar en förhöjd halt med den analysteknik man använder, trots att den verkliga halten är förhöjd (s.k. "*false negative error*").

3.8.1 Ange objektets form och storlek

Formen och storleken på det objekt man vill söka efter kommer att påverka hur sannolikt det är att man lyckas hitta objektet. I metodiken som Gilbert (1987) beskriver tänker man sig att objektet har en cirkulär eller elliptisk form (2-dimensionell sökning). Det viktiga är att man anger en realistisk area på objektet man söker, vilket i vissa fall kan vara mycket svårt. Man måste utgå från den historiska inventeringen och göra en rimlig skattning. Är man mycket osäker kan man prova olika storlekar för att få en känsla för hur många prover som krävs för olika förhållanden.

3.8.2 Välj modell och provtagningsmönster

Det sökbaserade angreppssättet kräver att man använder någon typ av modell. Vi kan skilja på två typer av modeller:

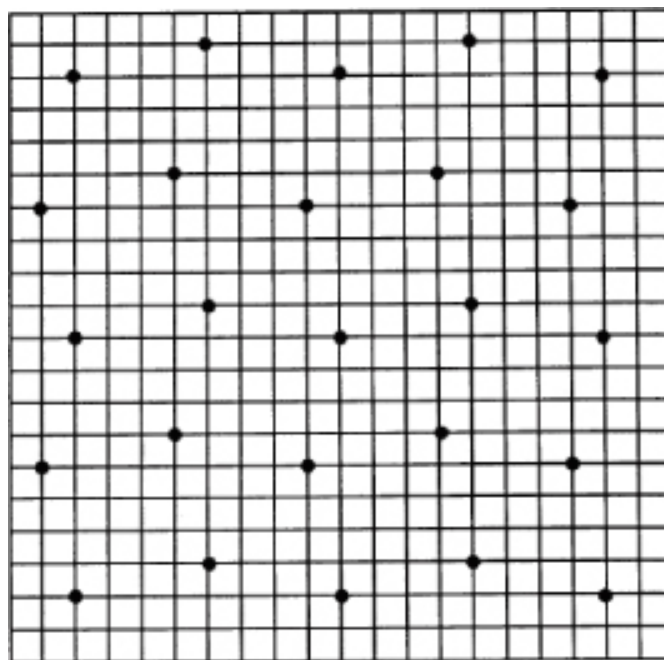
- Typ 1: Modeller som saknar eller ignorerar eventuell förhandskunskap om objektet
- Typ 2: Modeller som beaktar eventuell förhandskunskap om objektet

Modeller av båda typerna har beskrivits av Gilbert (1987) och metodiken där man bortser från eventuell förhandskunskap (typ 1 ovan) finns med i pro-

gramvaran VSP (2008). Fördelen med modeller som ignorerar förhandskunskap är att de är enkla att använda. Den stora nackdelen är att antalet prov som krävs kan bli orimligt stort eftersom ingen hänsyn tas till sådant som faktiskt är känt om det förorenade området.

I Gilberts modell där förhandskunskap beaktas kan man ange hur sannolikt det är att det finns en *hotspot* inom området. Bedömningen kan vara svår att göra, men bör exempelvis baseras på uppgifter om den tidigare verksamheten, platsbesök, erfarenheter från liknande undersökningar o.s.v. Bedömningen kan sedan uppdateras med ny informationen från provtagningen. I övrigt finns det inte många modeller för det sökbaserade angreppssättet som är allmänt tillgängliga och som tar hänsyn till förhandskunskapen (typ 2). Några modeller av typ 2 finns dock beskrivna av Nathanail et al. (1998) och Back (2006).

Oavsett vilken typ av modell som används så måste någon typ av provtagningsmönster användas. Systematiska provtagningsmönster är effektivare än rent slumpmässig provtagning. Ferguson (1992) har jämfört olika provtagningsmönster och kom fram till att det s.k. fiskbensmönstret är mest effektivt, se Figur 3-10. Flera provtagningsmönster beskrivs kortfattat i Bilaga C.



Figur 3-10. Fiskbensmönstret är ett effektivt provtagningsmönster för sökning av hotspots CLR (1994).

3.8.3 Bestäm önskad säkerhet i resultaten

Önskad säkerhet anges som en sannolikhet att träffa en *hotspot* av angiven storlek, givet att en sådan finns. Till exempel kan man kräva att en cirkulär *hotspot* med radien 5 meter skall kunna hittas med sannolikheten 0,95.

Svårigheten att ha en hög säkerhet på att detektera en *hotspot* uppstår givetvis när *hotspoten* är mycket liten i förhållande till sökområdet.

Alternativt kan sannolikheten att missa en sådan *hotspot* istället anges. Gilbert (1987) refererar till detta värde (som han kallar β) som ”*consumer’s risk*”, alltså den risk man som beställare är villig att ta att missa en *hotspot*. Man kan också, både i Gilberts metod (Gilbert, 1987) och i VSP (2008) ange hur stort avstånd man vill ha mellan provpunkterna (baserat på det antal prover som budgeten medger) och istället få hur stor sannolikheten är att detektera en *hotspot* med en viss storlek och form.

3.8.4 Bestäm antalet prov

Antalet prov som krävs beräknas med hjälp av grafer i Gilbert (1987) eller enkelt med Visual Sample Plan (2008) och baseras förhandsinformation gällande *hotspotens* storlek och form, önskad säkerhet och provtagningsmönster. Som nämndes ovan, kan man också specificera antalet prov (eller avstånd mellan provtagningspunkter) och få ut sannolikheten att hitta en *hotspot* av en viss storlek.

4 Alternativa angreppssätt

4.1 Datavärdesanalys

En provtagning kan planeras utifrån olika strategier att hantera osäkerheter och kostnader. Med osäkerheter menas här okunskap om föroreningsituationen, vilket i förlängningen kan leda till olika typer av oförutsedda direkta kostnader och indirekta miljökostnader. De fyra viktigaste strategierna är följande:

- 1) osäkerheterna minimeras för en given budget,
- 2) provtagningskostnaden minimeras utifrån en given accepterad osäkerhet (önskad säkerhet),
- 3) provtagningen följer ett givet regelverk eller myndighetsbeslut där strategi och omfattning fastställts, eller
- 4) provtagningen läggs upp så att provtagningskostnaden balanseras mot den ekonomiska risk som osäkerheterna leder till.

Det första angreppssättet är vanligt i små projekt med liten budget. Angreppssätt 2 är det som rekommenderas i olika provtagningsstandarder. Det är också det angreppssätt som avsnittet om sannolikhetsbaserad provtagning bygger på (se avsnitt 3.5.1). Det tredje angreppssättet innebär att man följer regelverk eller beslut från myndigheter men detta angreppssätt är mindre relevant i denna rapport. Angreppssätt 4 däremot, syftar till att optimera provtagningen så att nyttan av de nya proverna (nya data) uppväger provtagningskostnaden. Denna metod kallas datavärdesanalys, eller informationsvärdeanalys, och beskrivs något utförligare nedan.

Datavärdesanalys innebär att man försöker beräkna det ekonomiska värdet som en provtagning har, dvs. värdet av de nya data som man planerar att samla in. Data innehåller information och det är detta informationsvärde man vill värdera ekonomiskt. För att få ett mått på värdet måste man analysera hur stora osäkerheterna är. Som beskrivs i Bilaga D leder osäkerheter till potentiella felbedömningar och dessa felbedömningar kan leda till olika typer av kostnader. På motsvarande sätt kan korrekta bedömningar leda till nyttor, som exempelvis minskade miljörisker. När nya prover tas så minskar osäkerheterna och därmed ökar den potentiella nyttan (den potentiella kostnaden minskar), vilket även kan uttryckas som den förväntade nyttan (*expected benefit*). Den förväntade nyttan beräknas som sannolikheten att en kostnad eller nytta ska uppstå, multiplicerad med kostnaden/nyttan¹⁵. Notera att kostnader och nyttor kan vara både direkta åtgärds kostnader och indirekta nyttor som hälso- eller miljöförbättringar.

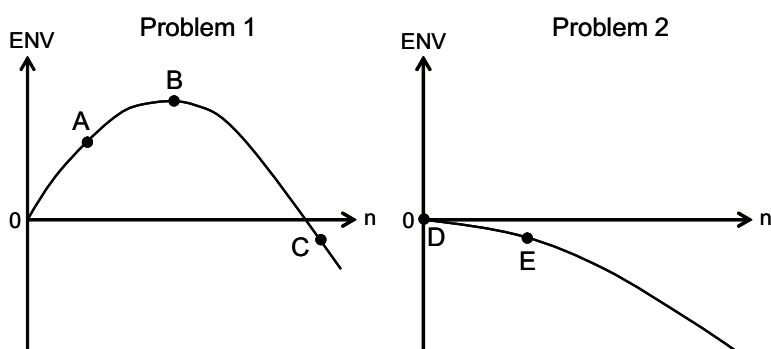
Genom att beräkna den förväntade nyttan för två olika fall, med respektive utan nya data, kan värdet av provtagningen (datavärdet) bestämmas:

¹⁵ Vad som betraktas som en kostnad eller nytta beror på vilken referensnivå som används. Det är lämpligt att använda dagens situation (nollalternativet) som referensnivå.

Datavärde = förväntad nytta med nya data – förväntad nytta utan nya data

Svårigheten i denna beräkning ligger främst i hur den förväntade nyttan med nya data ska kunna beräknas eftersom man ännu inte vet hur data kommer att se ut (analysen görs när provtagningen planeras, alltså innan nya data finns). Det finns dock statistiska metoder som kan användas till detta och som innebär att man kan beräkna hur osäkerheten förändras när man tar ett visst antal prov, trots att man inte på förhand känner till provernas halter.

Datavärdet är ett mått på värdet av den planerade provtagningen. Detta värde måste vägas mot provtagningskostnaden. Endast provtagningsprogram med ett värde som överstiger provtagningskostnaden är kostnadseffektiva och bör genomföras, dvs. nettovärdet måste vara positivt (datavärde minus provtagningskostnad). Det mest optimala provtagningsprogrammet är det som har störst nettovärde. Detta illustreras med provtagningsprogram B i Figur 4-1 nedan. Även program A är kostnadseffektivt, medan program C och E i praktiken innebär ett slöseri med resurser eftersom nettovärdet är negativt.



Figur 4-1. Principen för datavärdesanalys. Nettovärdet, Expected Net Value, på y-axeln och antalet prov på x-axeln. För provtagningsproblem 1 är provtagning B optimal. För problem 2 är provtagning D bäst men datavärdet och provtagningskostnaden är lika stora (skillnaden är noll). Provtagning C och E innebär slöseri med resurser eftersom nettovärdet är negativt.

Datavärdesanalys innebär att graden av säkerhet (jfr. avsnitt 3.6.3) automatiskt väljs till den ekonomiskt optimala nivån. Detta ligger inbyggt i metodiken och man behöver därför inte själv definiera önskad säkerhet, vilket är en stor fördel.

Metodiken kan användas för att optimera provtagningen för väldefinierade provtagningsproblem där syftet är tydligt uttryckt. Beräkningarna kan bli relativt omfattande, vilket gör att en beräkningsmodell krävs för det problem man vill lösa. Exempel på sådana modeller redovisas av Back (2006) för olika provtagningsproblem. Där ges också en utförlig beskrivning av hela metodiken. Nedan ges ett enklare exempel på datavärdesanalys.

Exempel datavärdesanalys

En miljöteknisk markundersökning har genomförts och ett antal jordprover har analyserats. Utvärderingen för ett egenskapsområde (delområde) tyder på att den verkliga medelhalten överskrider riktvärdet. Skattningen av medelhalten är dock ganska osäker och man bedömer att sannolikheten är 75% att egenskapsområdet har en verklig medelhalt som överskrider riktvärdet. Man vill minska osäkerheten och planerar därför ytterligare en provtagning med 10 nya data.

Vid en sanering bedömer man att nyttan i form av minskade hälso- och miljörisiker kan värderas till i storleksordningen 6 Mkr. Själva saneringskostnaden uppgår till 2 Mkr. Den förväntade nyttan med befintliga data blir då:

$$\text{Förväntad nytta utan nya data} = 0,75 \cdot (6-2) + 0,25 \cdot (-2) \text{ Mkr} = 2,5 \text{ Mkr}$$

Hur stor blir den förväntade nyttan med nya data? För att beräkna detta krävs att man tar hänsyn till alla de olika provtagningsresultat som de 10 nya proverna skulle kunna ge. En sådan beräkning kräver därför avancerade sannolikhetsberäkningar (se Back, 2006) eller simuleringar. Detta kan utföras med särskilda datorprogram men i exemplet beskrivs inte detta utan istället antar vi för enkelhetens skull att den förväntade nyttan med nya data blir 2,8 Mkr. Värdet av den kompletterande provtagningen kan nu beräknas:

$$\text{Datavärdet} = 2,8 \text{ Mkr} - 2,5 \text{ Mkr} = 0,3 \text{ Mkr}$$

Kostnaden för att den kompletterande provtagningsomgången bedöms bli 50 000 kr, dvs. 0,05 Mkr. Nettovärdet av provtagningen blir då $(0,3 - 0,05) \text{ Mkr} = 0,25 \text{ Mkr}$. Eftersom nettovärdet är positivt (datavärdet är större än provtagningskostnaden) är det lönsamt att genomföra den kompletterande provtagningen. Om nettovärdet hade blivit negativt hade det inte varit meningsfullt att genomföra provtagningen.

4.2 Dynamisk provtagning, TRIAD

Dynamisk provtagning är en målinriktad provtagning där mätningar utförs och utvärderas i realtid. Definition av mätresultaten utvärderas direkt i fält för att bedöma om fler mätningar behövs och, om så är fallet, var dessa mätningar skall utföras. En dynamisk provtagning kräver ett mycket tydligt provtagningssyfte och en klart definierad beslutsordning. Fördelarna är att svar på en frågeställning, t ex avgränsning av en förorening, erhålls inom ramen för fältarbetet. Efterbehandling kan ske parallellt med provtagningen och omfattningen styrs med ledning av fältresultaten och de i förväg definierade beslutskriterierna.

Konceptet har lanserats som Triad i USA (<http://www.triadcentral.org>) och består av tre huvudsakliga delar:

- 1) systematisk projektplanering,
- 2) ett dynamisk arbetssätt samt
- 3) mätningar i fält som ger resultat i realtid (Triad, 2008).

En förutsättning för korrekta och kostnadseffektiva beslut är att en konceptuell modell (CSM) upprättas för området.

I Triad definieras ett korrekt beslut som det beslut som skulle tas om ett fullständigt korrekt beslutsunderlag rörande föroreningens utbredning och uppförande samt exponering av skyddsobjekt var tillgängligt för beslutsfattarna (svår mening). Målet med Triad är att hantera beslutsosäkerheter, dvs. att öka tillförlitligheten i att beslut i projekt är korrekta och kostnadseffektiva. Det gäller beslut rörande: föroreningars förekomst, lokalisering, spridning, exponering samt val och design av riskreducerande alternativ.



Figur 4-2. Triads logotype. Konceptet består av tre huvudkomponenter: (1) systematisk projektplanering, (2) dynamiska arbetssätt, och (3) mätningar i fält som ger resultat i realtid.

Systematisk projektplanering är den del som är tänkt att stödja Triads mål rörande tryggt beslutsfattande. Triad uppmuntrar utveckling av:

- delaktighet från alla parter i projektet, tillit till varandras kompetens och erfarenheter
- konsensus hos alla parter gällande projektmål,
- en preliminär konceptuell modell
- en lista över tillsynsbeslut, vetenskapliga beslut samt ingenjörbeslut som måste fattas för att nå målet,
- en lista över de (besluts-)osäkerheter som förhindrar att man kan fatta dessa beslut
- strategier för att eliminera, reducera eller kringgå dessa osäkerheter, samt
- proaktiv kontroll över de största osäkerheterna i miljömätdata (dvs. provtagningsrelaterade variabler som support, kornstorlek, provtagningsdensitet).

Systematisk projektplanering är viktigt i alla provtagningar, men speciellt viktigt om en dynamisk provtagningsstrategi utförs. De kvantitativa nyckelfrågorna som skall besvaras måste definieras tillsammans med de beslut som skall kopplas till mätresultaten. Hypoteser kring föroreningens förekomst måste bekräftas och förkastas i realtid, vilket ställer höga krav på den konceptuella modellen och de personer som utför provtagningen. Osäkerheterna i den konceptuella modellen skall även hanteras allteftersom nya resultat erhålls. Vilka osäkerheter som kan accepteras måste vara klarlagt innan provtagningen påbörjas.

Ett dynamiskt arbetssätt är den del som skall tillåta att projekt kan slutföras snabbare och billigare än vid ett traditionellt arbetssätt. Detta innebär att arbetsplaner skall skrivas i en dynamisk eller flexibel form som hjälper projektet att anpassas i realtid (dvs.. när man fortfarande är i fält) när ny information blir tillgänglig. Detta skall i sin tur tillåta att den konceptuella modellen testas och utvecklas fullt ut tills den kan stödja önskad beslutssäkerhet i realtid. Tid och pengar sparas genom bättre kontroll av osäkerheterna för beslutet. Här är det tänkt att en processbaserad kvalitetskontroll är tänkt att bättra stödja analytisk kvalitetskontroll.

Den tredje delen, mätningar i fält som ger resultat i realtid, skall stödja det dynamiska arbetssättet genom att samla in, tolka, och presentera data tillräckligt snabbt för att kunna stödja beslut i realtid. Teknologi som stödjer realtidsmätningar är t.ex.:

- fältinstrument,
- in-situ övervakningssystem,
- geofysiska metoder,
- snabbare mätning/analyser i traditionella laboratorier såsom mobila lab,
- programvara som kan underlätta projektplanering och lagra, presentera, kartlägga, manipulera och dela data.

Det finns flera olika typer av fältinstrument som ger resultat i realtid och som kan utnyttjas vid dynamiska provtagningar. Kvalitetssäkring av mätningarna är viktigt för att kunna hantera osäkerheterna. Utvecklingen av fältinstrument har tagit stora steg de senaste åren, men det finns fortfarande föroreningar som är svåra att mäta i fält (Wadstein et al., 2008).

Provtagningsstätheten ökas för att hantera osäkerheter, vilket ger relativt precisa och detaljerade konceptuella modeller (CSM). Filosofin är att kostnaden per prov är mindre viktig än de realtidsbeslut som kan sänka totalkostnaden för projektet. Detta relaterar till det som skrevs i avsnitt 2.2.6 angående hur osäkerheter adderas.

En viktig del i Triadmetodikerna är att det finns enighet bland alla parter (beställare, utförare och tillsynsmyndighet) gällande projektmål och beslut som skall fattas i fält (med tolerabla beslutsfel) *innan* fältarbetet påbörjas. Den konceptuella modellen bör ta hänsyn till heterogeniteter och föroreningars fördelningar och en strategi för att förfinna modellen under projektets gång ska finnas. Dessutom ska även strategier för att hantera provtagnings- och analysosäkerheter samt kvalitetskontroll finnas framme innan fältarbetet startar .

Arbetet genomförs av multidisciplinära tekniska team med kunskaper inom skilda ämnesområden som t.ex. analytisk kemi, hydrogeologi, provtagningsstrategi och fältmätningar. Dessutom krävs flexibla sätt att skriva kontrakt med uppdragsgivare, vilket kan vara svårt att handla upp genom offentliga upphandlingar. Det krävs vidare någon form av beslutsstöd för beslutsfattande i realtid och inte minst är det ett krav med hög grad av medverkan från beställare och myndighet för att kunna driva ett lyckat projekt.

En dynamisk provtagning innebär även en del administrativa svårigheter jämfört med konventionell provtagning. Exempelvis är det svårt att i

förväg bestämma omfattning och kostnader, då detta styrs av den okända föroreningsituationen samt de beslutskriterier som förutbestämts. Beslut om omfattningen och när en provtagning kan avbrytas sker av personalen på plats och det finns inte alltid möjlighet att invänta utlåtanden från t ex tillsynsmyndighet eller beställare. Beslutsordningen och beslutskriterierna måste därför vara förutbestämda. En dynamisk provtagning ställer höga kompetenskrav på projektgruppen som utför provtagningen samt ett stort förtroende mellan utförare, beställare och myndighet. Den totala kostnaden för en dynamisk provtagning är svår att budgetera, t ex vid ett anbudsförfarande, men i många fall är förmodligen ett dynamiskt angreppssätt det mest kostnadseffektiva förfarandet.

Denna strategi har vuxit fram i USA efter en period med konventionella strategier som lett till oerhört långa och kostsamma undersökningar. Idag är bidragssystemet i Sverige inte riktigt anpassat till att arbeta på det här sättet men förmodligen kommer detta angreppssätt att bli mer attraktivt i en mognare marknad, det vill säga, när betydligt fler förorenade områden har undersökts och åtgärdats och man kan göra säkrare bedömningar av totalkostnaderna. Detta skulle dock kräva att utredningarna upphandlades på ett annat sätt än idag.

Nedan ges några exempel på specifika provtagningsstrategier som passar ihop med ett dynamiskt arbetssätt, där man arbetar med mätningar i realtid för att styra provtagningen.

4.2.1 Kombinerade dataset

Kombinerade dataset består av dels en stor mängd prover analyserade med snabb och billig analysteknik, dels av en mindre mängd prover analyserade med mer precisa och oftast långsammare metoder. Nedan anges ett antal principer för hur kombinerade dataset kan utnyttjas.

Det vanligaste sättet att hantera kombinerade dataset är att göra en regressionsanalys för att analysera hur väl data från de olika analysmetoderna stämmer överens. För att kunna göra detta behöver man ha ett antal som är analyserade på samma prov och inte mindre än 10. Sedan kombinerar man sin data för att beräkna medelhalt osv.

En annan vanligt metod är att låta realtidsmätningar (t. ex. PID eller XRF-mätning) styra vilka prover som ska gå vidare för noggrannare analys genom att skapa ett beslutsschema för vilken halt/nivå provet är med stor säkerhet kan klassas som rent och vid vilken halt/nivå som provet är förorenat. Alla prover som hamnar däremellan skickar man vidare för noggrannare analys.

Co-kriging, är en interpolationsmetod som kan kombinera data med olika ursprung t ex geofysiska data med haltanalyser. Metoden bygger på att man skapar en rumslig bild över en sorts data som är dyr att analysera med hjälp av att den parametern korrelerar med en annan typ av data som är billigare att analysera, dvs. man utnyttjar att datan samvarierar. Co-kriging har bland annat implementerats i den kommande versionen av SADA (2008), version 5.0.

Geofysik är en annan form av data som inte är helt kvantitativ men som kan användas t.ex. för att styra annan provtagning.

4.2.2 Sekventiell provtagning

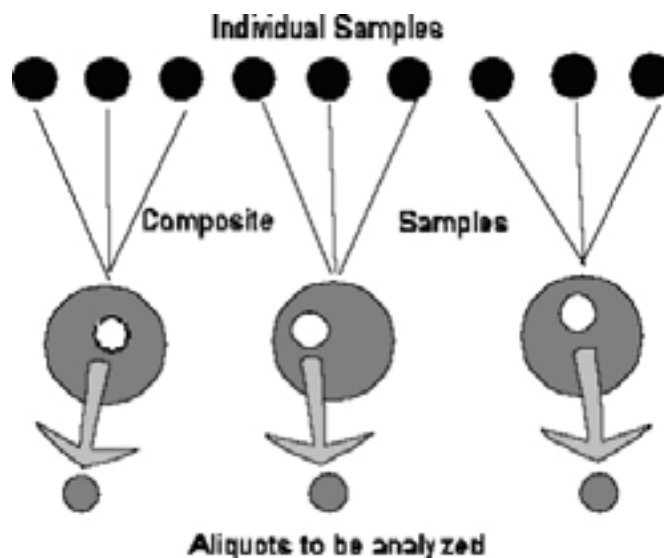
Syftet med en sekventiell provtagning är att avgöra om medelhalten i ett område ligger över eller under riktvärdet genom hypotestestning.

Metoden bygger kortfattat på att genom stegvisa realtidsmätningar i fält på enskilda prover värdera när tillräckligt med prover har mätts för att uppnå önskad säkerhet i resultatet. När den önskade säkerheten uppnåtts avbryts provtagningen. Proceduren stöds bland annat av Visual Sample Plan (VSP, 2008).

Innan sekventiell provtagning påbörjas behöver man specificera acceptabla felrisker för både alfa- och beta-fel samt bredden på effektstorleken (*gray region*). Provpunkter kan baseras på helt slumpmässig provtagning eller kombineras med någon annan typ av provtagningsregel, typ *adaptive fill*, som stöds av SADA (2008).

4.2.3 Anpassad samlingsprovtagning

Samplingsprover innebär att man kombinerar och homogeniserar två eller flera enskilda prover för att ta fram ett nytt sammansatt samlingsprov. Därefter väljs en mindre mängd av det nya provet och analyseras (det delprov som analyseras kallas på engelska *aliquot*). Samlingsprover förväntas representera den större jordvolym som de enskilda proverna hämtats från och på det sättet kan antalet analyser reduceras kostnaden minskas.



Figur 4-3. Huvudprincip för samlingsprovtagning: totalt nio enskilda prover (individual samples) tas i fält varefter blandas de och homogeniseras i 3 separata samlingsprov (composite samples). Därefter tar man ett delprov (aliquot) från varje samlingsprov för labanalyser (US EPA, 2002).

En anpassad samlingsprovtagning med syftet att identifiera *hotspot*områden har beskrivits av US EPA (2008). Först bestäms ett lämpligt antal enskilda prover som skall homogeniseras till ett samlingsprov och därefter beräknas ett beslutskriterium enligt:

$$Decision_criteria = \frac{(action_level - background)}{no_of_samples_in_composite} + background, \text{ (Ekv. 4-1)}$$

där *action level* är riktvärdet och *background* är bakgrundshalten. Om resultatet från samlingsprovet är lägre än beslutskriteriet så accepteras detta, men om halten är högre analyseras varje enskilt prov för sig. Två viktiga frågeställningar är: (a) avståndet mellan de enskilda proverna och (b) hur många enskilda prover som ingår i samlingsprovet. Avståndet är en funktion av hur man tror att *hotspot*området ser ut. Antalet enskilda prover i varje samlingsprov är en funktion av sannolikheten att föroreningen ska finnas och ju större chans att påträffa föroreningen, desto färre enskilda prover bör ingå i samlingsprovet.

Metoden fungerar bra om föroreningen uppträder fläckvis, ju mer fläckvis desto bättre, samt då det är stor skillnad mellan bakgrundshalt och riktvärde (ju större skillnad desto bättre). Vidare är det bättre ju större skillnad det är mellan riktvärdet och medelhalten över hela området. I bästa fallet behöver man inte analysera några enskilda prover och i ett värsta fall behöver man analysera alla enskilda prover i samtliga samlingsprov. Metoden är endast lämplig då provtagningskostnader och hanteringskostnader är betydligt lägre än de analytiska kostnaderna, vilket de dock sällan är. Vidare krävs att analysmetoden är så känslig att den kan användas för att detektera den relativt låga påverkan på medelhalten ett enskilt prov med halt över riktvärdet kan ge.

4.2.4 Rankbaserad provtagning

Rankbaserad provtagning är en innovativ metod som på senare år har tillämpats inom olika miljörelaterade undersökningar. Metoden utvecklades i sin ursprungliga form av McIntyre (1952). Principen går ut på att en slumpmässig provtagning kombineras med bedömningar, eventuellt utifrån resultat från screeningundersökningar, för att bestämma var nya provpunkter skall placeras. Utifrån screeningmätningar kan platser med låga, medelhöga och höga föroreningshalter lokaliseras (skalan kan delas i fler än tre klasser). Man strävar efter att antalet prover för laboratorieanalys ska vara samma för varje rank/klass.

Rankbaserad provtagning ökar sannolikheten att data från valda provpunkter återger hela variationsspänn av föroreningshalter inom det undersökta området. Med andra ord, det är mer sannolikt att data blir mer representativa jämfört med enkel slumpmässig provtagning. Dessutom är rankbaserad provtagning i många situationer mer kostnadseffektiv, eftersom färre antal prover behöver analyseras på labbet än vid slumpmässig provtagning. US EPA (2002) rekommenderar följande skanningsmetoder för datarankning:

- Fluorescensmetoder för BTEX och PAH.
- XRF för metaller.
- TOX (*total organic halid*) för flyktiga klorerade organiska lösningsmedel.

Se även ett exempel som finns beskrivet i NV (2008). Visual Sample Plan (VSP, 2008) stödjer rankbaserad provtagning ("ranked set sampling") med syftet att beräkna ett medelvärde, både baserat på professionella bedömningar och baserat på skanningmetoder i fält.

4.2.5 Anpassad klusterprovtagning

Anpassad klusterprovtagning (eng. *adaptive cluster sampling*) tillämpas då syftet är att bestämma en medelhalt i ett område samt att avgränsa en förorening. För att anpassad klusterprovtagning skall vara lämpligt så bör man förvänta sig *hotspots* inom undersökningsområdet. Klusterprovtagningen styrs av observationer gjorda direkt i fält framförallt med olika screeningmetoder. Konceptet bygger på två huvudprinciper:

- 1) Val av initiala provenheter. Det kan t ex vara en ruta av en viss storlek inom området där man misstänker koncentrationer över riktvärde eller slumpmässigt valda provpunkter.
- 2) Provtagning av intilliggande enheter; provtagningen i intilliggande rutor fortsätter så länge de uppmätta halterna inte uppfyller de fördefinierade kriterierna.

Anpassad klusterprovtagning är fördelaktigt när halter kan mätas snabbt på plats med fältanpassade instrument, t ex med en lämplig screeningmetod. Det finns stöd för utföra en klusterprovtagning i Visual Sample Plan (VSP, 2008), där man dock antar att data är normalfördelade. Metoden finns utförligare beskriven med ett illustrerande exempel i NV (2008).

4.2.6 Geobayesiansk metod

När syftet är att hitta *hotspots* och avgränsa dessa kan en geobayesiansk metod användas. Metoden bygger på att det finns ändamålsenliga instrument för realtidsmätningar, att provresultaten tydligt kan klassas som ja/nej (dvs. över eller under riktvärdet), att det föreligger en signifikant autokorrelation i området samt att man gärna vill använda kombinerade dataset. Metoden är olämplig att använda om föroreningen förväntas uppträda mycket fläckvis, dvs. i små osammanhängande enheter med liten rumslig korrelation, t.ex. utfyllnadsområden.

Metoden använder geostatistik och Bayesiansk analys av kombinerade dataset för att utforma provtagningsstrategin och för att uppskatta sannolikheten för förorening på en specifik plats. En programvara för att utföra en provtagning enligt Geobayesiansk strategi har utvecklats av Argonne National Laboratory (www.evs.anl.gov). Programvaran, *Bayesian Approaches for Adaptive Spatial Sampling* (Johnson et al., 2005) finns tillgänglig på nätet (<ftp://ftp.ead.anl.gov/pub/BAASSVersion1-0/BAASS-Setup.exe>).

Mycket kortfattat bygger metoden på att området delas in i rutor och vid varje nod i rutnätet så uppskattar man en sannolikhet för att platsen är förorenad, baserad på tillgänglig kunskap. Därefter översätts de skattade sannolikheterna till en betafunktion. Man behöver specificera dels acceptabla risker

för beslutsfel samt även den rumsliga korrelationens utsträckning (eng *range*), dvs. det avstånd över vilket det finns ett rumsligt samband.

4.3 Beviskedjor - Lines of evidence

Att använda sig av beviskedjor i provtagningen, eller lines-of-evidence som det även kallas, är inte någon renodlad provtagningsstrategi för jord. Tvärtom, själva tanken är att man samlar in olika typer av data från flera olika medier. Man kan tänka sig två principiella syften med att undersöka en beviskedja: (1) att belägga att förorening finns, eller (2) att belägga att det föreligger en miljö- eller hälsorisk.

Provtagning med det första syftet rör att ta prover i flera media för att utnyttja det faktum att föroreningar typiskt sprider sig till flera olika media (jord, porvatten, porluft, grundvatten, ytvatten, sediment och ev. växter och träd). Genom att undersöka fler media får man en säkrare indikation på att det verkligen finns förorening, trots att man kanske inte hittar någon förorening i jord.

Provtagning med syfte att belägga miljö- eller hälsorisker innefattar prover på olika nivåer i en orsakskedja, från förekomst av en förorening till effekt på det aktuella skyddsobjektet. Här är alltså tanken att man kan få en bättre riskbedömning genom att undersöka olika nivåer i orsakskedjan (förorenad matris, biotillgänglig fraktion, upptag, biokemisk/fysiologisk effekt (individ), effekt på population eller effekt på ekosystem o.s.v). Detta finns skrivet om i remissversionen av den nya vägledningen för riskbedömning från Naturvårdsverket (NV, 2007).

De ingående delarna av beviskedjan har ofta olika stort värde för det principiella undersökningssyftet. Exempelvis så kan arkivstudier som visar att sjunkande vätskor (DNAPL se 3.2.2) har använts i betydande omfattning för ett objekt, ha ett större värde för att belägga att förorening finns i marken, jämfört med ett fåtal rena jordprover. Genom att tillskriva olika vikt/värde/poäng till de ingående delarna (*weight of evidence*) i beviskedjan kan undersökningresultaten vägas samman för att få en slutlig bedömning, vilket möjliggör någon form av kvantifiering och jämförelse av resultaten.

5 Diskussion och rekommendationer

I denna rapport har ett strukturerat angreppssätt beskrivits för att utforma provtagningsstrategier för förorenad jord. Metodiken kan tillämpas på olika typer av problem, dvs. för provtagningar med olika typer av syften: belägga förekomst av förorening, ta fram beskrivande statistik för ett område eller en jordvolym, avgränsning av *hotspot* eller *hotspotsökning*. Metodiken som föreslås för att utforma en bra provtagningsstrategi innehåller följande sex generella moment:

- 1) Definiera provtagningens syfte
- 2) Bestäm hur förhandskunskap ska hanteras
- 3) Definiera rumsliga och tidsmässiga avgränsningar
- 4) Bestäm provtagningskala och typ av prov
- 5) Välj angreppssätt
- 6) Bestäm antal prov och placering.

De tre olika angreppssätten som beskrivs är:

- sannolikhetsbaserat angreppssätt,
- bedömningsbaserat angreppssätt och
- sökbaserat angreppssätt,

där ett sannolikhetsbaserat och ett sökbaserat angreppssätt tillåter att man ställer krav på säkerhet i resultaten för att beräkna antalet prov.

Det som framförallt är nytt för ett sannolikhetsbaserat angreppssätt jämfört med tidigare rapporter från Naturvårdsverket är att det ges praktiska metoder för att beräkna antalet prover som behövs under vissa krav på säkerhet i resultaten, både för normal- och lognormalfördelad data. Graferna som har introducerats i rapporten kan användas dels för att dimensionera antalet prover utifrån önskade krav på säkerhet. De kan även användas för att avläsa vilken säkerhet man kan förvänta sig med ett givet antal prover, dvs. i fall där provtagningsbudgeten eller provtagningens omfattning är låst. För det bedömningsbaserade angreppssättet betonas vikten av att sätta upp en hypotes för hur föroreningsituationen ser ut samt att redan före provtagningen definiera hur data bör tolkas beroende på analysvar. För det sökbaserade angreppssättet betonas vikten av att använda sig av kompletterande metoder, och eventuell förhandskunskap, för att antalet prover skall bli rimligt.

Metodiken skall inte ses som en enkel kedja av steg som man successivt arbetar sig igenom utan ett iterativt arbetssätt rekommenderas där man går tillbaka en eller flera gånger och ser över och eventuellt reviderar de antaganden och skattningar som gjorts. Olika syften och angreppssätt illustreras i exempelbilagan där beskrivningarna följer den stegvisa beskrivning som ges i huvudtexten (Kapitel 3). Flera olika metoder används ibland för samma problem för att visa att det inte bara finns EN lösning som är den rätta.

Exempelsamlingen kan vara ett stöd när man vill tillämpa metodiken på ett verkligt problem.

Metodiken riktar sig främst mot ”traditionella efterbehandlingsprojekt”, dvs. projekt där man planerar sin provtagning på kontoret, etablerar sig i fält och samlar in data, för att därefter återvända till kontoret för att utvärdera och analyserad data. Detta är fortfarande det vanligaste sättet att arbeta, men några alternativa angreppssätt beskrivs även. Syftet med detta är att rapporten även ska introducera alternativa sätt som dels bygger på kostnads-nyttoaspekter, dels på ett mer dynamiskt arbetssätt, arbetssätt som kanske kommer att vara vanligare i framtiden.

För vissa problemställningar kan beräkningar behövas göras där de framtagna graferna inte är tillräckliga. Visual Sample Plan (VSP, 2008) är en användarvänlig programvara som är lätt att lära sig för de vanligaste tillämpningarna. Dock bör man se upp med att använda default-värdena för t.ex. standardavvikelse som kommer upp i dialogrutorna eftersom dessa är lågt satta. Man bör alltid göra en platsspecifik uppskattning av variabiliteten, även om den baseras på erfarenheter från liknande områden. Kalkylprogrammet Excel är också användbart och något de flesta PC-användare har tillgång till.

Parallellt med arbetet med denna rapport har arbete pågått med en rapport som beskriver olika statistiska metoder för utvärdering av data och ett ramverk för detta (Norrman et al., 2009).

Den föreslagna metodiken syftar till att vara ett stöd till alla som arbetar med planering av undersökningar av förorenad jord. Specifikt kan rapporten vara ett stöd för konsulter för att kommunicera önskad säkerhet i resultat med beställare och tillsynsmyndigheter, och den koppling detta har till provtagningens omfattning.

6 Referenser

- Back, P.-E., 2003. On Uncertainty and Data Worth in Decision Analysis for Contaminated Land. Thesis for the degree of licentiate of engineering. Institutionen för geologi, Chalmers Tekniska Högskola, Göteborg
- CLR, 1994: Sampling strategies for contaminated land. Report by The Centre for Research into the Built Environment, Nottingham Trent University. DoE, 1994. http://www.eugris.info/envdocs/CLR04_00.pdf
- Domburg, P., de Gruijter, J.J. and Brus, D.J. 1994. A structural approach to designing soil survey schemes with prediction of sampling error from variograms. *Geoderma*, 62(1994):151-164
- Engelke, F., Norrman, J., Starzec, P., Andersen, L., Grøn, C., Overgaard, J. & Refsgaard, A., 2009. Inventering av provtagningsstrategier för jord, grundvatten och porgas. Rapport 5894. Naturvårdsverket, Stockholm
- Ferguson, C.C., 1992. The statistical basis for spatial sampling of contaminated land. *Ground Engineering*, June
- Gilbert, R. 1987. Statistical methods for environmental pollution monitoring. John Wiley & Sons, Inc. New York.
- Gilbert, R.O., Wilson, J.E., O'Brien, R.F., Carlson, D.K., Bates, D.J., Pulsipher, B.A. and McKinstry, C.A., 2002. Version 2.0 Visual Sample Plan (VSP): Models and Code Verification PNNL-13991, Pacific Northwest National Laboratory, Richland, Washington. <http://vsp.pnl.gov/docs/pnnl13991.pdf> (08-06-29)
- Grandin, U., 2006. Statistisk analys av möjligheter att kunna upptäcka regionala trender i de nuvarande programmen för nationell övervakning av sötvatten. Ett underlag för revisionen av programmen år 2006. Rapport Sveriges lantbruksuniversitet, Miljöanalys vol 18. http://publikationer.slu.se/Filer/IMA2006_18.pdf (2008-05-26)
- Johnson, R., LePoire, D., Huttenga, A. and Quinn, J., 2005. Bayesian Approaches to Adaptive Spatial Sampling: An Example Application. Report ANL/EAD/TM/05-1. Environmental Assessment Division, Argonne National Laboratory. <http://www.ipd.anl.gov/anlpubs/2005/05/53231.pdf> (08-07-01)
- McIntyre, G. A., 1952. A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research* 3: 385-390
- Nathanail, C.P., Ferguson, C.C. and Tucker, P. 1998. The use of Bayesian statistics and prior information to design sampling strategies. In *Eight International Congress, International Association for Engineering Geology and the Environment*. More D. and Hungr O. (Editors), Balkema, Rotterdam: Vancouver; 369-376

- 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>)
- Norrman, J., Purucker, T., Engelke, F., Back, P-E., och Stewart, R., 2009. Ramverk för statistisk utvärdering av miljötekniska undersökningar i jord. Rapport 5932. Naturvårdsverket, Stockholm. (Under publicering)
- NV, 1994. Vägledning för miljötekniska markundersökningar. Del I Strategi. Rapport 4310. Naturvårdsverket, Stockholm. <http://www.naturvardsverket.se/Documents/publikationer/620-4310-2.pdf> (08-06-26)
- NV, 1996. Rätt datakvalitet. Rapport 4667. Naturvårdsverket, Stockholm. <http://www.naturvardsverket.se/Documents/publikationer/620-4667-5.pdf> (08-06-29)
- NV, 1997. Åtgärdskrav vid efterbehandling. Rapport 4807. Naturvårdsverket, Stockholm. <http://www.naturvardsverket.se/Documents/publikationer/620-4807-4.pdf> (08-06-26)
- NV, 2007. Riskbedömning av förorenade områden. En vägledning från förenklad till fördjupad riskbedömning. Remissversion 2007-10-19. Naturvårdsverket, Stockholm. http://www.naturvardsverket.se/upload/30_global_meny/02_aktuellt/Remisser/vagledningsmaterial_om_foroerade_omraden/Riskbedomning_av_foroerade_omraden_remissversion_2007-10-19.pdf (08-06-29)
- NV, 2008. Efterbehandling av förorenade områden: Manual för användning och hantering av bidrag till efterbehandling och sanering. Kvalitetsmanual - Utgåva 4. Naturvårdsverket, Stockholm. <http://www.naturvardsverket.se/Documents/publikationer/1234-7.pdf> (08-07-01)
- SADA, 2008. Spatial Analysis and Decision Assistance. Software. University of Tennessee, UT. <http://www.tiem.utk.edu/~sada/index.shtml> (08-06-29)
- SGF, 2004. Fälthandbok, Miljötekniska Markundersökningar, SGF Rapport 1:2004
- Starks, TH., 1986. Determination of Support in Soil Sampling. *Mathematical Geology*, 18(6): 529-537
- Triad, 2008. Triad Resource Center. Federal/State Interagency Partnership, USA. <http://www.triadcentral.org> (08-06-29)
- UNEP (United Nations Environment Programme), 2006. Groundwater and its susceptibility to degradation: A global assessment of the problem and options for management. http://www.unep.org/DEWA/water/groundwater/groundwater_report.asp (2008-07-11)

US EPA, 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)

US EPA, 2006a. Data Quality Assessment: A Reviewer's Guide. EPA QA/G-9R. EPA/240/B-06/002, February 2006, Office of Environmental Information, US EPA, Washington DC. <http://www.epa.gov/quality/qs-docs/g9r-final.pdf> (08-06-26)

US EPA, 2006b. Guidance on Systematic Planning Using the Data Quality Objective Process. EPA QA/G-4. EPA/240/B-06/001, February 2006, Office of Environmental Information, US EPA, Washington DC. <http://www.epa.gov/quality/qs-docs/g4-final.pdf> (08-06-26)

US EPA, 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)

VSP, 2008. Visual Sample Plan. Software. Pacific Northwest National Laboratory, PNNL. <http://vsp.pnl.gov/> (2008-05-26)

Wadstein, E., Axelström, K., Larsson, L., Nilsson Påledal, S. och Stark, M., 2008. Bättre markundersökningar. Delprojekt 1. In situ-metoder för undersökningar av förorenad mark. Statens geotekniska institut, SGI. Varia 591. Linköping. <http://www.swedgeo.se/upload/publikationer/Varia/pdf/SGI-V591.pdf> (09-01-13)

Bilagor

A Exempelsamling

A.1 Fallstudie med förslag till metodik

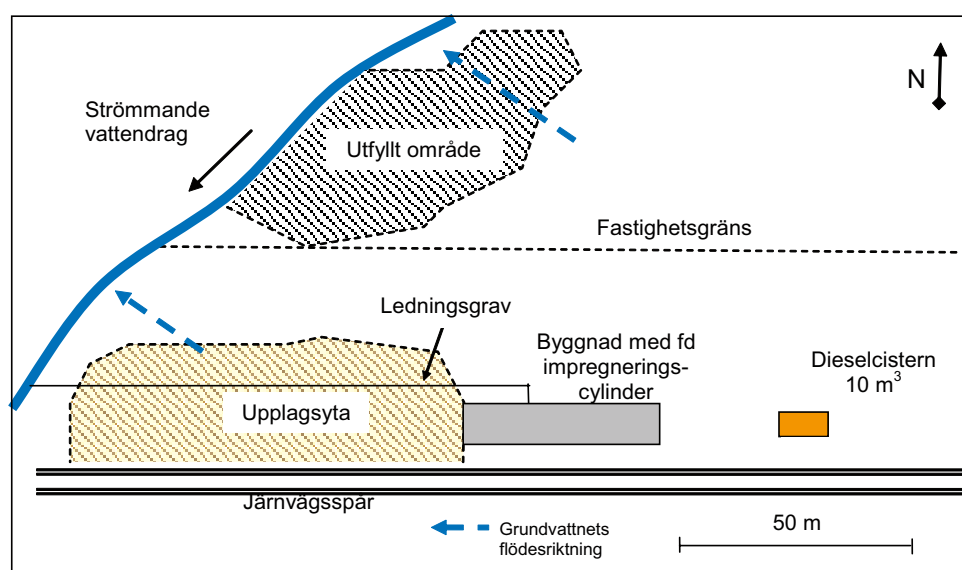
I föreliggande exempelsamling används ett och samma hypotetiska område för alla exempel. I den första delen beskrivs området och de ingående delområdena. Därefter beskrivs några olika frågeställningar och hur lämpliga provtagningsstrategier kan tas fram för respektive delområde.

A.2 Områdesbeskrivning

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 har enligt uppgift alltid bedrivits 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 A-1. Grundvattnenytan 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 okänt.



Figur A-1. Plankarta över området för fallstudien.

A.3 Delområden

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

- 1) Diesalcisternen
- 2) Impregneringsanläggningen,
- 3) Upplagsytan,
- 4) Eventuellt nedgrävt impregneringssalt
- 5) Utfyllnadsområdet

Delområdena har olika karaktär och kommer förmodligen, beroende på syftet, att kräva olika provtagningsstrategier. Det bör även påpekas att exemplen endast behandlar provtagning i jord och inte provtagning av grundvatten, dagvattenledningar, porgas, ytvatten eller liknande som kan vara intressant för att bedöma den totala föroreningsituationen.

I Tabell A-1 sammanfattas föroreningsituationen med delområden, typ av förorening samt en hypotes kring föroreningsituationen.

Tabell A-1. Sammanfattning av föroreningskällor, spridning och skyddsobjekt.

Delområde	Typ av förorening	Hypotes kring föroreningsituationen
Diesalcisternen	Diesel, brännolja Känd punktkälla	Punktkälla med föroreningsplym i grundvatten. Om spill/läckage har skett bör föroreningen kunna återfinnas i samtliga matriser (jord/porgas/grundvatten) i närheten av cisternen.
Impregneringsanläggningen	Koppar, krom, arsenik. Känd men heterogen källa över hela ytan.	Risk finns att förorening kan påträffas på olika djup. CCA-medel bör finnas i de övre jordlagren, men om omfattande spill eller läckage skett kan föroreningarna även förekomma djupare och ha nått grundvattnet där lokal kraftig förorening kan förväntas. Ämnena fastläggs dock i jord och spridningen förväntas vara måttlig.
Upplagsytan	Koppar, krom, 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 urlakning av förorening går långsamt förväntas en marginell påverkan på grundvattnet.
Ev. nedgrävt impregneringssalt/utfällningar	Koppar, krom, arsenik. Okänd punktkälla	Om impregneringssalt verkligen grävts ner bedöms djupet vara relativt litet, max 1-2 meter. Mycket höga metallhalter i jord kan förväntas inom det begränsade område där saltet/utfällningarna ev har grävts ner. Ytan på området bedöms vara 2-4 m ² .
Utfyllnadsområdet	Diverse föroreningar. Känd till utbredning men ej till innehåll. Heterogen och diffus.	Området med utfyllda massor ligger utanför fasthetsgränsen. Det är dock inte omöjligt att restprodukter från impregneringen lagts här tillsammans med t ex spån, bark, rivningsrester osv. Sammansättningen i det utfyllda området är okänd.

A.4 Exempel A. Diesalcisternen – att avgränsa en känd *hotspot*

A.4.1 PROVTAGNINGENS SYFTE

Historiken i området indikerar att det är relativt säkert att någon form av utsläpp i av diesel har skett i närheten av cisternen. Syftet med provtagningen är att avgränsa dieselföroreningen inför en eventuell efterbehandling.

A.4.2 TILLGÄNGLIG FÖRHANDSKUNSKAP

Diesalcisternens närmaste omgivning kan betraktas som en känd *hotspot* eftersom vi känner dess läge. Däremot är kunskapen ofullständig om hur föroreningen spridits i marken och hur långt föroreningen nått. Större utsläpp runt diesalcisternen antas främst ha rört sig vertikalt ned mot grundvattenytan och därefter länkats av och spridits längs grundvattenytan mot vattendraget. Eftersom grundvattenytan fluktuerar kan den föroreningen vara ”utsmetad” i zonen där grundvattnet fluktuerar. Detta innebär att även om ytskiktet är rent kan jordlager längre ner i markprofilen uppvisa förhöjda halter pga. att föroreningen kan transporteras längs grundvattenytan.

A.4.3 RUMSLIGA OCH TIDSMÄSSIGA AVGRÄNSNINGAR

Baserat på förhandskunskapen avgränsas sökområdet i horisontalled till cisternens direkta närområde, dock något utökat som en ellips med utsträckning i nordvästlig-sydöstlig riktning. I vertikalled görs en ganska fin indelning i halvmetersskikt ned till ca. 0,5 m under grundvattenytan (belägen ca 2-3 m under markytan) för att kunna avgränsa föroreningen med en hög noggrannhet.

A.4.4 PROVTAGNINGSSKALA OCH TYP AV PROV

Endast enskilda prover är av intresse. Små volymer jord provtas med hjälp av en skruvborr. Samtliga prover skickas inte in till laboratorium utan ett fältinstrument används för mätning direkt i fält (t.ex. PetroFLAG) trots att mätmetoden är förknippad med vissa osäkerheter.

A.4.5 ANGREPPSSÄTT

Här väljs ett bedömningsbaserat angreppssätt eftersom syftet är att avgränsa en *hotspot* med känt läge.

A.4.6 ANTAL PROV OCH PUNKTERNAS PLACERING

Det exakta antalet prov som sänds för laboratorieanalys beror av fältmätningen.

Det bästa är att definiera haltgränser för när ett prov bör betraktas som rent (mätvärde < haltgräns X_1) respektive förorenat (mätvärde > haltgräns X_2). Mätvärden mellan dessa haltgränser (haltgräns X_1 < mätvärde < haltgräns X_2) skickas till laboratorium för kompletterande analys.

Provpunkternas lägen

Proverna placeras i sex radiella linjer från cisternens läge med 5 meter mellan varje provpunkt längs varje snitt.

Tolkning av data

Provtagningen avbryts när man har funnit minst två rena prover i följd. Dock bör man vara uppmärksam på att även om ytskiktet är rent, kan jordlager längre ner i profilen uppvisa förhöjda halter p.g.a. att föroreningen kan transporteras längs grundvattenytan, som i sin tur kan fluktuera.

A.4.7 KOMMENTARER TILL EXEMPEL A

Exemplet är tillämpbart i de fall det är önskvärt att först avgränsa en känd *hotspot* och därefter besluta om efterbehandling. I NV (1994) beskrivs detta som snittmetoden.

A.5 Exempel B. Impregneringsanläggningen

A.5.1 PROVTAGNINGENS SYFTE

Det övergripande syftet är att bekräfta misstanke om förorening som kan ha spridits i samband med impregnering i den f.d. anläggningens impregneringscylinder. Provtagningens huvudsakliga syfte är således att belägga förekomst av förorening. Ett delsyfte är även att beräkna representativa halter för områdets föroreningar. De representativa halterna ska senare kunna jämföras med riktvärden i en riskbedömning.

A.5.2 TILLGÄNGLIG FÖRHANDSKUNSKAP

Det är många år sedan impregneringsanläggningen lades ner och byggnaderna är sedan länge rivna. Den exakta placeringen av impregneringscylindern är därför inte känd, även om man tror sig veta ungefär var den funnits. Kunskapen om området sammanställs i form av en konceptuell modell som beskriver de förväntade förhållandena på platsen ur förorenings-, spridnings- och risksynpunkt.

Koppar, krom och arsenik är adsorberande ämnen som bedöms förekomma bundet till jordpartiklarna där spill av impregneringsmedel skett. Har stora spill skett kan dock impregneringsvätska ha nått grundvattenytan. Bedömningen är att halterna av dessa ämnen är lognormalfördelade.

A.5.3 RUMSLIGA OCH TIDSMÄSSIGA AVGRÄNSNINGAR

Det område som misstänks vara förorenat vid markytan har en yta på ca 50×20 m². I vertikalled görs en avgränsning vid djupet 1 m under markytan.

A.5.4 PROVTAGNINGSSKALA OCH TYP AV PROV

Eftersom en av de föroreningar som förväntas förekomma i högst halter är akuttoxisk (arsenik)¹⁶ är det viktigt att inte provtagningsskalan väljs för stor. Om den volym proverna representerar blir för stor kan den akuta risken

¹⁶ Den akuttoxiska referenshalten för arsenik baserat på TDAE (Tolerable Dose for Acute Effects) är 0,05 mg/kg kroppsvikt (NV, 1996a; NV, 2007). Med ett intag av 5 gram jord för ett barn på 10 kg blir den akuttoxiska halten i jord 100 mg/kg.

underskattas. Därför väljs provtagningskalan som en cylinderformad volym med diametern ca 0,1 m och höjden 1 m. Detta gör det möjligt att ta proverna med skruvborr.

A.5.5 ANGREPPSSÄTT

Det finns förhandskunskap som bedöms vara viktig, både beträffande placeringen av tidigare föroreningskällor på området och hur de aktuella föroreningarna uppträder. Därför väljs ett bedömningsbaserat angreppssätt med riktad provtagning för att belägga förekomst av förorening. Den bedömningsbaserade provtagningen bör dock likna sannolikhetsbaserad provtagning som möjligt eftersom data ska användas för att beräkna representativa halter för området. Provtagningen baseras på enskilda prov.

A.5.6 ANTAL PROV OCH PROVPUNKTERNAS PLACERING

Ingen riktigt bra metodik finns för att beräkna nödvändigt antal prov för en bedömningsbaserad provtagning. I detta exempel ska dock mätdata inte bara användas för att bekräfta föroreningen utan även för att ta fram representativ statistik. Därför kan graferna för sannolikhetsbaserad provtagning i avsnitt 3.6.5 användas som utgångspunkt. Metodiken att beräkna nödvändigt antal prov baserat på konfidensintervall väljs.

Den representativa halt som senare ska tas fram för exempelvis arsenik måste ha en viss säkerhet. Önskad säkerhet väljs genom att man bestämmer hur stort avståndet ska vara mellan UCL och det verkliga, men okända, medelvärdet. Här bedömer vi det som acceptabelt om UCL är 50% högre än det verkliga medelvärdet, vilket innebär att $D=1,5$ enligt ekvation 3-3 ($UCL = 1,5 \times$ verklig medelhalt). Mätdata antas vara lognormalfördelade med $CV=1,2$. Detta ger enligt Bilaga G, Figur G-2 att ca 24 prov krävs, vilket motsvarar 2,4 prov per 100 m².

Ungefär hälften av de 24 proverna riktas subjektivt mot punkter där förorening förväntas eller där osäkerheten bedöms vara störst. Syftet med dessa provpunkter dokumenteras. För att få en så god spridning av provpunkterna som möjligt slumpas övriga provpunkter ut med hjälp av systematiskt slumpmässig provtagning, se Bilaga C.

Ett rutnät med 24 rutor läggs ut över området och provpunkter slumpas ut i de rutor som saknar provpunkter efter den riktade provtagningen.

A.5.7 KOMMENTARER TILL EXEMPEL B

Bedömningsbaserad provtagning kan leda till problem när representativa halter ska beräknas för ett område. Den metodik som föreslås i detta exempel visar på en enkel metod att ändå kunna hantera detta. Lösningen ligger i att den riktade provtagningen fått ett stort inslag av slumpmässig provtagning och att provpunkterna fördelas någorlunda jämnt över området. Detta minskar de negativa effekterna av riktad provtagning. Trots detta finns det en risk att medelvärdet av mätdata kan bli för högt, d.v.s. att långtidsrisker i området kan överskattas.

Notera att antalet prov har beräknats med utgångspunkt från konfidensintervall för medelvärde, vilket kan kopplas till långtidsrisker. Andra metoder krävs om fokus istället ligger helt och hållet på akuta risker.

Bestämningen av medelhalten i det avgränsade området är inte kopplad till områdets storlek. Dock finns det en koppling mellan storleken på området och de osäkerheter man är beredd att ta, dvs. de kostnader som en felbedömning kan ge upphov till.

A.6 Exempel C. Upplagsytan

A.6.1 PROVTAGNINGENS SYFTE

Det övergripande syftet med provtagningen är att bedöma om upplagsytan är förorenad över riktvärden samt att bedöma hur stor mängd förorening som finns. Därför ska representativa halter tas fram för området och dessa ska användas i en riskbedömning samt för mängdberäkningar. Detta kräver att representativ statistik tas fram.

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

A.6.2 HANTERING AV FÖRHANDSKUNSKAP

Viktig förhandskunskap om objektet är följande:

- De ämnen som har hanterats på platsen är As, Cu och Cr.
- Förorening har förmodligen hamnat i marken genom spill och droppning.
- Föroreningen kan vara spridd över hela upplagsytan, men hypotesen är att endast de ytliga jordlagren förorenats (0-0,5 m under markytan).
- Föroreningshalterna kan antas vara lognormalfördelade.

A.6.3 RUMSLIGA OCH TIDSMÄSSIGA AVGRÄNSNINGAR

Hela området betraktas som en enhet. Inga tidsmässiga avgränsningar görs eftersom föroreningshalterna i jord antas förändras mycket långsamt över tiden.

A.6.4 PROVTAGNINGSSKALA OCH TYP AV PROV

Arsenik har akuttoxiska egenskaper, vilket innebär att även små volymer jord behöver kartläggas. Provgropsgrävning utförs för att få en bättre överblick över jordlagerföljen men enskilda prover från groparna analyseras. Varje prov representerar en volym på endast några få liter. Samlingsprov från provgrop bedöms inte vara lämpligt i detta fallet då provtagningsskalan bör vara liten för att ta hänsyn till de akuttoxiska egenskaperna.

A.6.5 ANGREPPSSÄTT

Eftersom provtagningen syftar till att ta fram representativ statistisk för området är det lämpligt att använda ett sannolikhetsbaserat angreppssätt.

A.6.6 SANNOLIKHETSBASERAT ANGREPPSSÄTT

Val av statistisk parameter

Medelhalten i området bedöms vara lämplig att använda för mängdberäkningar och medel- respektive maxhalter för riskbedömning. Maxhalter är intressant att titta på ur akuttoxisk synpunkt. Provtagningen läggs dock upp med avseende på medelhalt.

Uppskattning av variabilitet

Den lilla provtagnings-skalan innebär att man kan förvänta sig en relativt hög variabilitet i data. Vidare förväntas föroreningen förekomma relativt heterogent, dvs. förorenade ytor blandade med rena ytor, till följd av det spill och dropp som förekommit. Variabiliteten skattas som en variationskoefficient, CV. Här skattar vi CV till 1,5 för arsenik som beräkningarna baseras på.

A.6.7 ÖNSKAD SÄKERHET

I exemplet används två olika angreppssätt för att bestämma provtagningens omfattning; ett som bygger på konfidensintervall och ett som bygger på hypotestest. Vid praktiska tillämpningar bör man välja en metod där man förstår bakomliggande teori. Det kan även vara lämpligt att använda sig av båda metoderna för att se om slutsatserna blir ungefär desamma

Konfidensintervall

I rapporten Åtgärdskrav vid efterbehandling (NV, 1997) används en 95%-ig konfidensgrad vid bestämning av medelhalten. Denna konfidensgrad används även här.

Önskad säkerhet bestäms med kvoten D, se ekvation 3-3 i rapportens huvudtext. Kvoten D är ett mått på hur många gånger över den verkliga medelhalten som man accepterar att det uppmätta medelvärdet kan hamna. I detta fall vill vi beräkna hur många prover som behövs för två olika fall: då $D = 1,4$ respektive $D = 1,5$. Detta innebär att vi accepterar att uppmätt medelhalt kan vara upp till 40% respektive 50% högre än det verkliga medelvärdet.

Hypotestest

Hypoteserna om förorenings-situationen formuleras på följande sätt:

- Nollhypotesen H_0 : medelhalten \geq riktvärdet
- Alternativhypotesen H_A : medelhalten $<$ riktvärdet

α -felet definieras som risken att felaktigt förkastar en sann nollhypotes.

Konfidensgraden 95% (se ovan) innebär att man accepterar en felrisk på 5% att lämna ett förorenat område utan åtgärd, dvs. $\alpha = 0,05$.

Den andra feltypen (β -felet), dvs. att felaktigt acceptera nollhypotesen, är svårare att kvantifiera. Här väljs en säkerhet på 80%, vilket rekommenderas som lägsta nivå av Grandin (2006). Detta innebär att $\beta = 0,2$. Storleken på β -felet är något som problemägaren själv är intresserad av att hålla så låg som möjligt, medan en tillsynsmyndighet är mindre oroad för storleken på detta fel.

En minsta detekterbar skillnad (Δ) måste också väljas, se avsnitt 4.4 i huvudrapporten samt Bilagorna D och F för diskussion och förklaring. Riktvärdet för arsenik är 15 mg/kg TS. Här väljer vi Δ som är 30% av riktvärdet, dvs. $\Delta = 0,3$. Detta innebär att om den verkliga medelhalten på området är högre än 10,5 mg/kg ($15 - 0,3 \times 15 = 10,5$) så kommer hypotestestet sannolikt att indikera att H_0 inte kan förkastas, dvs. området kommer att klassas som förorenat. En beräkning görs även för $\Delta = 0,4$.

A.6.8 PROVTAGNINGSMÖNSTER

Här väljer vi en systematisk slumpmässig provtagning. Systematiken skapar vi genom att dela in området i celler (delytor, block) och proverna slumpas ut inom varje cell. Cellindelningen baseras på hur många prover som skall tas och vi väljer att dela in området i 10 lika stora celler. För att kunna undersöka den rumsliga korrelationen med hjälp av variogramanalys är det viktigt att placera provpunkter på *olika* avstånd från varandra; några på korta avstånd och andra på längre avstånd (se även Bilaga B i NV, 1996b). Detta kan göras genom att slumpa ut olika antal provpunkter i varje cell, där även cellerna väljs slumpmässigt.

Antal prov baserat på konfidensintervall

Antalet prov som krävs för den önskade säkerheten kan tas fram med hjälp av Figur G-2 i Bilaga G eller Figur 3-8 i huvudrapporten. Med värdena $CV = 1,5$ samt $D = 1,4$ respektive $D = 1,5$ så krävs 60 respektive 44 prover för att nå önskad säkerhet i medelvärdet, vilket motsvarar ca 2,9 respektive 2,1 prov per 100 m².

Antal prov baserat på hypotestest

Antalet prov som krävs för den önskade säkerheten tas fram genom att gå in i Figur G-4 i Bilaga G eller Figur 3-9 i huvudrapporten. Variationskoefficienten CV är 1,5 och de acceptabla felen är $\alpha = 0,05$ respektive $\beta = 0,20$. Om $\Delta = 0,3$ så krävs 59 prover och om $\Delta = 0,4$ krävs 30 prover. I det fall man redan tidigare har tagit prover kan dessa också användas i utvärderingen senare. Det nya antal prover som i så fall måste samlas in är 59 eller 30 minus det antal som redan finns. Detta motsvarar ca 2,8 respektive 1,4 prov per 100 m².

A.6.9 KOMMENTARER TILL EXEMPEL C

Beroende på vilken metod som används och vilka krav på säkerhet som ställs varierar nödvändigt antalet prover från 10 till 59 stycken, vilket är ett ganska stort spann. Detta illustrerar att det är viktigt att testa sig fram och resonera omkring kraven på säkerhet och tillgänglig budget för att komma fram till ett rimligt antal provpunkter.

I avsnitt A.6.1 anges att ett delsyfte med provtagningen är att undersöka om det finns någon rumslig korrelation hos föroreningshalterna i området. Med detta syfte är 59 prover det bästa alternativet, men omfattningen bör vägas mot kostnaderna för provtagningen. .

Såsom nämndes i Exempel B så är bestämningen av medelhalten i det avgränsade området inte kopplad till områdets storlek. Men det finns en relation mellan storleken på området och vilka osäkerheter man är beredd att ta, dvs. de kostnader är förknippad med en felbedömning på grund av osäkerheterna.

A.7 Exempel D. Nedgrävt impregneringssalt

A.7.1 SYFTE

Det övergripande syftet är att försöka lokalisera en *hotspot* men man har bara väg information om, och i så fall var, den kan finnas¹⁷.

A.7.2 FÖRHANDSKUNSKAP

Den förhandskunskap som finns om det eventuellt nedgrävda impregneringssaltet används dels för att avgränsa sökområdet, dels för att bestämma storleken på området där saltet är nedgrävt.

A.7.3 RUMSLIGA OCH TIDSMÄSSIGA AVGRÄNSNINGAR

Utifrån vad som skrivits tidigare om förhandskunskaperna antas att det nedgrävda saltet finns inom tomtavgränsningen, inte vid impregneringsanläggningen och inte inom upplagsytan. Sökområdet är drygt 5500 kvadratmeter stort.

A.7.4 PROVTAGNINGSSKALA

Skruvborr används.

A.7.5 ANGREPPSSÄTT

Här väljs ett sökbaserat angreppssätt pga. att vi inte har någon förhandskunskap om var det nedgrävda saltet kan ligga i området.

A.7.6 SÖKBASERAT ANGREPPSSÄTT

Form och storlek

Hotspoten antas ha en cirkulär form med radien 1 m.

Provtagningsmönster

Ett triangulärt provtagningsmönster väljs i VSP (2008) för att detta är något mer effektivt än ett rektangulärt provtagningsmönster i samband med *hotspotsökning*.

¹⁷ Rapporten är inriktad på provtagning i jord men normalt skulle man försöka lokalisera saltet med hjälp av grundvattenprover. Exemplet får dock illustrera arbetsmetodiken för jord.

Önskad säkerhet

Det antas vara rimligt att ha en önskad säkerhet på 90% att detektera *hotspoten*, vilket innebär att sannolikheten att missa *hotspoten* är 10%. Om man lyckas borra precis i *hotspoten* förväntas man direkt kunna identifiera att materialet skiljer sig från det naturliga isälvsavlagringarna, dvs. det antas att det inte sker någon falsk negativ detektering.

Antal prov

Antalet prov som behövs för att hitta *hotspoten* med 90% säkerhet beräknas med hjälp av VSP (2008) till ca 1 600 prover (!) med ungefär 2 meter mellan provpunkterna

A.7.7 KOMMENTARER TILL EXEMPEL D

Det är naturligtvis orimligt att provta och analysera 1 600 prover för att hitta saltet i detta fall. Ett mer rimligt angreppssätt kan vara att försöka göra en leta efter saltet med hjälp av andra metoder än traditionell jordprovtagning, t.ex. grundvattenprovtagning eller geofysiska metoder. Man kan också fundera över vilken säkerhet man vill ha att detektera *hotspoten*. Utgör den en stor risk? Om så är fallet kan 90% säkerhet i resultaten t.o.m. anses vara ganska osäkert. Om den inte utgör en stor risk kan man fundera över hur nödvändigt det är att lokalisera den.

Det är generellt väldigt svårt att hitta en liten *hotspot* genom jordprovtagning, det blir som att leta efter en nål i en höstack. För att sådan här sök teori skall vara lämplig att använda bör *hotspoten* vara relativt stor i förhållande till sökområdet. Hade t.ex. *hotspoten* haft en radie på 5 m och sökområdet en area på ca 450 m² så hade endast 7 prover krävts för att hitta *hotspoten* med 95% säkerhet.

A.8 Exempel E. Utfyllt område

A.8.1 SYFTE

Det övergripande syftet med att undersöka det utfyllda området är att avgöra om det är förorenat över riktvärdet (dvs. om området utgör en oacceptabel risk eller ej) samt att översiktligt bedöma mängden förorening. Båda syftena innebär att *beskrivande statistik* av föroreningsituationen i området ska tas fram. Eftersom området är utfyllt med massor med olika ursprung förväntar man sig att den rumsliga korrelationslängden är förhållandevis kort. Därför förväntar man sig inte att kunna göra någon interpolering mellan provpunkterna, dvs. det upprättas inte någon karta med interpolerade föroreningshalter. Om oacceptabel förorening konstateras i området kan det senare bli aktuellt med förtätad provtagning för att avgränsa och klassa föroreningen inför åtgärd (rutnätsschaktning).

A.8.2 FÖRHANDSKUNSKAP

Man misstänker att fyllnadsmassorna i huvudsak utgörs av massor med viss grad av PAH- och metallförorening. Även andra typer av massor kan dock förekomma. Förhandskunskapen är därför relativt svag. Inga akutttoxiska halter förväntas dock.

A.8.3 RUMSLIGA OCH TIDSMÄSSIGA AVGRÄNSNINGAR

Det finns ingen information om att vissa delar av det utfyllda området skulle vara mer förorenade än andra. Därför betraktas hela området som en enhet (x- och y-koordinater). I djup (z-koordinat) görs däremot en indelning i 1-metersintervall ner till naturliga jordlager. På grund av områdets läge i terrängen bedöms att djupet till grundvattenytan varierar mellan 0,5 m och 1,5 m inom området. Delar av fyllnadsmassorna kan därför befinna sig under grundvattenytan.

A.8.4 PROVTAGNINGSSKALA

Det är önskvärt att proverna som tas representerar en relativt stor volym jord så att inte småskaliga haltvariationer i fyllnadsmaterialet påverkar resultatet alltför mycket. Samlingsprov från provgrop är därför lämpligt¹⁸ eftersom ett sådant prov representerar någon eller några kubikmeter jord, om det tas på rätt sätt. Detta är en acceptabel provtagningskala även ur risksynpunkt eftersom inga akuttoxiska halter förväntas.

A.8.5 ANGREPPSSÄTT

Det finns ingen förhandskunskap som skulle kunna utnyttjas angående var eventuell förorening främst finns. Därför bör ett sannolikhetsbaserat angreppssätt väljas.

A.8.6 SANNOLIKHETSBASERAT ANGREPPSSÄTT

Val av statistisk parameter

Medelhalten ska användas när nödvändigt antal samlingsprov ska beräknas. När provdata utvärderas planerar man att beräkna en representativ halt för området. Som representativ halt har man tänkt använda övre 95%-konfidensgränsen för medelvärdet (95% UCL). Detta värde kommer att jämföras mot riktvärdet i riskbedömningen.

Uppskattning av variabilitet

För att bestämma antalet prover behöver variabiliteten uppskattas. Observera att det är variabiliteten i provgropsskalan, dvs. variationen mellan olika provgropar, som ska skattas. Det finns inget sätt att beräkna variabiliteten utan att ha några data från området. Enda lösningen är att använda erfarenheter från liknande områden och rimlighetsbedömningar. Då området är utfyllt förväntas en förhållandevis stor variabilitet. Å andra sidan förväntas inte heller några direkta *hotspots*. Variationskoefficienten (CV) från områden med liknande föroreningsbild brukar ligga mellan 1 och 2. Vi har här antagit att CV är 1,5. Detta innebär att standardavvikelsen är 150% av medelvärdet (se avsnitt 3.6.2).

¹⁸I de fall föroreningsämnen är flyktiga, är samlingsprov inte lämpliga.

A.8.7 ÖNSKAD SÄKERHET

Konfidensintervall

Då en relativt stor variation i föroreningshalterna bedöms finnas inom utfyllnadsområdet, är det inte heller rimligt att ställa alltför höga krav på säkerhet för medelvärdet. En rimlig grad av säkerhet bedöms i detta fall till 80% påslag på medelhalten. Detta innebär att om den verkliga medelhalten är 100 mg/kg så accepterar man det uppmätta (dvs. det beräknade) medelvärdet får vara upp till 180 mg/kg. Detta innebär att $D = 1,8$ i ekvation 3-3 i avsnitt 3.6.3.

Notera att när man väl får provresultaten så garderar man sig för osäkerheterna genom att använda sig av 95% UCL vid jämförelsen mot riktvärdet, se ovan. Därför kan 80% osäkerhet i medelvärdet vara fullt acceptabelt i detta fall.

Hypotestest

För α - respektive β -fel väljer man en maximal felrisk på 5% respektive 20%. För minsta detekterbara skillnad (effektstorleken) väljs 50% av riktvärdet. Om effektstorleken väljs till 50% av riktvärdet så innebär detta i praktiken att om den verkliga medelhalten ligger på hälften av riktvärdet så kommer området att klassas som förorenat.

A.8.8 PROVTAGNINGSMÖNSTER OCH ANTAL PROV

För att få en jämn spridning av provpunkterna över hela området, men ändå en viss variation i avstånd mellan punkterna, väljs systematisk slumpmässig provtagning. Detta provtagningsmönster eliminerar många av nackdelarna med rent slumpmässig eller systematisk provtagning (NV, 1994).

Antal prov baserat på konfidensintervall

Antalet prov som krävs för att nå önskad säkerhet beräknas med antagandet att populationen är lognormalfördelad.

Antalet prov bestäms med hjälp av grafen i Figur 3-8. Med värden $CV = 1,5$ och $D = 1,8$ indikerar figuren att det krävs ca 31 prover. Om budgeten inte medger så stort antal så måste man sänka kravet på säkerhet.

Antal prov baserat på hypotestest

Data antas vara lognormalfördelad. Grafen i Figur 3-9 används för att avläsa antalet prover. Grafens x-axel har en logaritmisk skala, vilket gör det lite svårt att läsa av exakt, men ca 20 prover kan utläsas av grafen om $\delta = \text{riktvärdet} \times 0,5$ och ca 60 prover om $\delta = \text{riktvärdet} \times 0,3$ (dvs. samma som i Exempel 3).

A.8.9 KOMMENTARER TILL EXEMPEL E

Beräkning av antal prov gjordes med två olika metoder som båda bygger på lognormalfördelade data. Resultaten är inte helt jämförbara eftersom de bygger på olika metoder med olika antaganden. Det som påverkar resultaten starkt är vilken variabilitet man förväntar sig i området samt vilken säkerhet man vill uppnå i resultatet.

A.9 Referenser

- Back, P-E., 2003. On Uncertainty and Data Worth in Decision Analysis for Contaminated Land. Licentiate thesis. Publ A 105. Department of Geology, Chalmers University of Technology.
- Grandin, U., 2006. Statistisk analys av möjligheter att kunna upptäcka regionala trender i de nuvarande programmen för nationell övervakning av sötvatten. Ett underlag för revisionen av programmen år 2006. Rapport Sveriges lantbruksuniversitet, Miljöanalys vol 18.
http://publikationer.slu.se/Filer/IMA2006_18.pdf (2008-05-26).
- NV, 1994. Vägledning för miljötekniska markundersökningar. Del I Strategi. Rapport 4310. Naturvårdsverket, Stockholm. <http://www.naturvardsverket.se/Documents/publikationer/620-4310-2.pdf> (08-06-26)
- NV, 1996a. Development of generic guideline values Rapport 4639. Naturvårdsverket, Stockholm. <http://www.naturvardsverket.se/Documents/publikationer/620-4639-X.pdf> (08-07-01)
- NV, 1996b. Rätt datakvalitet. Rapport 4667. Naturvårdsverket, Stockholm. <http://www.naturvardsverket.se/Documents/publikationer/620-4667-5.pdf> (08-06-29)
- NV, 1997. Åtgärdskrav vid efterbehandling. Rapport 4807. Naturvårdsverket, Stockholm. <http://www.naturvardsverket.se/Documents/publikationer/620-4807-4.pdf> (08-06-26)
- NV, 2007. Riktvärden för förorenad mark - Modellbeskrivning och vägledning. Bilaga 1 Sammanställning av indata till riktvärdesmodellen. Remissversion 2007-10-19.
- US EPA, 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)
- VSP, 2008. Visual Sample Plan. Software. Pacific Northwest National Laboratory, PNNL. <http://vsp.pnl.gov/> (2008-05-26)

B Ordlista

Action Limit (AL)	<i>Eng.</i> begrepp som ofta används av US EPA och anger den halt vid vilken åtgärder av något slag bör vidtas. I denna rapport används ”riktvärde” 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
Bakgrundshalt	Naturlig halt plus antropogent diffust tillskott.
Datakvalitetsmål	Ett systematiskt angreppssätt för att definiera de kriterier 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 provtagnings omfattning som bygger på att provtagningskostnaden balanseras mot den nytta som provtagningen genererar. 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 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. 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> .
Fyllning	Av människan påförda lösa massor som kan bestå av 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. Vanligen används platsspecifika riktvärden vid bedömningen.
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.
Grey region	Begrepp använt av US EPA. För förklaring, se <i>Effektstorlek</i> .
Hotspot	Eng. begrepp för starkt förorenat och till yta/volym begränsat område i mark eller sediment.
Inferens	I statistik: att dra slutsatser från stickprov till population.
Konceptuell modell	En beskrivning av objektet som skall undersökas, baserad på tillgänglig information och utformad som 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. Ibland även konfidensnivå.
Konfidensgräns	De gränser som avgränsar ett konfidensintervall. Man skiljer på övre konfidensgräns och nedre konfidensgräns.
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 .

Känslighet/skyddsvärde	En av de fyra delar som tillsammans bildar grunden för bedömning av risker. Här bedöms hur människor, växter och djur kan exponeras för föroreningarna och hur allvarligt man ser på denna exponering.
LCL	<i>Eng.</i> lower confidence limit. Den undre gränsen för ett konfidensintervall (se <i>Konfidensgräns</i>). Inom förorenad mark är oftast denna mindre intressant än UCL.
LCL95	<i>Eng.</i> 95% lower confidence limit. Den undre 95%iga gränsen av ett konfidensintervall.
Markanvändning	Det ändamål för vilket ett mark- eller vattenområde utnyttjas eller kommer att utnyttjas.
Medelhalt	Här används medelhalten för att beteckna den verkliga men okända medelhalten i en målpopulation.
Medelvärde	Här används medelvärdet för att beteckna <i>skattningen</i> av medelhalten, dvs. det medelvärde man observerar från sitt stickprov.
Medium	Avser i rapporten mark, grundvatten, sediment, ytvatten, byggnader och anläggningar.
Minsta detekterbara skillnad Målpopulation	Se <i>Effektstorlek</i> . Den totala mängd av koncentrationvärden i delområdet som vi önskar uttala oss om (i princip en oändlig mängd).
Naturlig halt	Den halt som skulle finnas utan antropogen påverkan.
Platsspecifikt riktvärde	Riktvärde framtaget för ett objekt med dess speciella förutsättningar.
Provtagningsmönster	En beskrivning av det sätt på vilket provpunkter placeras i rummet.
Provtagningsplan	Den färdiga produkten som beskriver provtagningsstrategin och hur provtagningen ska utföras.
Provtagningsskala	<i>Eng. support.</i> Den volym som ett prov representerar, inklusive dess form och orientering. Provtagningsskalan kan bl.a. ökas genom att ta samplingsprov.
Provtagningsstrategi	Innefattar själva metodiken och upplägget av provtagningen.
Provtagningsvolym	Den jordvolym som fysiskt tas ut som prov. Provtagningsvolymen är alltid mindre eller lika med provtagningsskalan.
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.
Riktvärde	Den halt av förorening över vilken risk för oönskade effekter på människor eller miljö kan föreligga. Se <i>Generellt riktvärde</i> respektive <i>Platsspecifikt riktvärde</i> .
Riskbedömning	Identifiering och bedömning (kvantitativ eller kvalitativ) av de risker för människors hälsa eller miljön som ett förorenat område kan ge upphov till.
Screeningtest/metod Slumpmässig	Metod för snabb översiktlig undersökning. 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 fördelningens varians.
Statistisk styrka	Sannolikheten att förkasta en nollhypotes som är felaktig. Definieras som $1-\beta$.
Stickprov	En samling slumpmässigt utvalda prov som skall representera den målpopulation man vill undersöka.
Stratifierad provtagning	Provtagning där den volym som ska provta 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 Tillsynsmyndighet	Se <i>Provtagningskala</i> . 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 arbeten, allt enligt miljöbalken.
UCL	<i>Eng. upper confidence limit</i> . Den övre gränsen av ett konfidensintervall (se <i>Konfidensgräns</i>).
UCL95	<i>Eng. 95% upper confidence limit</i> . Den övre 95%-iga gränsen av ett konfidensintervall.
Utredningsstrategi	Innefattar själva planeringen och upplägget av utredningen.

Varians	Varians och standardavvikelse är exempel på spridningsmått för en sannolikhetsfördelning, det vill säga ett mått på hur utspridd fördelningen är kring väntevärdet.
Åtgärds mål	Funktionsmål som anger hur området kan användas efter åtgärderna, samt kvantitativa 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.

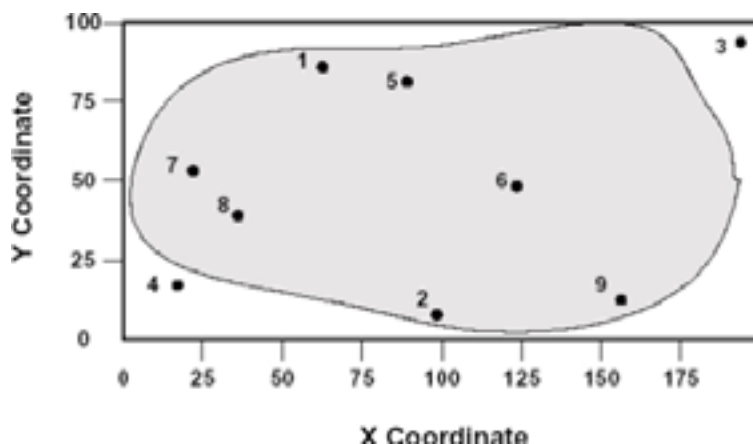
C Provtagningsmönster

C.1 Slumpmässigt provtagningsmönster

Med ett slumpmässigt provtagningsmönster placeras provtagningspunkterna inom undersökningsområdet oberoende av förhandskunskap eller någon form av systematik. Det är viktigt att placeringen blir slumpmässig och någon form av slumpgenerator bör användas t.ex. i Excel, VSP (2008) eller i SADA (2008).

Efter att antalet prover har bestämts väljs deras koordinater (helst i tre dimensioner) av en slumpgenerator. I Figur C-1 presenteras ett exempel där X och Y-koordinater för 9 provpunkter har genererats med slumpgenerator inom en rektangel och lagts ovanpå undersökningsområde. Eftersom provtagningsområdet har en annan geometri än rektangel faller vissa punkter utanför – de punkterna används inte i provtagningen.

Fördelarna med provtagningsmönstret är resultatets oberoende av provpunkters koordinater (*selection bias*). Detta gäller dock under förutsättning att provmängden är representativ i relation till syftet med provtagningen. En annan fördel med denna typ av provtagning är att statistisk analys av provresultat följer väl etablerade och förhållandevis enkla principer. En nackdel är dock att slumpen kan innebära att provpunkter ansamlas på en viss plats (klusterbildning) eller att stora delar av området inte täcks av provtagningen. Viss förhandskunskap om området kan utnyttjas genom att dela in området i egenskapsområden och utslumpningen kan därefter göras separat för varje sådant område.



Figur C-1. Randomisering av 9 provpunkter inom en rektangel innehållande ett undersökningsområde. Punkter 3 och 4 faller utanför området (Engelke & Norrman, 2009).

C.2 Systematiska provtagningsmönster

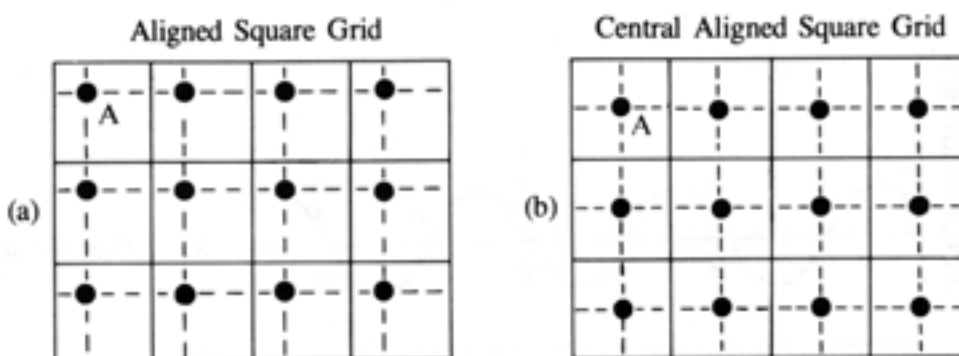
Systematisk provtagning bygger på ett specifikt provtagningsmönster där huvudprincipen är att prover tas med jämna intervall. Den ger en jämn täckning av området och är enkelt att genomföra. Systematisk provtagning kan användas både för rumsliga- och tidsrelaterade problem (t ex nedbrytning av

petroleumprodukter i jord kan övervakas genom provtagning inom samma område vid jämna tidsintervall under ett antal år).

Systematiska provtagningsmönster kan ha olika konfigurationer: rutnät, triangulär nät, rektangulär, eller sk. fiskbensmönster, se Figur C-2. Systematisk provtagning används i följande situationer:

- När syftet med provtagning är att skatta statistiska parametrar (t ex medelvärden) för ett område med heterogen föroreningsbild.
- När syftet är att undersöka trender i data eller att studera om data är rumslig- eller tidkorrelerade. Genom att studera korrelationsstruktur (om sådan kan påvisas att existera i data) kan vissa slutsatser om storleken och geometri på förorenade volymer dras.
- När syftet är att hitta *hotspots*. Konfigurationen på provtagningsnät anpassas till önskad sannolikhet att detektera en potentiell *hotspot* med en given storlek och geometri.

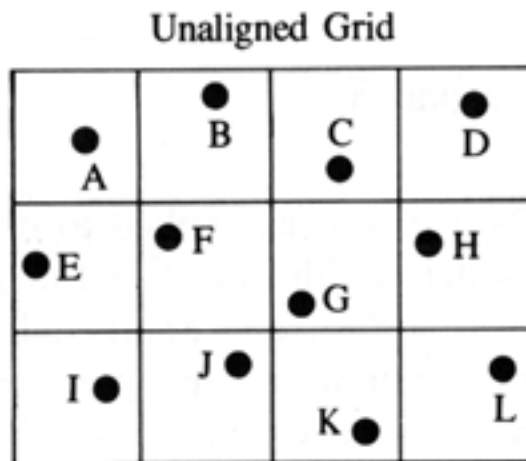
Nackdelarna med en systematisk provtagning är t.ex. att förhandskunskapen inte utnyttjas i någon högre grad och att proverna hamnar samma avstånd från varandra. Att det blir samma avstånd mellan provtagningspunkterna innebär att det inte går att kontrollera hur halterna varierar vid olika avstånd mellan provtagningspunkterna, något som bör göras för att kunna interpolera mellan provtagningspunkterna. Det finns även risk att det blir systematiska fel beroende på hur föroreningsförekomsten ser ut geometriskt.



Figur C-2. Två varianter på systematiskt provtagningsmönster med ett kvadratisk rutnät (Gilbert 1987).

C.3 Slumpmässigt systematiskt provtagningsmönster

Med ett slumpmässigt systematiskt provtagningsmönster läggs ett systematiskt rutnät över undersökningsområdet och ett (eller fler) prov slumpas inom respektive ruta, se Figur C-3. Fördelen med metoden är att hela undersökningsområdet täcks in samtidigt som risken för klusterbildning försvinner. Nackdelen är att provpunkternas placering inte blir helt slumpmässiga och att geometrin hos mönstret kan påverka resultatet.



Figur C-3. Exempel på slumpmässigt systematiskt provtagningsmönster (Gilbert 1987).

C.4 Referenser

Gilbert, R., 1987. Statistical methods for environmental pollution monitoring. John Wiley & Sons, Inc. New York.

SADA, 2008. Spatial Analysis and Decision Assistance. Software. University of Tennessee, UT. <http://www.tiem.utk.edu/~sada/index.shtml> (08-06-29)

Engelke, F., Norrman, J., Starzec, P., Andersen, L., Grøn, C., Overgaard J. & Refsgaard, A., 2009. Inventering av provtagningsstrategier för jord, grundvatten och porgas. Rapport 5894. Hållbar Sanering. Naturvårdsverket, Stockholm.

VSP, 2008. Visual Sample Plan. Software. Pacific Northwest National Laboratory, PNNL. <http://vsp.pnl.gov/> (2008-05-26)

D Hypotestest, styrka och antalet prov

Statistiska hypoteser följer ett strikt mönster, där man först formulerar en nollhypotes och därpå en eller flera mothypoteser eller alternativhypoteser. Det finns i princip två olika sätt att ställa upp sin nollhypotes (StatSoft, 2008): (1) nollhypotesen formuleras så att den beskriver att det inte finns någon skillnad eller effekt (RS, *reject-support*), eller (2) nollhypotesen formuleras så att den beskriver att det finns en effekt (AS, *accept-support*). Nollhypotesen är den slutsats som kommer att bli gällande om det visar sig att det saknas tillräckliga bevis för att den kan förkastas i en hypotesprövning. Om däremot bevisen att förkasta nollhypotesen är tillräckligt starka blir slutsatsen att alternativhypotesen är gällande. Mothypotesen kan formuleras en- eller tvåsidigt. När hypotestest används i samband med utvärdering av förorenad mark formuleras vanligen en ensidig mothypotes.

Det första alternativet, AS, dvs. att nollhypotesen beskriver att det inte finns någon effekt, är det vanligaste sättet att formulera sig i statistiska sammanhang (jfr Grandin, 2003). Inom förorenad mark använder man sig dock ofta av det andra alternativet, AS, dvs. att nollhypotesen formuleras så att den beskriver att det finns en effekt. Nollhypotesen betecknas ofta H_0 och mothypotesen H_1 . En nollhypotes för förorenad mark formuleras vanligen som: H_0 : Medelhalten av arsenik inom området är lika med eller högre än riktvärdet för arsenik. En ensidig mothypotes formuleras då som: H_1 : Medelhalten av arsenik i området är signifikant lägre än riktvärdet för arsenik.

Grandin (2003) beskriver teorin bakom hypotesprövning och redovisar en hierarkisk ordning av beräkningssteg i 8 punkter. I testproceduren kan nollhypotesen antingen förkastas eller inte förkastas, men aldrig accepteras! Några viktiga förutsättningar för statistiska tester är att de data man vill pröva skall vara ett slumpmässigt urval ur den bakomliggande populationen och att naturen hos data avgör vilka tester man kan använda. Normalfördelade eller lognormalfördelade data kan analyseras med *parametriska* tester medan data från andra fördelningar måste prövas med *icke-parametriska* metoder.

D.1 Felrisiker och styrka

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 det stickprov man samlat in, på grund av slumpen, visar ett annat mönster än hela populationen och att man därigenom tar ett felaktigt beslut när man förkastar nollhypotesen. Vid all hypotesprövning kan man hamna i fyra olika situationer, se Tabell D-1. Om nollhypotesen förkastas medan den *de facto* är sann begår man ett α -fel (typ I-fel). Alternativt, om nollhypotesen inte förkastas medan den *de facto* är falsk begår man ett β -fel (typ II-fel).

Tabell D-1. Möjliga utfall från statistisk hypotesprövning.

		Verkligt förhållande (okänt)	
		H ₀ sann	H ₀ falsk
Resultat av statistiskt test	H ₀ förkastas	Typ I-fel, α	Korrekt beslut, $1-\beta$ (styrka)
	H ₀ behålls	Korrekt beslut, $1-\alpha$	Typ II-fel, B

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 (US EPA (2006a) kallar det för *false rejection rate*), dvs. sannolikheten att man har fel om man förkastar nollhypotesen. Om $p = 0,05$ så är sannolikheten att man har fel om man förkastar nollhypotesen 5%. Detta är i allmänhet (men inte alltid) en acceptabel felrisk, och om det är en acceptabel felrisk så förkastas nollhypotesen och slutsatsen blir 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 som 40% är i de flesta sammanhang inte acceptabelt och slutsatsen blir att man inte kan förkasta nollhypotesen.

Typ II-fel, eller β , kallas i US EPA (2006a) för "false acceptance rate", dvs. felaktig acceptansnivå, och är alltså det fel som görs om man felaktigt accepterar en nollhypotes.

Beroende på hur nollhypotesen formuleras (RS eller AS i förra avsnittet) så innebär α och β olika saker, se sammanfattning i Tabell D-2 och D-3.

Tabell D-2. RS, reject-support: H₀: medelhalten av X är lika med eller lägre än riktvärdet.

		Verkligt förhållande (okänt)	
		H ₀ sann (området är rent)	H ₀ falsk (området är förorenat)
Resultat av statistiskt test	H ₀ förkastas (området klassas som förorenat)	Typ I-fel, α	Korrekt beslut, $1-\beta$ (styrka)
	H ₀ behålls (området klassas som rent)	Korrekt beslut, $1-\alpha$	Typ II-fel, B

Tabell D-3. AS, accept-support: H₀: medelhalten av X är lika med eller högre än riktvärdet.

		Verkligt förhållande (okänt)	
		H ₀ sann (området är förorenat)	H ₀ falsk (området är rent)
Resultat av statistiskt test	H ₀ förkastas (området klassas som rent)	Typ I-fel, α	Korrekt beslut, $1-\beta$ (styrka)
	H ₀ behålls (området klassas som förorenat)	Korrekt beslut, $1-\alpha$	Typ II-fel, B

Som nämnts tidigare är det inom förorenad mark vanligt att utgå från att ett område är förorenat (jfr Tabell D-3) och nollhypotes och mothypotes definieras enligt: H_0 : Medelhalten av X inom området är lika med eller högre än riktvärdet, respektive H_1 : Medelhalten av X i området är signifikant lägre än riktvärdet för arsenik. Detta innebär, som även indikeras i Tabell D-3, att om *inte* nollhypotesen förkastas så antas området vara förorenat¹⁹. Typ I-felet (α) är förknippat med konsekvenserna av att lämna ett förorenat område utan åtgärd. 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 testet definieras som $1-\beta$ och är sannolikheten att klassa området som rent givet att det är rent (här: sannolikheten att uppmätt medelhalt i området är lägre än riktvärdet givet att verklig medelhalt är det), se Tabell D-3. Liksom typ I-fel, kan styrkan endast beräknas för de fall man har valt att förkasta nollhypotesen. En tumregel från Grandin (2006) säger att styrkan ska vara högre än 80%, men några fastställda nivåer för styrkan finns inte i miljöövervakningssammanhang. Hur stor styrka man bör eftersträva beror istället på de konsekvenser som är kopplade till typ II-fel, se ovan. Grandin (2006) beskriver konceptet med styrka och följande är hämtat därifrån. Den statistiska styrkan är direkt beroende av fyra faktorer. Tre av dessa kan man i viss mån styra över (punkt 1 – 3), medan den fjärde beror av egenskaper hos det man studerar.

1. Signifikansnivå (α). Väljs vanligtvis till 5 procent, men det valet är beroende av konsekvenserna av ett Typ I-fel.
2. Provstorlek, dvs. stickprovets storlek (n). Generellt gäller att ju större stickprov desto högre styrka.
3. Effektstorlek (Δ), eller den minsta förändring eller skillnad man anser vara av vikt att upptäcka. Ju mindre skillnad man vill kunna upptäcka, desto svårare är det. Teoretiskt gäller en nollhypotes bara i de fall då noll- och mothypoteserna är exakt lika, vilket sällan är fallet och i praktiken arbetar man därför med den minsta skillnad som man anser vara av vikt för det man studerar. Om skillnaderna i mothypotesen är mindre än den specificerade effektstorleken behåller man nollhypotesen. Generellt gäller att ju större effektstorlek desto högre styrka, vilket betyder att det är lättare att upptäcka en förändring på 50% än en på 5%. Grandin betecknar effektstorleken också

¹⁹ Inom sötvattensövervakningen som Grandin (2006) utgår från är det vanligt att formulera mothypotesen på det motsatta sättet, H_0 : Alkaliniteten i grundvattenmiljö 3 och typområde F är densamma som för fem år sedan, respektive H_1 : Alkaliniteten i grundvattenmiljö 3 och typområde F är skilljer sig signifikant från nivån för fem år sedan. Detta innebär att om *inte* nollhypotesen förkastas så antas alkaliniteten vara opåverkad.

som ”Minsta Detekterbara Skillnad”. US EPA (2006a) kallar effektstorleken för *gray region*.

4. Variationen hos den studerade variabeln. Om stickprovet uppvisar en låg variation är det lättare att upptäcka en skillnad än om stickprovet har en stor variation. Detta innebär att det är viktigt att dela in ett förorenat område i delområden för att avgränsa olika populationer och därmed begränsa variationen i data. Hypotestester får därefter göras separat för varje delområde.

En femte faktor som också påverkar styrkan är vilken typ av statistiskt test som utförs, där vissa statistiska tester leder till högre styrka än andra (StatSoft, 2008).

D.2 Beräkning av styrka och provstorlek för normalfördelad data

För att skatta vilken styrka ett visst antal prover ger för att upptäcka en given effektstorlek från en given halt för normalfördelad data kan följande formel användas för ett ensidigt test enligt Grandin (2006):

$$1 - \beta = P\left(Z < -z_{1-\alpha} / \nu + \frac{\Delta \times \sqrt{n}}{\sigma}\right) \quad (\text{Ekv. D-1})$$

$z_{1-\alpha}$ = z-koefficienten för den standardiserade normalfördelningen för $1 - \alpha$

α = vald signifikansnivå

n = antal prov

σ = standardavvikelsen för det datamaterial som studien baseras på

Δ = effektstorlek i absoluta tal

P = p-värdet som ges av det beräknade värdet av uttrycket efter mindre än-tecknet inom parenteserna i z-fördelningen (standardiserad normalfördelning).

I Excel kan P beräknas med funktionen NORMSFÖRD(*värde*). I svenska versioner av Excel skrivs ekvation D-1 som:

= NORMSFÖRD(-NORMSINV(1 - α)+ Δ *ROT(n)/ σ),
där kursiverad text anger de värden som skall fyllas i.

I ekvation D-1 antar man att den verkliga standardavvikelsen är känd men i de fall där man har en relativt liten mängd data är den typiska situationen att man har en osäker skattning av standardavvikelsen. Det normala sättet att kompensera för detta är att använda sig av Students t-fördelning istället. Då skrivs ekvation D-1 om till:

$$1 - \beta = P\left(T_{n-1} \leq t_{n-1, \alpha} + \frac{\Delta \sqrt{n}}{s}\right) \quad (\text{Ekv. D-2})$$

där

s = den skattade standardavvikelsen från stickprovet

$t_{n-1,\alpha}$ = t-koefficienten för Students t-fördelning för sannolikheten α och $n-1$ frihetsgrader.

P = p-värdet som ges av det beräknade värdet av uttrycket efter mindre än-tecknet inom parentes i t-fördelningen.

I svenska versioner av Excel kan P beräknas med funktionerna TFÖRD(*värde; frihetsgrader; sidor*) och TINV(*sannolikhet; frihetsgrader*). Eftersom TFÖRD bara är definierad för positiva värden och TINV för tvåsidiga test måste ekvation 9-2 formuleras om. I Excel²⁰ kan ekvation D-2 skrivas på följande sätt för ensidigt test:

För positiva värden: = TFÖRD(1-TINV(2* α / v)+ Δ *ROT(n)/ s ; $n-1$; 1)

För negativa värden: = TFÖRD(-TINV(2* α / v)+ Δ *ROT(n)/ s ; $n-1$; 1)

För att istället beräkna det antal prover som behövs för att kunna upptäcka en given förändring, med α risk för ett typ I-fel och $1-\beta$ styrka, används följande formel enligt (Grandin, 2006):

$$n = \frac{\sigma^2 (z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2} \quad (\text{Ekv. D-3})$$

Här är:

σ = standardavvikelsen för den population som studien baseras på, och

$z_{1-\beta}$ = z-koefficienten för den valda styrkan.

Övriga variabler är desamma som i ekvation D-1.

Ekvation D-3 gäller när den verkliga standardavvikelsen är känd men på samma sätt som ovan har man ofta en osäker skattning av standardavvikelsen. Detta måste man kompensera för, vilket kan göras genom att på motsvarande sätt som ovan använda värden från t-fördelningen. Detta ger:

$$n = \frac{s^2 (t_{n-1,\alpha} + t_{n-1,\beta})^2}{\Delta^2} \quad (\text{Ekv. D-4})$$

Variablerna är desamma som i ekvationerna ovan. Om man använder ekvation D-4 så finns n på båda sidor om likhetstecknet och därför måste ekvationen lösas genom iteration. US EPA (2006b) redovisar en ekvation baserad på den standardiserade normalfördelningen men där man kompenserar för osäkerheten i skattningen av standardavvikelsen genom en extra faktor så att iteration undviks:

$$n = \frac{s^2 (z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2} + \frac{z_{1-\alpha}^2}{2} \quad (\text{Ekv. D-5})$$

²⁰ På grund av att t-fördelningens definition i Excel måste man vara mycket noggrann så att inte beräkningarna blir fel.

Ekvation D-4 och D-5 ger likvärdiga resultat men ekvation D-5 är betydligt lättare att använda eftersom den inte kräver någon iteration. Det är viktigt att notera att ju längre från riktvärdet som den verkliga medelhalten är, desto färre prover behövs för att uppnå acceptabla fel-nivåer (α och β). Det omvända gäller även, dvs ju närmre riktvärdet områdets verkliga medelhalt ligger desto fler prov krävs.

Exempel 1.

Man önskar undersöka ett område m.a.p. blyförorening. Det generella riktvärdet för by för mindre känslig markanvändning (MKM) är 300 mg/kg. Om vi ställer kravet att vi vill kunna detektera en effektstorlek på 10% (dvs. 30 mg/kg), med en signifikansnivå på 0,05 och en styrka på 0,80, hur många prover behöver vi då? Standardavvikelsen antas vara 128 mg/kg, data antas vara normalfördelade och ekvation D-5 används för att beräkna antalet prov.

Antalet prov som skulle behövas är 115 stycken.

Om man t.ex. skulle anta en lägre standardavvikelse så minskar antalet prover: $s = 80$ och i övrigt samma värden skulle ge att $n = 46$, dvs. 46 prover skulle behövas.

D.3 Decision Performance Curve

En *Decision Performance Curve* (DPC) är ett grafiskt verktyg som gör det möjligt att visualisera sambandet mellan de två typerna av beslutsfel (α och β) i statistiska hypotestest, effektstorleken (Δ) och antalet prover. För att ta fram kurvan använder man sig av möjligheten att beräkna styrkan i ett test som beskrivits ovan.

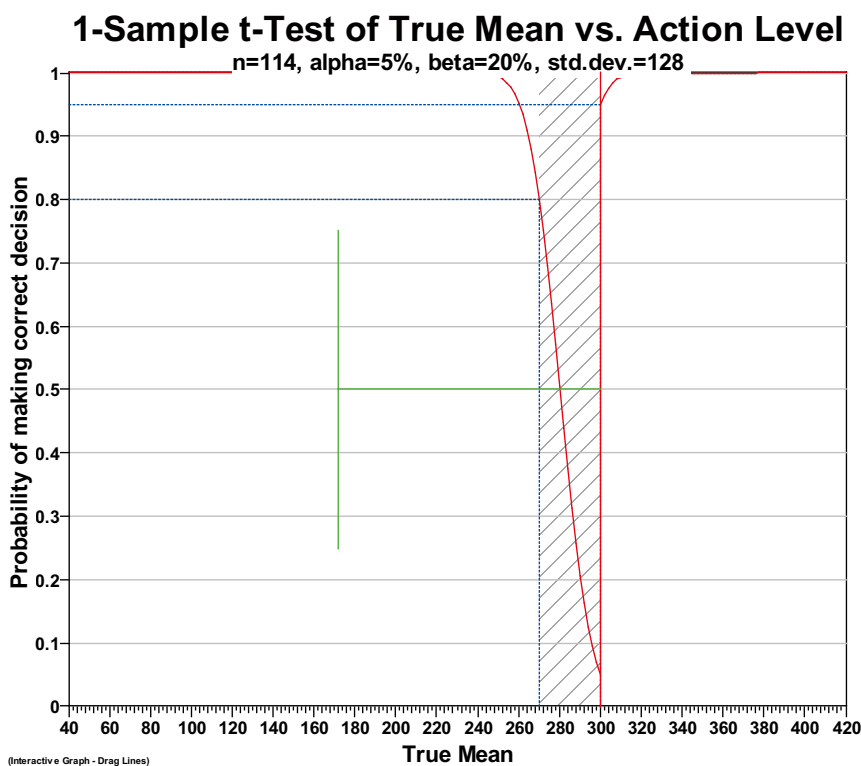
Kurvan (se Figur D-1) kan användas av för att ta fram en provtagningsplan med tillräckligt många provpunkter för att sannolikheterna för de olika typerna av beslutsfel ska bli acceptabla. Den horisontella axeln visar parametrans verkliga värde (vanligen medelvärde för en målpopulation) och den vertikala axeln visar sannolikheten att man ska fatta ett korrekt beslut. Kurvan är typiskt S-formad, med den brantast lutande delen i *gray region*. Detta område är ett intervall som bör väljas så att konsekvenserna av att fatta ett felaktigt beslut blir relativt små.

I VSP används Decision Performance Curves för att planera en provtagning med avseende på antalet prov som behövs för att uppnå tillräcklig säkerhet i slutsatserna. Acceptabla nivåer på α och β bestäms av beslutsfattare och måste relateras till potentiella konsekvenser av de olika felen. Eventuella beslut om att tillåta större osäkerhet måste relateras till konsekvenserna och på motsvarande sätt, om osäkerheterna ska minskas. Vid valet av lämpliga nivåer på α och β bör de miljömässiga konsekvenserna av ett α -fel vägas mot kostnaderna som ett β -fel ger upphov till.

Exempel 2.

Resonemanget omkring exempel 1 kan visualiseras mha. en DPC, se Figur D-1. Figuren är gjord i programvaran Visual Sample Plan (VSP, 2008). Nollhypotesen är formulerad som i exempel 1, att området har en medelhalt som är lika med eller över riktvärdet för bly. Figur D-1 kan tolkas på följande sätt:

1. Om den verkliga (och okända) medelhalten av bly i området är upp till 270 mg/kg, kommer vi på basis av ett ensidigt t-test dra slutsatsen att området är rent, dvs. att nollhypotesen kan förkastas givet att vi tar 114 prover. Sannolikheten att vi har fel när vi drar den slutsatsen är maximalt 20% och minskar ju lägre den verkliga medelhalten är.
2. Om den verkliga medelhalten av bly i området ligger i *grey region*, dvs. mellan 270 och 300 mg/kg så kommer sannolikheten att vi drar fel slutsats på basis av det ensidiga t-testet vara stor, givet våra 114 prover. Testet kommer att visa att nollhypotesen inte kan förkastas, dvs. att medelhalten av bly i området är lika med eller högre än riktvärdet, men β -felet kommer att vara större än 20%. Beta-felet kommer att vara större ju närmre riktvärdet den verkliga medelhalten ligger.
3. Om den verkliga medelhalten av bly i området ligger över 300 mg/kg kommer vi på basis av ett ensidigt t-test och 114 prover dra slutsatsen att medelhalten av bly i området är högre än eller lika med riktvärdet, dvs. att området är förorenat (nollhypotesen kan inte förkastas). Sannolikheten att vi har fel när vi drar den slutsatsen är maximalt 5% och minskar ju högre den verkliga medelhalten är.



Figur D-1. Decision Performance Curve för Exempel 1. På den horisontella axeln visas den verkliga (och okända) medelhalten av bly i mg/kg i området och på den vertikala axeln visas sannolikheten att dra en korrekt slutsats på basis av ett ensidigt t-test, givet en standardavvikelse på 128 mg/kg och ett stickprov med 114 observationer.

D.4 Referenser

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).

Grandin, U., 2006. Statistisk analys av möjligheter att kunna upptäcka regionala trender i de nuvarande programmen för nationell övervakning av sötvatten. Ett underlag för revisionen av programmen år 2006. Rapport Sveriges lantbruksuniversitet, Miljöanalys vol 18. http://publikationer.slu.se/Filer/IMA2006_18.pdf (2008-05-26).

StatSoft, 2008. Electronic textbook. Power Analysis. <http://www.statsoft.com/textbook/stathome.html> (2008-05-26).

VSP, 2008. Visual Sample Plan. Software. Pacific Northwest National Laboratory, PNNL. <http://vsp.pnl.gov/> (2008-05-26)

US EPA, 2006a. Guidance on Systematic Planning Using the Data Quality Objectives Process. EPA QA/G-4. EPA/240/B-06/001, February 2006. <http://www.epa.gov/QUALITY/qs-docs/g4-final.pdf> (2008-05-26)

US EPA, 2006b. Data Quality Assessment: Statistical Methods for Practitioners. EPA QA/G-9S. EPA/240/B-06/003, February 2006. <http://www.epa.gov/QUALITY/qs-docs/g9s-final.pdf> (2008-05-26)

E Antal prov från en lognormalfördelad population baserat på konfidensintervall

E.1 Metod baserad på konfidensintervall för medelvärdet

Enligt Gilbert (1987) kan den ensidiga övre $1-\alpha$ konfidensgränsen $UCL_{1-\alpha}$ beräknas enligt följande (ekvation 13.13 på sidan 170):

$$UCL_{1-\alpha} = \exp\left(\bar{y} + \frac{s_y^2}{2} + \frac{s_y \cdot H_{1-\alpha}}{\sqrt{n-1}}\right) \quad (\text{Ekv. E-1})$$

där \bar{y} och s_y är det aritmetiska medelvärdet respektive standardavvikelsen av n stycken logtransformerade värden $y_i = \ln x_i$. $H_{1-\alpha}$ är H-statistiken enligt Land (1971; 1975); se Gilbert (1987). Vidare gäller följande samband (Strom and Stansbury, 2000):

$$\bar{x} = \exp\left(\bar{y} + \frac{s_y^2}{2}\right) \quad (\text{Ekv. E-2})$$

samt

$$s_y = \sqrt{\ln(CV^2 + 1)} \quad (\text{Ekv. E-3})$$

Vi kan nu applicera (E-2) och (E-3) i (E-1):

$$UCL_{1-\alpha} = \bar{x} \cdot \exp\left(\frac{\sqrt{\ln(CV^2 + 1)} \cdot H_{1-\alpha}}{\sqrt{n-1}}\right) \quad (\text{Ekv. E-4})$$

Kvoten $\frac{UCL_{1-\alpha}}{\bar{x}}$ anger hur många gånger större än medelvärdet som UCL är, dvs. önskad grad av säkerhet i medelvärdet. Vi betecknar denna kvot med D . Det är förhållandevis enkelt att ange ett rimligt värde på D . Exempelvis betyder $D=1,5$ att vi accepterar att uppmätt medelhalt får vara upp till 50% högre än den verkliga medelhalten (med 95% säkerhet om $\alpha=0,05$). Vidare är variationskoefficienten CV möjlig att uppskatta från tidigare genomförda undersökningar. Antalet prov n kan lösas ut från ekvation (E-4):

$$n = \frac{\ln(CV^2 + 1) \cdot H_{1-\alpha}^2}{(\ln D)^2} + 1 \quad (\text{Ekv. E-5})$$

Antalet nödvändiga prov kan nu beräknas med ekvation (E-5), givet att rimliga värden anges för D och CV . Notera att $H_{1-\alpha}$ är en funktion av både α , CV (alltså även σ_y) och n . Detta innebär att ekvation (E-5) måste lösas genom iteration. I avsnitt 3.6.5 samt i Bilaga G redovisas en graf som utgår från ekvation (E-5) och som kan användas för att snabbt beräkna antalet prov.

E.2 Referenser

Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring*. John Wiley & Sons, New York.

Land, C.E., 1971. Confidence intervals for linear functions of the normal mean and variance. *Annals of Mathematical Statistics*, 42:1187-1205.

Land, C.E., 1975. Tables of confidence limits for linear functions of the normal mean and variance, in *Selected Tables in Mathematical Statistics*, vol. III. American Mathematical Society, Providence, R.I., pp. 385-419.

Strom, D.J. and Stansbury, P.S., 2000. Determining Parameters of Lognormal Distributions from Minimal Information. *AIHAJ*, 61:877-880.

F Determining the sample size required for testing the mean of lognormal data

F.1 Determining the Sample Size

In this report I set forth a relatively simple approach for determining the number of samples required to test the mean of a single population of lognormal data. The approach uses well-known transformations of design parameters that allow us to take advantage of normal theory, while still being able to draw conclusions in the natural scale. In what follows, the subscript L (for lognormal) will denote a parameter that is in the natural measurement units, and the subscript N to denote a parameter value that has been transformed to the space where we can use the standard normal theory. Also, “log” denotes the natural logarithm.

Let us assume that the data in their natural scale are lognormal with some mean μ_L and standard deviation σ_L . Hence, the natural logarithm of the data are distributed normally with mean μ_N and standard deviation σ_N . The relationship between (μ_L, σ_L) and (μ_N, σ_N) is well known and discussed by a number of authors (Singh et al 1997, Blackwood 1992, Land 1988). Specifically, if $Y = \log X$ is distributed normally with mean μ_N and standard deviation σ_N , then X is lognormally distributed with parameters μ_N and σ_N . The expected value (mean) and variance of X are given by

$$E(X) \equiv \mu_L = \exp(\mu_N + \sigma_N^2/2) \quad (\text{Ekv. F-1})$$

$$\text{Var}(X) \equiv \sigma_L^2 = \mu_L^2 [\exp(\sigma_N^2) - 1] \quad (\text{Ekv. F-2})$$

and solving and for μ_N and σ_N gives

$$\mu_N = \log \mu_L - \frac{1}{2} \log(\eta^2 + 1) \quad (\text{Ekv. F-3})$$

$$\sigma_N = \sqrt{\log(\eta^2 + 1)} \quad (\text{Ekv. F-4})$$

where $\eta = \sigma_L/\mu_L$ is the coefficient of variation.

Now, suppose we wish to test the null hypothesis that the population mean μ_L is equal to or greater than some action level, or threshold, T_L :

$$H_0 : \mu_L \geq T_L \quad \text{vs.} \quad H_A : \mu_L < T_L \quad (\text{Ekv. F-5})$$

Note that the hypothesis in is formulated in terms of the natural measurement scale. Consequently, if we reject H_0 , we may conclude that the mean contamination level is lower than the threshold (in the original units of measurement). Adhering to the standard approach for creating a suitable

rejection rule for H_0 , the objective is to pick the smallest sample size which satisfies the following criteria:

1. The probability of rejecting H_0 when $\mu_L = T_L$ is no more than α , the false positive, or type I error rate.
2. The probability of rejecting H_0 when $\mu_L = T_L - \Delta_L$ is at least $1 - \beta$, assuming η remains unchanged. Here, Δ_L is the width of the gray region, and β is the false negative, or type II error rate.

In environmental sampling applications, investigators typically wish to demonstrate, with a prescribed level of confidence, that the mean concentration of a certain analyte or contaminant is below a regulatory threshold. For this reason, in this report I will only focus on the one-sided lower alternative hypothesis as shown in . However, this methodology can be readily adapted to one-sided upper and two-sided alternative hypotheses as well.

A standard formula (EPA 2006) for determining the sample size required to test the mean of a normal random variable is given by

$$n^* = \left\lceil \frac{\sigma_N^2 (z_{1-\alpha} + z_{1-\beta})^2}{\Delta_N^2} + 0.5z_{1-\alpha}^2 \right\rceil \quad (\text{Ekv. F-6})$$

where $\lceil \square \rceil$ is the ceiling function (i.e. round up to the nearest integer). Here, z_p is the quantile of the standard normal distribution with lower tail probability p .

Determining the sample size requires that the investigator identify the values of α , β , Δ_N and σ_N^2 . Values of α and β are chosen to achieve the desired error rates, based on the consequences of committing type I and type II errors. The value of Δ_L is chosen so has to have high likelihood of rejecting H_0 if μ_L is Δ_L or more units below T_L . Values of Δ_N can be found by mapping the hypothesized values of μ_L , T_L and $T_L - \Delta_L$, to the normal parameter space using :

$$T_L \Rightarrow T_N = \log T_L - 0.5 \log(\eta^2 + 1) \quad (\text{Ekv. F-7})$$

$$T_L - \Delta_L \Rightarrow T_N - \Delta_N = \log(T_L - \Delta_L) - 0.5 \log(\eta^2 + 1) \quad (\text{Ekv. F-8})$$

and subtracting from gives

$$\Delta_N = \log\left(\frac{T_L}{T_L - \Delta_L}\right) \quad (\text{Ekv. F-9})$$

Estimates of σ_N^2 can be obtained by using applicable historical data or by taking a small pilot sample from the site. Let v_1, v_2, \dots, v_{n_p} represent the sample of n_p measurements from either reliable historical data or a pilot study and let $w_i = \log v_i$ for $i = 1, \dots, n_p$. The preliminary estimate of σ_N^2 is then given by the sample variance of the log transformed measurements:

$$s_w^2 = \frac{1}{n_p - 1} \sum_{i=1}^{n_p} (w_i - \bar{w})^2 \quad \text{where} \quad \bar{w} = \frac{1}{n_p} \sum_{i=1}^{n_p} w_i \quad (\text{Ekv. F-10})$$

To summarize the sample size calculation procedure:

- 1.1. Identify the values for T_L , Δ_L , α and β .
- 1.2. Calculate Δ_N using .
- 1.3. Using , estimate the variance in the normal scale, σ_N^2 , using n_p observations from a pilot study or reliable historical data. I recommend $n_p \geq 10$.
- 1.4. Calculate the sample size n^* using , replacing σ_N^2 with s_w^2 .

If a pilot study was conducted , the investigator may wish to collect only $n^* - n_p$ samples and then combine the pilot study data with the data from the main sample. This approach may be justified if 1) both samples come from the same site and 2) the time elapsed between collecting the two sets of samples is not expected to influence the results.

Because the sample size calculation relies on the assumptions that the data are lognormal and the coefficient of variation is constant, it would be prudent to take additional samples (within the constraints of the project budget) to ensure that the test will have sufficient power in the event that the data mildly violate the assumptions.

F.2 Data analysis: various approaches for testing $H_0 : \mu_L \geq T_L$

Once the data are collected, there are a number of ways to test $H_0 : \mu_L \geq T_L$. One common approach is to calculate the upper confidence limit in the natural units, UCL_L , and then reject H_0 if $UCL_L < T_L$. Singh *et al* (1997) discuss a number of ways to calculate the UCL for lognormal data, as does Gilbert (1987). Krishnamoorthy (2002) sets forth a Monte-Carlo approach for calculating p-values in order to test the mean of a single lognormal population. Chen's test (Chen 1995, EPA 2006, pp 52) also provides a good alternative for testing the mean of a skewed distribution.

In the following, I discuss three possible methods for testing $H_0 : \mu_L \geq T_L$ which are relatively simple to calculate. I did not consider the recommended UCL based on the H statistic (Land 1971, 1975) because Singh *et al* (1997) clearly demonstrated that while the UCL based on the H statistic controls the type I error rate very well, it can be so conservative that it has very little power to reject H_0 when H_A is true.

For each of the three methods, let n denote the total number of measurements that will be used to test $H_0 : \mu_L \geq T_L$. Furthermore, let x_1, x_2, \dots, x_n represent the sample of n measurements, and let $y_i = \log x_i$ for $i=1, \dots, n$.

Method 1: UCL_L based on the lognormal UMVUEs and t -distribution

The uniformly minimum variance unbiased estimators (UMVUEs) of μ_L and its standard error, $SE(\hat{\mu}_L)$ can be used to construct a UCL. Let μ_L and $SE(\hat{\mu}_L)$ denote these UMVUEs, respectively. The formulas for these estimators are given by equations 13.3 and 13.5 of Gilbert (1987). The UCL is then given by

$$UCL_L^t = \hat{\mu}_L + t_{1-k\alpha, n-1} \times SE(\hat{\mu}_L) \quad (\text{Ekv. F-11})$$

Where $t_{1-k\alpha, n-1}$ is the quantile of the t distribution with $n-1$ degrees of freedom and lower tail probability $1 - k\alpha$, with $0 < k \leq 1$. Because the distribution of $\hat{\mu}_L / SE(\hat{\mu}_L)$ is only approximated by a t distribution when $\mu_L = T_L$, a careful choice of k helps ensure that the observed type I error rate is close to the desired rate α . Values of $k < 1$ give rise to larger values of $t_{1-k\alpha, n-1}$, making the UCL less likely to fall below the threshold T_L when the null hypothesis is true—thereby resulting in increased protection against a type I error. Simulation studies (described below) demonstrate that $k = 0.50$ results in fairly good control of the type I error rate over a range of values for α , β , Δ_L , and T_L .

Method 2: UCL_L based on the lognormal UMVUEs and the Vysochanskii-Petunin inequality

This approach is very similar to Method 1, the only difference being that a critical value derived from the Vysochanskii-Petunin inequality (Vysochanskii, 1980) is used instead of a quantile from the t distribution. The inequality only requires that the data arise from a continuous distribution that has a single mode with finite variance. Consequently, it can be described as a non-parametric procedure. The unimodal assumption is reasonable because 1) the lognormal distribution is unimodal, and 2) loosely speaking, even if the underlying data were not lognormal, the distribution of estimates of the mean (such as $\hat{\mu}_L$) are approximately normal, per the central limit theorem. The Vysochanskii-Petunin bound results in smaller critical values than those produced by both the one-sided and two-sided varieties of the Chebyshev inequality. This results in a procedure that is more powerful than a UCL based on the Chebyshev bounds. The UCL is given by

$$UCL_L^v = \hat{\mu}_L + \frac{2}{3\sqrt{\alpha}} SE(\hat{\mu}_L) \quad (\text{Ekv. F-12})$$

The Vysochanskii-Petunin bound is valid so long as $\alpha \leq 1/6$. Simulation studies demonstrate UCL_L^v provides very good protection against type I errors. However, this protection comes at the price of reduced sensitivity (power) to detect the situation where the true mean, μ_L , is indeed less than the threshold T_L .

Method 3: Adaptation of the standard t -test

This approach is very similar to a standard t -test, with some adaptations which will permit drawing conclusions in the natural measurement scale. Using \hat{T}_N and s_y , an estimate of the threshold in the normal space is given by

$$\hat{T}_N = \log T_L - \frac{1}{2} s_y^2 \quad (\text{Ekv. F-13})$$

where s_y^2 is the sample variance of the log transformed measurements, calculated in the same manner as the log transformed sample variance of the pilot (or historical) data as shown in . The adapted t statistic is given by

$$t_a = \frac{\sqrt{n} (\bar{y} - \hat{T}_N)}{s_y} \quad (\text{Ekv. F-14})$$

where \bar{y} is the sample average of the log transformed measurements and $s_y = \sqrt{s_y^2}$. Hence, $H_0 : \mu_L \geq T_L$ is rejected if $t^a < t_{k\alpha, n-1}$, where $t_{k\alpha, n-1}$ is the quantile of the t distribution with $n - 1$ degrees of freedom and lower tail probability $k\alpha$, with $0 < k \leq 1$. As with Method 1, a suitable choice of k helps ensure that the observed type I error rate is close to the desired rate α , since the distribution of t^a is only approximated by the t distribution. Values of $k < 1$ give rise to smaller values of $t_{k\alpha, n-1}$, making t^a less likely to fall below $t_{k\alpha, n-1}$ when the null hypothesis is true—thereby providing increased protection against a type I error. Simulation studies demonstrate that $k = 0.50$ results in fairly good control of the type I error rate over a range of values for α , β , Δ_L and T_L .

F.3 Evaluation of the methods

To evaluate each of the three methods proposed in Section F.3, I performed a simulation study for 486 combinations the design parameters which are shown in Table F-1. Figure F-1 shows the shapes of the lognormal densities for various values of η when $T_L = 20$. (The shapes of the densities for $T_L = 5$ and $T_L = 50$ are virtually identical, just scaled differently).

Table F-1: Values of design parameters.

Parameter	Description	Values
η	Coefficient of variation	0.25, 0.50, 0.75, 1, 1.5, 2
T_L	Threshold in the natural measurement scale	5, 20, 50
$\delta_L = (T_L - \Delta_L) / T_L$	Gray region expressed as a fraction of T_L	0.7, 0.8, 0.9
α	Type I error rate	0.01, 0.05, 0.10
β	Type II error rate	0.05, 0.10, 0.20

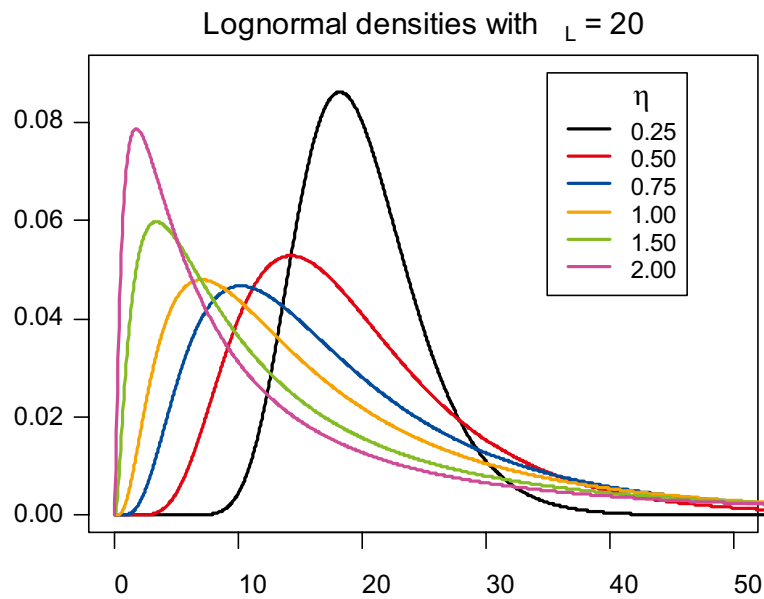


Figure F-1: Lognormal densities for the values of η used in the study

For each of the 486 cases, the performance of ten different tests (based on the three methodologies presented in Section F.2) was evaluated using 100,000 simulated data sets. Methods 1 and 3 were applied to each simulated data set for values of $k = 1.00, 0.90, 0.75,$ and 0.50 . Method 2 was also applied to each simulated data set. Additionally, a control test, which applied Method 3 with $k = 1$ using the known value of σ_N instead of s_y for both the sample size calculation and the calculation of t^a was also applied to each simulated data set. The following procedure was used to estimate the observed type I error rate of the ten tests for each of the 496 cases:

- 3.1 To simulate the pilot study, $n_p = 10$ observations were drawn from a lognormal distribution with $\mu_L = T_L$ and coefficient of variation η . These pilot study data were used to estimate the variance in the normal scale by calculating s_w^2 .
- 3.2 The sample size n^* was then calculated as described in steps 1.1 to 1.4 of Section F.1.
- 3.3 If $n^* > 10$, the main sample of $n^* - 10$ observations were then drawn from a lognormal distribution with $\mu_L = T_L$ and coefficient of variation η . The data from the main sample and the pilot study were combined together. If $n^* \leq 10$, no additional observations were generated and the 10 pilot data observations were used as the main sample.

3.4 The data set of the main sample was analyzed using the ten tests discussed above. For each test, I recorded whether the null hypothesis was rejected or not.

3.5 Steps 3.1 through 3.4 were repeated 100,000 times. The percentage of rejections, $\hat{\alpha}$, was then calculated for each of the ten tests to estimate the false alarm rate.

The procedure to estimate the observed type II error was virtually the same, except $\mu_L = \delta_L T_L \equiv T_L - \Delta_L$ was used in steps 3.1 and 3.3. In step 3.5, the percentage of failures to reject H_0 , $\hat{\beta}$, was tabulated to estimate the false acceptance rate.

Figure F-2 illustrates the difference between the observed and nominal type I error rates, $\hat{\alpha} - \alpha$, for each of the ten tests. The control method performs as expected, with a negligible amount of deviance from 0 due to simulation error. Clearly, Method 2 controls the type I error the best, and in many instances, it provides type I error rates lower than the nominal value. This is not surprising, since the less restrictive assumptions of nonparametric procedures typically make them more conservative than parametric approaches. Method 1 and Method 3 achieve very similar results for each value of k that was considered. Both of these methods appear to achieve a balance between exceeding or falling short of the nominal α when $k = 0.50$. Table F-2 confirms that Methods 1 and 3 have practically identical type I error performance when $k = 0.50$. In addition, for the 486 cases considered, the type I error performance, was, at worst, no more than 0.06 larger than the nominal value of α .

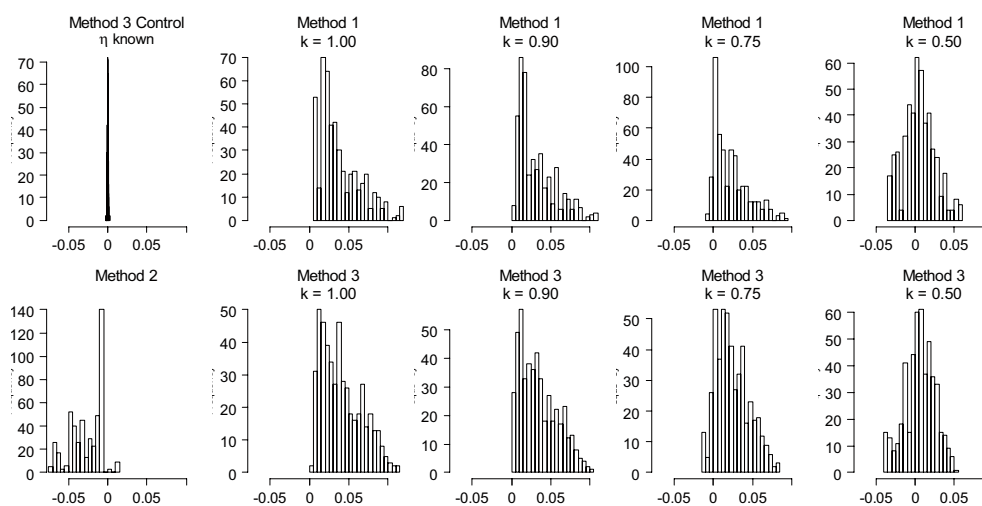


Figure F-2: Histograms of the difference in the observed and nominal type I error rates, $\hat{\alpha} - \alpha$, calculated for each of the 486 cases.

Table F-2: Numerical summary of difference in observed and nominal type I error rates for Methods 1 and 3 when $k = 0.5$.

Method	Summaries of $\hat{\alpha}_i - \alpha_i$ for cases $i = 1, \dots, 486$						$\sum_{i=1}^{486} (\hat{\alpha}_i - \alpha_i)^2$
	Minimum	1 st quartile	Median	Mean	3 rd quartile	Maximum	
1	-0.033	-0.007	0.004	0.005	0.017	0.059	0.202
3	-0.039	-0.005	0.006	0.006	0.019	0.050	0.193

Figure F-3 illustrates the difference between the observed and nominal type II error rates, $\hat{\beta} - \beta$, for each of the ten tests. Once again, the control method performs as expected, with a negligible amount of deviance from 0 due to simulation error. While Method 2 provided strong control of $\hat{\alpha}$, this clearly comes at the price of very poor type II error performance, suggesting that considerably larger sample sizes would be required to improve the power of Method 2. The tighter histograms of Method 3 suggest that it has better power than Method 1 for each value of k that was considered. This conclusion is further corroborated by the numerical summary in Table F-3. While Method 3 has higher type II error rates for $k = 0.50$ than for $k > 0.5$, the larger values of k result in considerably worse type I error rates.

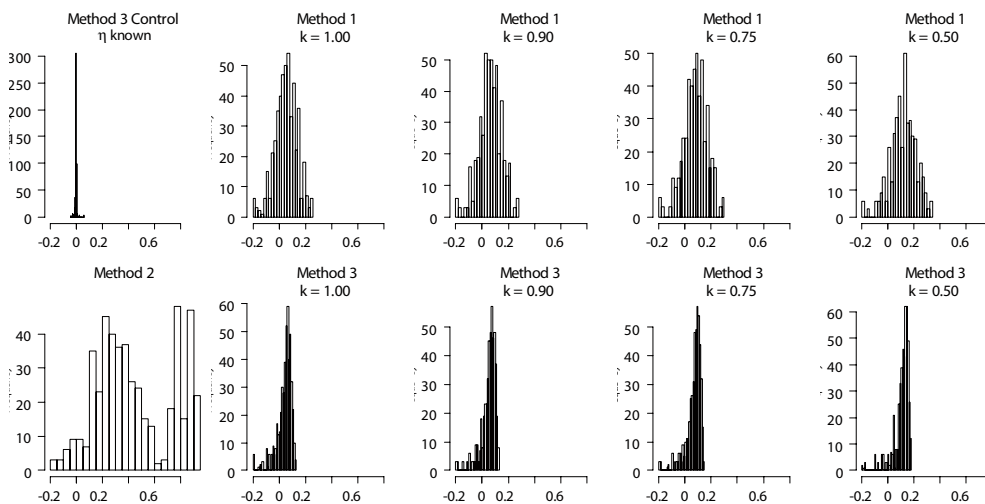


Figure F-3: Histograms of the difference in the observed and nominal type II error rates, $\hat{\beta} - \beta$, calculated for each of the 486 cases.

Table F-3: Numerical summary of difference in observed and nominal type II error rates for Methods 1 and 3 when $k = 0.5$.

Method	Summaries of $\hat{\beta}_i - \beta_i$ for cases $i = 1, \dots, 486$						$\sum_{i=1}^{486} (\hat{\beta}_i - \beta_i)^2$
	Minimum	1 st quartile	Median	Mean	3 rd quartile	Maximum	
1	-0.195	0.062	0.123	0.117	0.183	0.333	11.266
3	-0.191	0.094	0.124	0.106	0.145	0.177	7.293

Note in Figure F-2 and Figure F-3 that the deviations of $\hat{\alpha}$ and $\hat{\beta}$ from α and β result almost entirely from having to estimate the variability from the data. This is clearly demonstrated by comparing the histogram of the Method 3 control approaches (when η is known) to the histograms of the remaining nine tests. Recall that the variability is estimated twice: first to obtain a preliminary estimate of the log transformed variability in order to calculate the sample size, and secondly during the analysis to draw a conclusion about H_0 . Another parallel study I performed (described briefly in Section F.4.2) suggests that the deviance of the observed false positive and false negative error rates from their nominal values is due primarily to the estimation of σ_N^2 or $SE(\hat{\mu}_L)$ during the analysis phase. Said differently, the error in sample size calculation (due to error in the preliminary estimate of σ_N^2) has relatively minimal impact on the observed values of $\hat{\alpha}$ and $\hat{\beta}$.

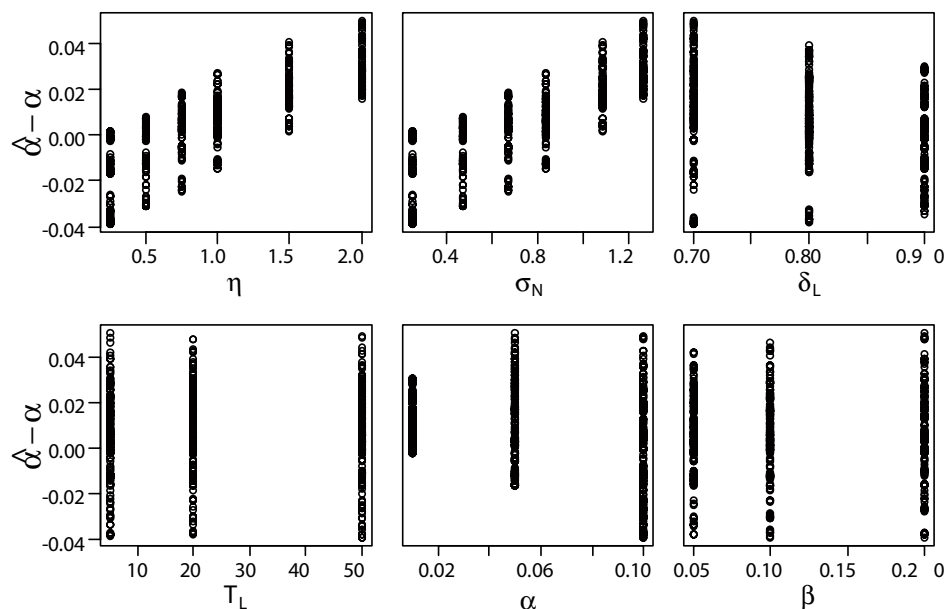


Figure F-4: Plots of the difference in the observed and nominal type I error rates versus the design parameters for Method 3, $k = 0.50$.

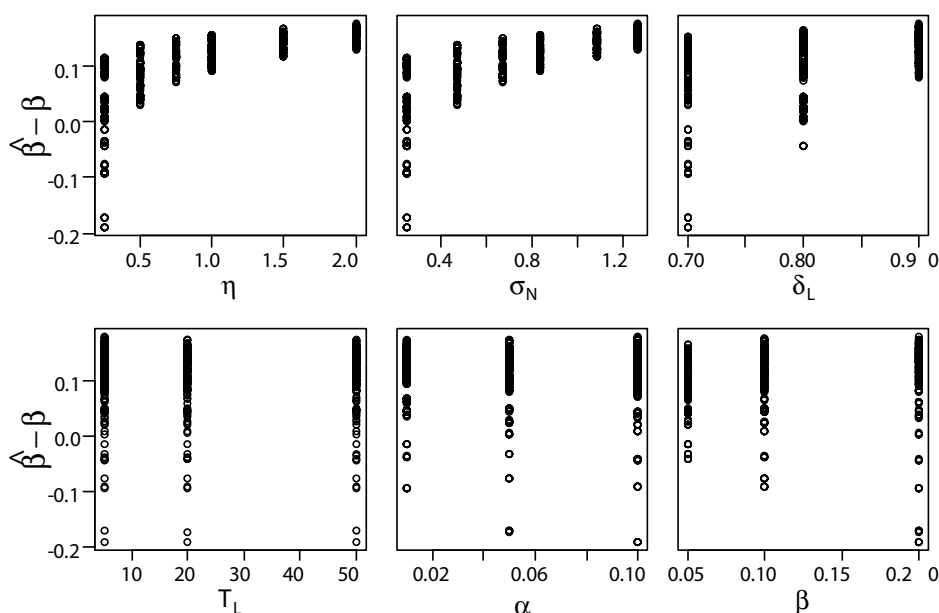


Figure F-5: Plots of the difference in the observed and nominal type II error rates versus the design parameters for Method 3, $k = 0.50$.

The influence of the design parameters on the type I and type II error performance is depicted in Figure F-4 and Figure F-5 for Method 3 when $k = 0.50$. Most notable in Figure F-4 is that differences in the observed and nominal type I error rates are most influenced by the coefficient of variation, η (and naturally σ_N as well, since it is an increasing function of η). Increases in η are directly related to increases in $\hat{\alpha} - \alpha$. The lower, middle panel of suggests that sample designs with small α result in less overall variation in $\hat{\alpha} - \alpha$ than do larger values of α . Figure F-5 demonstrates that large values of η result in values of $\hat{\beta}$ which are increasingly larger than the desired value β . Increasing values of δ_T also result in larger values of $\hat{\beta} - \beta$, but to a lesser extent than increases in η .

F.4 Discussion

F.4.1 REGARDING THE ASSUMPTION OF CONSTANT COEFFICIENT OF VARIANCE

As is typically the case in most sample size calculations, the number of samples is chosen to simultaneously achieve a nominal type I error rate (under a specific null hypothesis) and the nominal type II error rate for a specific, hypothetical alternative hypothesis. For the sample size calculation outlined in Section F.1, the specific null and hypothetical alternative hypotheses are:

$$\begin{aligned}
 H_0 : \mu_L = T_L, \quad \sigma_L = \sigma_0 \\
 H_1 : \mu_L = T_L - \Delta_L, \quad \sigma_L = \left(\frac{T_L - \Delta_L}{T_L} \right) \sigma_0
 \end{aligned}
 \tag{Ekv. F-15}$$

where σ_0 is the population standard deviation (in the natural measurement scale) that we expect when $\mu_L = T_L$.

Note that implies that the coefficient of variation, η , is the same for both hypotheses. Consequently, the approach used to determine the sample size, as well as the observed type II error performance, $\hat{\beta} - \beta$, depend on the assumption that η remains constant, regardless of the hypothesized value of μ_L . Regarding this assumption, Blackwood (1990) notes that Eberhardt and Gilbert (1973) argue that “while means range over many orders of magnitude across different [environmental] data sets, coefficients of variation stay relatively constant.” We can visualize this assumption graphically in Figure F-6, where the joint parameter space of μ_L and σ_L is given by $(0, \infty) \times (0, \infty)$, the upper right quadrant of the two-dimensional plane. The assumption that η is constant restricts this parameter space to a line given by the equation $\sigma_L = \eta\mu_L$.

The restriction of the parameter space is a common technique in sample size determination. For example, in one and two-sample tests of hypothesis regarding the mean of a normal distribution, the assumption is often made that the standard deviation remains constant, which restricts the two-dimensional parameter space of $(-\infty, \infty) \times (0, \infty)$ for μ_N and σ_N to a single horizontal line. In fact, from it is clear that the assumption that η remains constant in the original measurement scale implies that σ_N remains constant when the data are transformed using the natural logarithm.

With regard to the observed type I error performance, the assumption of constant coefficient of variation can be thought of heuristically as follows. We observe log transformed data which arise from a distribution with some unknown population variance, σ_N^2 (which we estimate using s_y^2). We then assess the likelihood that these log transformed data came from a null distribution with mean T_N and the same variance, σ_N^2 . If it appears unlikely that the data have arisen from this null distribution, then we reject H_0 with $(1 - \alpha) \times 100\%$ confidence. We can likewise formulate the heuristic in terms of the data in their natural measurement scale: We observe data which have some unknown coefficient of variation, η . We then assess the likelihood that these data came from a null distribution with mean T_L and the same value of η . If it appears unlikely that the data have arisen from the null distribution, then we reject H_0 with $(1 - \alpha) \times 100\%$ confidence.

Each of the three methods presented in section F.2, as well as most, (if not all) of the analysis methods existing in the literature, rely on the implicit assumption of constant coefficient of variation in the natural scale, which is equivalent to constant variance in the log transformed scale. This naturally begs question (which I leave for additional research), how do these methods perform when the coefficient of variation is not constant?

F.4.2 DEPARTURES FROM LOGNORMALITY

To examine the sensitivity of the procedures discussed in Sections F.1 and F.2 to the assumption of lognormality, I performed a small simulation study wherein the sample size calculation and the analysis were conducted as if the data were lognormal, when in fact, the data were sampled from a family of gamma distributions. I chose the gamma distribution because, like the log-

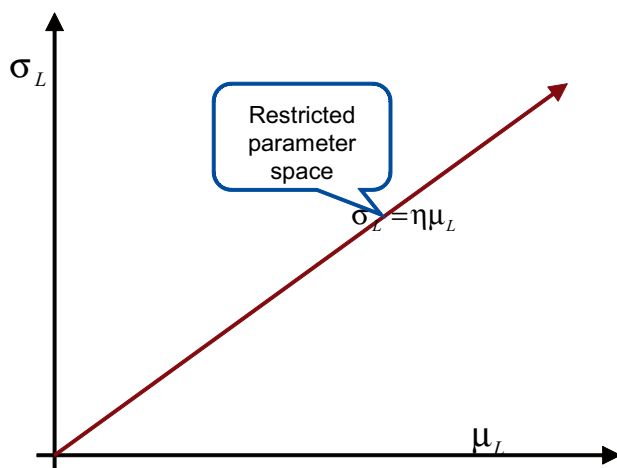


Figure F-6: Parameter space for the mean and standard deviation of the lognormal distribution. The assumption that η remains constant restricts the space to a straight line with slope equal to η .

normal, it has support on the positive real line and its family of densities are somewhat similar in appearance to the family of lognormal densities. To simplify the investigation, I conducted the study assuming that η was known. Consequently, the simulation results shown below *do not* reflect the added uncertainty of having to estimate η (or, in the log transformed scale, σ_N).

For a particular gamma distribution with shape parameter φ and scale parameter γ , I determined the sample size (using the steps outlined in Section 1) required to test $H_0 : \mu_L \geq T_L$ for $T_L = 50$, $\eta = 1/\sqrt{\varphi}$, $\Delta_L = 10$ or 15 , $\alpha = 0.05$ and $\beta = 0.20$. This was done for 30 different gamma distributions, using values of φ in the sequence $0.75, 1.25, \dots, 14.75, 15.25$ and corresponding values of γ equal to T_L/φ , so that each of the gamma distributions had a common mean of 50. (A subset of these gamma densities are shown in Figure F-7). Because the mean of a gamma distribution is $\varphi\gamma$ and the variance is $\varphi\gamma^2$, η is given by $1/\sqrt{\varphi}$.

Having determined the sample size n for each gamma distribution, 100,000 samples of size n were drawn from each gamma distribution under the assumption that the null hypothesis was true, with φ being one of $0.75, 1.25, \dots, 14.75$, or 15.25 , and $\gamma = 50 / \varphi$. The Method 3 control approach (Method 3 using $k = 1$ and the known value of σ_N instead of s_y) was performed for each sample and the number of (incorrect) rejections of H_0 was recorded, giving rise to the observed type I error rate which is plotted in the left panel in Figure F-8. The maximum standard error of the simulated Type I error rate was 0.0016 and the median standard error was 0.00066.

An analogous approach was followed to evaluate the type II error rate for gray regions widths of $\Delta_L = 10$ and 15 . One hundred thousand samples of size n were drawn from gamma distributions under the assumption that $\mu_L = T_L - \Delta_L$ with φ being one of $0.75, 1.25, \dots, 14.75$, or 15.25 , and $\gamma = (50 - \Delta_L) / \varphi$. After drawing each sample, the Method 3 control approach (Method 3 using $k = 1$ and the known value of σ_N instead of s_y) was performed and the

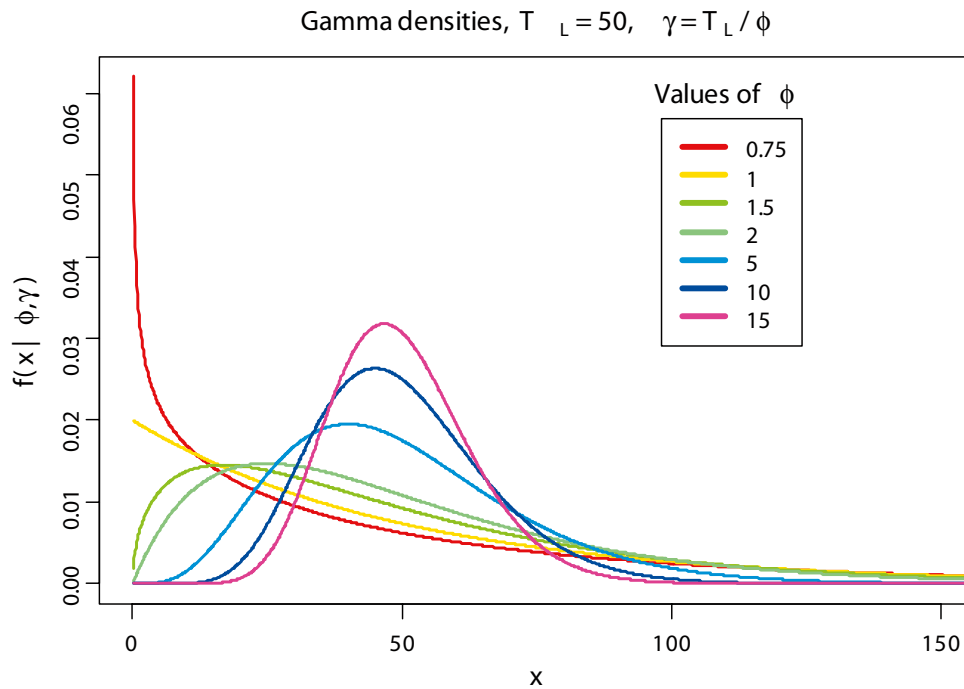


Figure F-7: Gamma densities with shape parameter ϕ and scale parameter. Each of these densities has an expected value (mean) of $\phi\gamma = 50$

number times H_0 was (incorrectly) not rejected was recorded, giving rise to the observed type II error rate which is plotted in the right panel in Figure F-8. The maximum standard error of the simulated Type II error rate was 0.00132 and the median standard error was 0.00126.

The results of the simulation study suggest that the sample size methodology in Section F.1 and the analysis approach of Method 3 are fairly robust to gamma-based departures from the lognormal assumption provided the actual distribution does not have too much weight near zero, as is the case when $\phi < 3$. Note in the left panel of Figure F-8 that the simulated type I error rate achieves (or almost achieves) the nominal value of 0.05 when $\phi > 3$. The right panel in Figure F-8 indicates that, in some instances, the type II error only mildly exceeds the nominal value of 0.20. In fact, for $\phi < 3$, the type II error is considerably lower than 0.20—but that extra power comes at the cost of unacceptably large type I error rates.

This small study suggests that the smaller the width of the gray region, the larger the value of ϕ (and consequently, the smaller the value of η) needs to be in order to achieve the nominal α level. All things being equal, larger values of Δ_L will require smaller sample sizes, which in turn, make it less likely that the t statistic will be well approximated by the t distribution. This may also explain why the blue line ($\Delta_L = 15$) in the right panel of Figure F-8 is more erratic than its red counterpart.

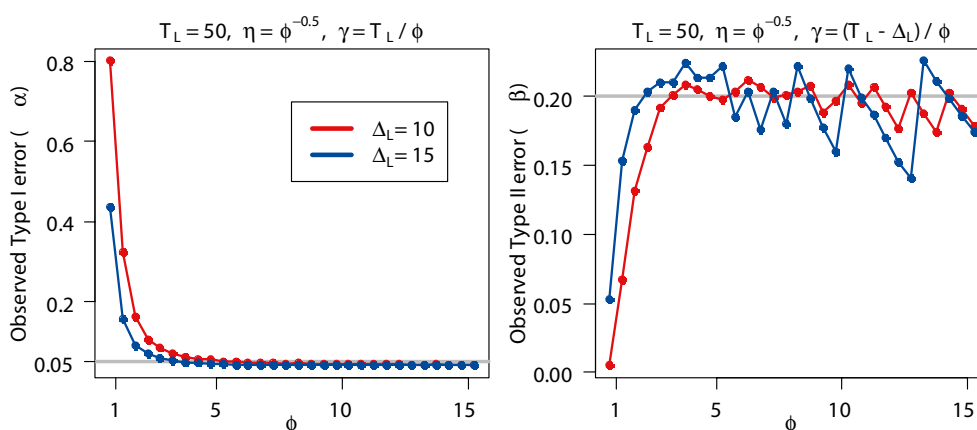


Figure F-8: Simulated type I and type II errors for data sampled from Gamma densities with shape parameter ϕ and scale parameter γ (as indicated in the title of each plot). The nominal values of $\alpha = 0.05$ and $\beta = 0.20$ are indicated by the horizontal gray lines.

F.4.2 SAMPLE SIZE FOR PILOT STUDIES

It should be remembered that sample size calculations are performed with the objective of achieving a desired level of power to reject the null hypothesis when the alternative hypothesis is correct. Consequently, error in calculating the number of samples principally affects the observed type II error rate—but it does not directly affect the observed type I error rate. That is, errors in determining the sample size can either result in diminished power (too few samples) or in excessive power (too many samples). However, since the consequence of a type I error is often considered more severe than the consequence of a type II error, the need for extreme precision in determining the sample size is somewhat mitigated.

The simulation results in Section F.3 use $n_p = 10$, which seemed to be a reasonable number of samples for a pilot study. I also performed something of a duplicate of the study described in Section F.3, the only difference being that no pilot samples were taken and the sample sizes n^* were calculated based on the known value of σ_N . Histograms of $\hat{\alpha} - \alpha$ and $\hat{\beta} - \beta$ from the duplicate study are very similar to Figure F-2 and Figure F-3, with the exception that the distributions are a bit narrower. This suggests that the error in estimating σ_N^2 during the analysis phase has more impact on the observed false positive and false negative error rates than the error in the sample size that arises from estimating σ_N^2 from pilot data. Determining the optimal number of samples for a pilot, n_p , required to reliably estimate the variance of the log transformed measurements in order to calculate the total sample size is an area for further study.

F.4.3 SEQUENTIAL APPROACH

Conducting a pilot study to assist in determining the sample size for the main study leads naturally to the concept of using a sequential analysis to test the hypothesis in . After all, what if, after conducting a pilot study, there was sufficient evidence to reject the null hypothesis? The concepts discussed in this report could, with adaptation and further research, be used in a sequential analysis.

F.5 Practical Recommendations

Investigations should begin by calculating the samples size as discussed in steps 1.1 through 1.4 which appear near the end of Section F.1. I recommend that at least 10 samples be taken when conducting a pilot study. If samples are to be spread over a spatial area, sample locations can be determined with the help of software such as Visual Sample Plan (Matzke 2007), which is freely available. If possible, it is always prudent to take a larger number of samples than indicated by the sample size formula in order to protect against mild violation of the assumptions that the data are lognormal and that the coefficient of variation is constant.

Once the main sample of data are collected, the assumption of lognormality should be evaluated. The standard QQ plot of the log transformed data can be a useful tool in assessing the assumption of lognormality (Figure F-9). If the log transformed data fall largely on a straight line in the QQ plot, they can be considered lognormal. If the data do not appear to be lognormal, another analysis approach should be used.

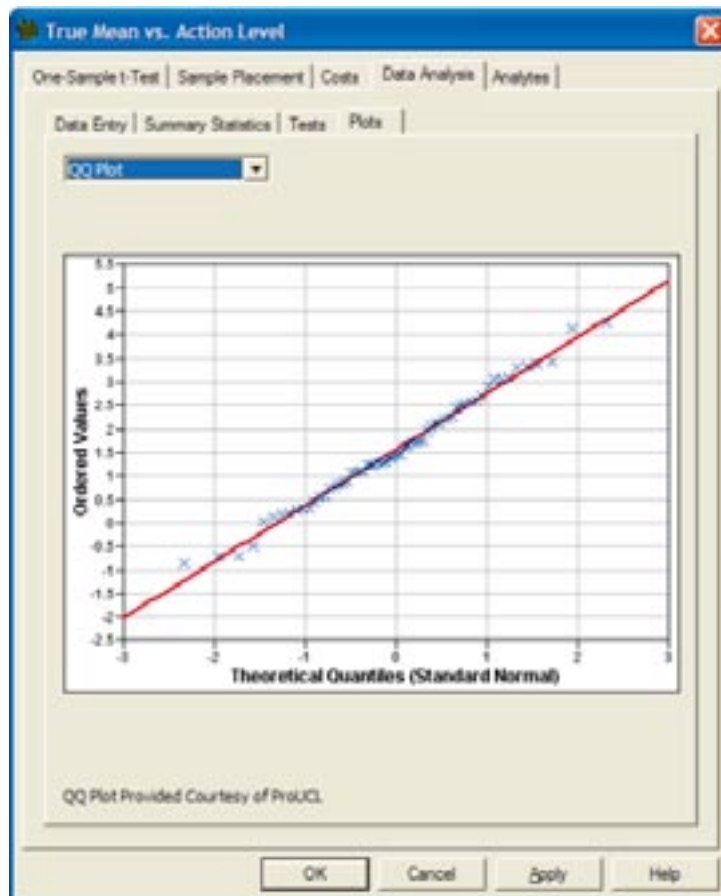


Figure F-9: Assessing the assumption of lognormality using a QQ plot of the log transformed data produced by VSP.

As discussed in Section F.2, there are a number of ways to analyze the data in order to conclude whether the true population mean falls beneath a threshold. I do not recommend UCLs based on the H-statistic nor the Chebyshev approach because they have been demonstrated to be overly conservative, resulting in sharply higher false negative rates. With regard to the three methods proposed in Section F.2, if having tight control of the false positive (type I) rate is the most important objective, then I recommend using Method 2. To implement Method 2, ProUCL (2008), another freely available software tool, can be used to calculate the UMVUEs of μ_L and its standard error. While it is better than the UCLs based on the H-statistic or the Chebyshev bound, Method 2 is still quite conservative, and is much more prone to result in a false negative (type II) error than Methods 1 or 3.

Because Method 3 results in similar type 1 error control as Method 1 but has increased power, I recommend in general that Method 3 be used with $k = 0.5$. Results from Section F.3 demonstrate that over a large number of possible scenarios, the observed type I error rate for Method 3, $k = 0.5$ was never more than 0.05 larger than the nominal rate—and in many instances, it was lower than the nominal rate. If extra control of the type I error is desired, smaller values of k , such as $k = 0.25$ could also be used, probably with minimal impact to the type II error rate.

F.6 References

Blackwood, L.G. (1992) The Lognormal Distribution, Environmental Data, and Radiological Monitoring. *Environmental Monitoring and Assessment* 21:193-210.

Chen, L. (1995) Testing the Mean of Skewed Distributions. *Journal of the American Statistical Association*. 90:767-772.

Eberhardt, L.L. and R.O. Gilbert. (1973) Gamma and Lognormal Distributions as Models in Studying Food-Chain Kinetics. Battelle, Pacific Northwest National Laboratory, BNWL-1747.

EPA 2006. *Data Quality Assessment: Statistical Methods for Practitioners* EPA QA/G-9S, EPA/240/B-06/003, U.S. Environmental Protection Agency, Office of Environmental Information, Washington DC. <http://www.epa.gov/quality/qs-docs/g9s-final.pdf>.

Gilbert, R.O. (1987) *Statistical Methods for Environmental Pollution Monitoring*. Wiley & Sons, Inc., New York, NY.

Krishnamoorthy, K. and T. Mathew. (2003) Inferences on the Means of Lognormal Distributions Using Generalized p-values and Generalized Confidence Intervals. *Journal of Statistical Planning and Inference* 115:103-121.

Land, C.E. (1971) Confidence Intervals for Linear Functions of the Normal Mean and Variance. *Annals of Mathematical Statistics*. 42:1187-1205.

Land, C.E. (1975) “Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance” in *Selected Tables in Mathematical Statistics*, vol III, American Mathematical Society, Providence, R.I., 385-419.

Land, C.E. (1988) Hypothesis Tests and Interval Estimates. In: Crow, E.L. and K. Shimizu (Eds.) *Lognormal Distributions*. Marcel Dekker, Inc., New York, 87-112.

Matzke, B.A., J.E. Wilson, L.L. Nuffer, S.T. Dowson, R.O. Gilbert, N.L. Hassig, J.E. Hathaway, C.J. Murray, L.H. Segó, B.A. Pulsipher, B. Roberts, and S. McKenna. (2007) *Visual Sample Plan Version 5.0 User's Guide*, PNNL-16939, Pacific Northwest National Laboratory, Richland, WA.

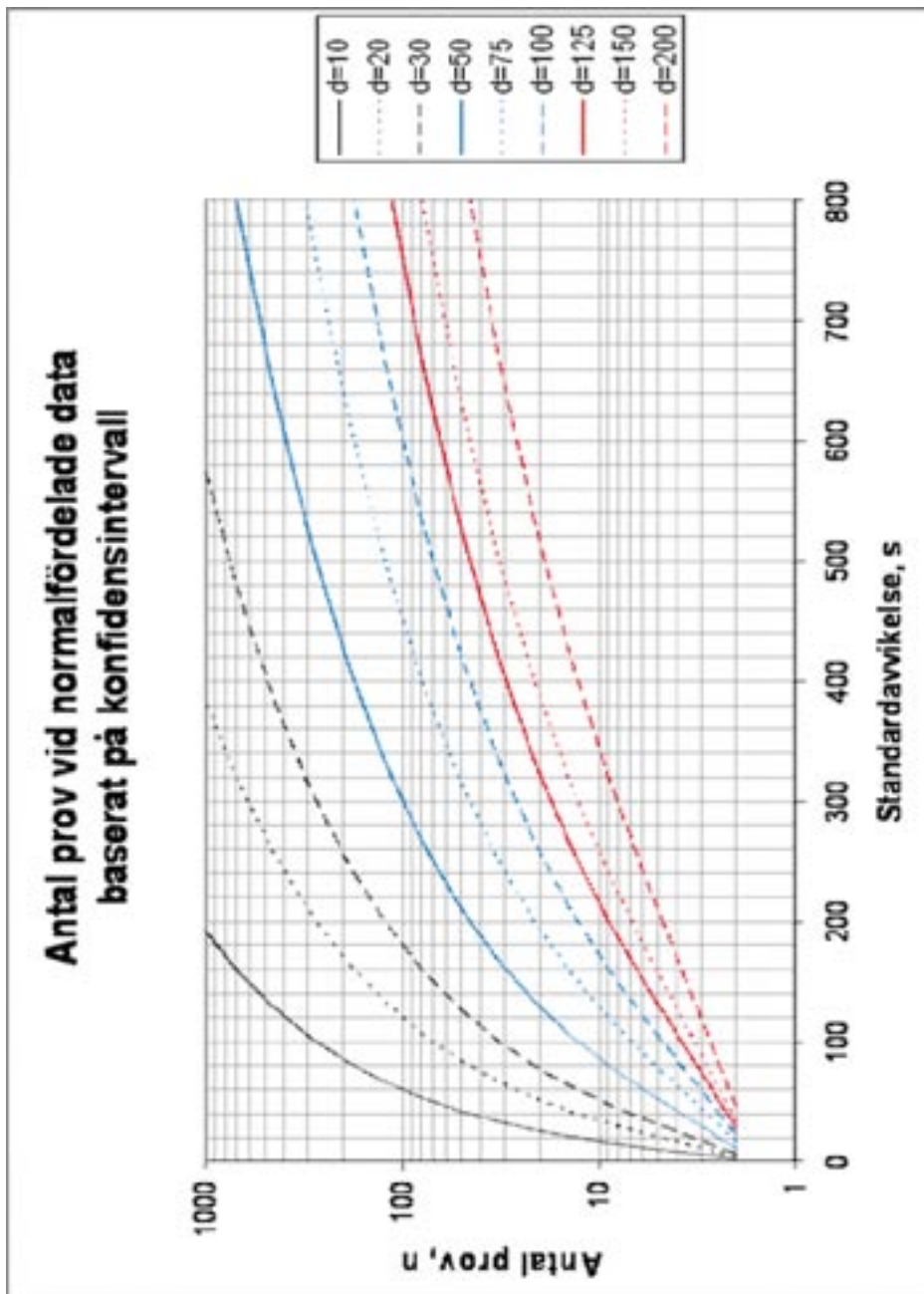
<http://vsp.pnl.gov/docs/pnnl16939.pdf> Also see: <http://vsp.pnl.gov>

ProUCL version 4.00.02 (2008) <http://www.epa.gov/esd/tsc/software.htm> [11 Nov 2008].

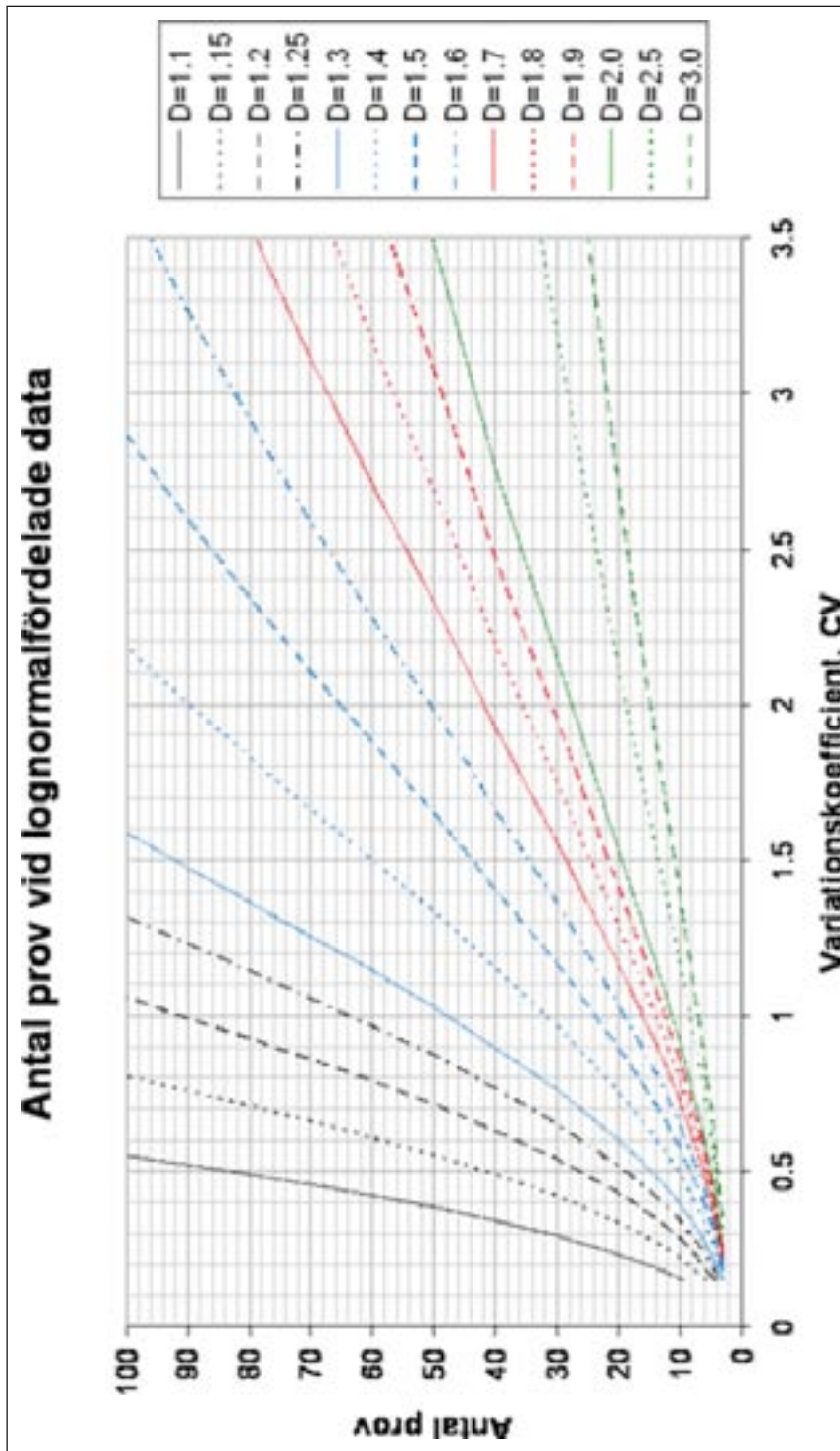
Singh, A.K., A. Singh, and M. Engelhardt. (1997) The Lognormal Distribution in Environmental Applications. EPA/600/R-97/006. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington DC.

Vysochanskii, D.F. and Y.I. Petunin. (1980) Justification of the 3σ rule for unimodal distributions, *Theory of Probability and Mathematical Statistics*. 21:25-36. See also: http://en.wikipedia.org/wiki/Vysochanski%C3%AF-Petunin_inequality [11 Nov 2008].

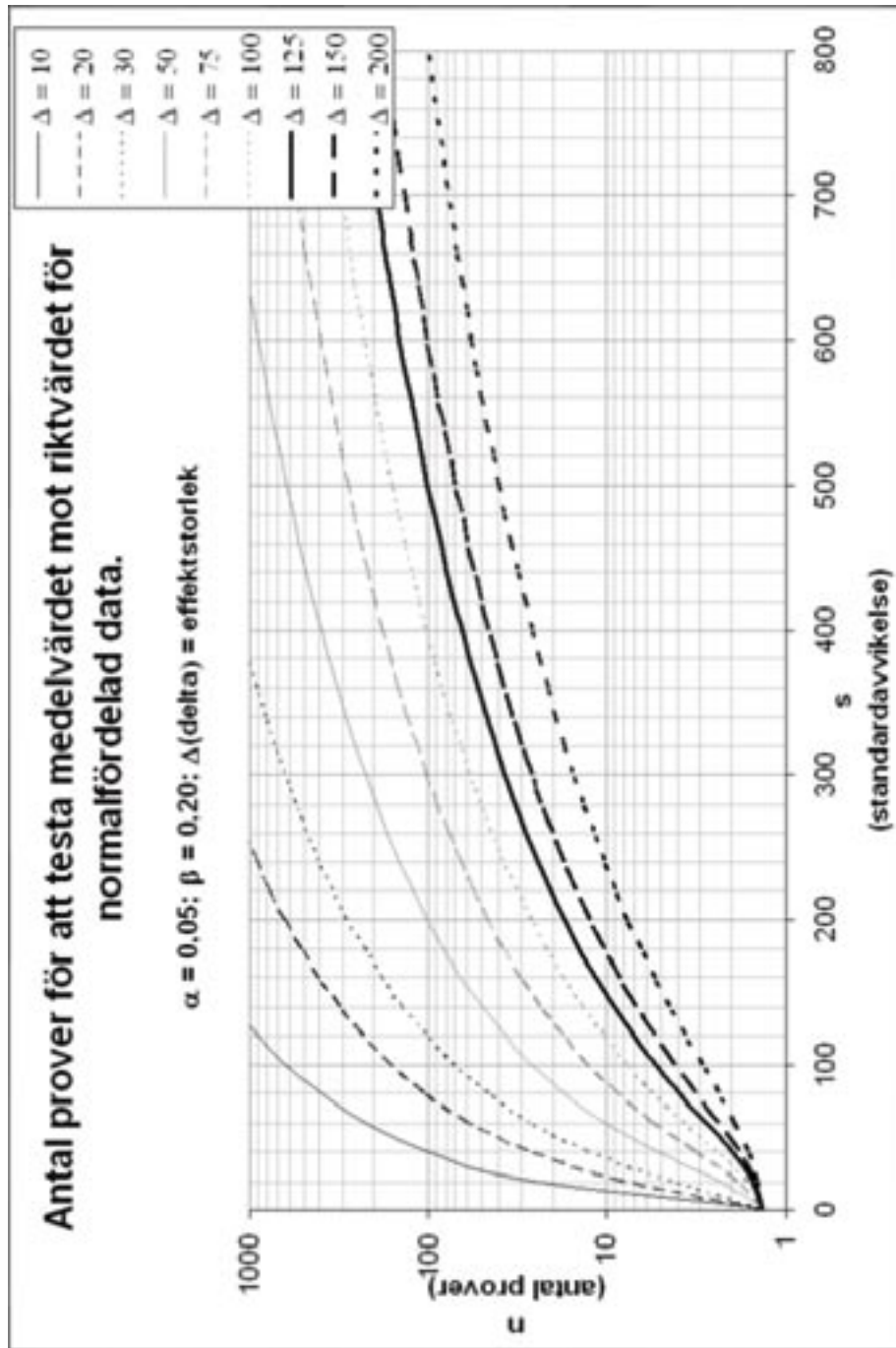
G Hjälpgrafer för att avgöra lämpligt antal prover



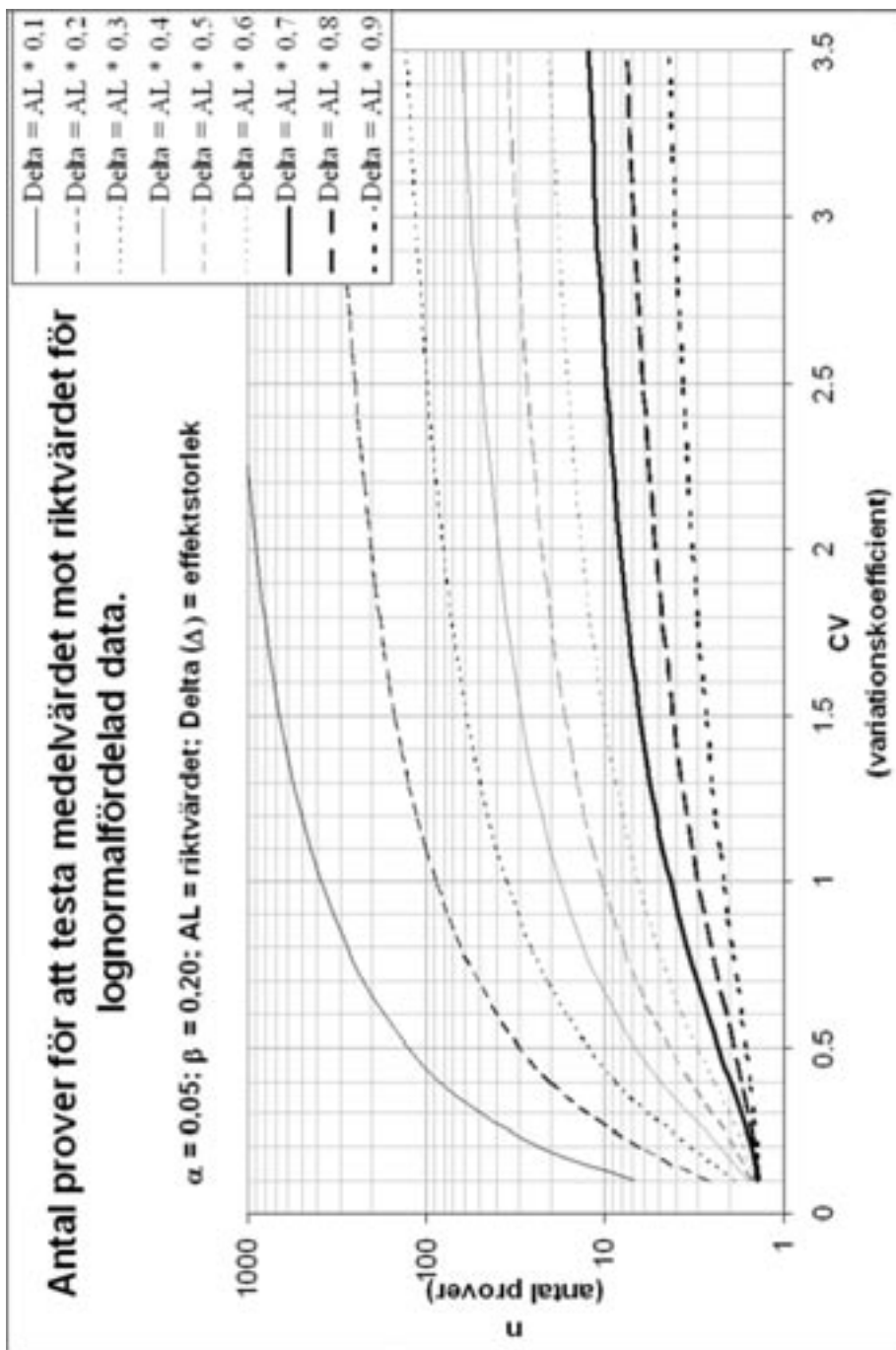
Figur G-1. Graf för att bestämma antal prov med hjälp av konfidensintervall (95% konfidensgrad) om data bedöms vara normalfördelade.



Figur G-2. Graf för att bestämma antal prov med hjälp av konfidensintervall (95% konfidensgrad) om data bedöms vara lognormalfördelade.



Figur G-3. Graf för bestämning av antal prov för att bestämma medelvärdet med önskad grad av säkerhet angivet som acceptabla α - och β -fel samt effektstorlek (Δ) när data bedöms vara normalfördelade.



Figur G-4. Graf för bestämning av antal prov för att bestämma medelvärdet med önskad grad av säkerhet angivet som acceptabla α - och β -fel samt effektstorlek (delta, Δ), när data bedöms vara lognormalfördelade. AL står för engelska Action Level, motsvarande riktvärdet. Effektstorleken (Δ) och riktvärdet (AL) anges i samma enhet, t.ex. mg/kg eller $\mu\text{g}/\text{kg}$. Variationskoefficienten CV är ett mått på föroreningens variabilitet i området.

Provtagningsstrategier för förorenad jord

RAPPORT 5888

NATURVÅRDSVERKET
ISBN 978-91-620-5888-3
ISSN 0282-7298

Rapporten beskriver en metodik för att planera provtagning av jord inom förorenade områden, där syftet av provtagningen är av överordnad karaktär och där man ställer krav på att hantera de stora osäkerheter som föreligger i samband med karaktärisering av förorenad jord. I rapporten beskrivs tre olika angreppssätt för provtagning: (i) sannolikhetsbaserat angreppssätt, (ii) bedömningsbaserat angreppssätt och (iii) sökbaserat angreppssätt. Metodiken riktar sig främst mot projekt där man planerar sin provtagning i förväg, etablerar sig i fält och samlar in data för att sedan utvärdera och analysera data i efterhand.

Rapporten innehåller även en exempelsamling där metodiken tillämpas på modellområden.

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.

