

# Osäkerheter i riskbedömning och beslutsprocess

RAPPORT 5804 • FEBRUARI 2008



Kunskapsprogrammet

**HÅLLBAR  
SANERING**



# Osäkerheter i riskbedömning och beslutsprocess

Peter Starzec (SGI), Tom Purucker och Robert Stewart (University of Tennessee)

**Beställningar**

Ordertel: 08-505 933 40

Orderfax: 08-505 933 99

E-post: [natur@cm.se](mailto:natur@cm.se)

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

Internet: [www.naturvardsverket.se/bokhandeln](http://www.naturvardsverket.se/bokhandeln)

**Naturvårdsverket**

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

E-post: [natur@naturvardsverket.se](mailto:natur@naturvardsverket.se)

Postadress: Naturvårdsverket, SE-106 48 Stockholm

Internet: [www.naturvardsverket.se](http://www.naturvardsverket.se)

ISBN 978- 91-620-5804-3.pdf

ISSN 0282-7298

Elektronisk publikation

© Naturvårdsverket 2008

Tryck: CM Gruppen AB

Omslagsfoto: Ola Arvidslund, SGI

# Förord

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

Den här rapporten redovisar projektet ”Osäkerheter i riskbedömning och beslutsprocess”. I projektet har osäkerheten i olika delar av riskbedömningen studerats, vilken inverkan osäkerheten kan ha på de beslut som fattas och vilka faktorer som påverkar osäkerheten.

Arbetet har utförts av Peter Starzec på Statens geotekniska institut (SGI) samt Tom Purucker och Robert Stewart vid University of Tennessee, USA. Peter Starzec som ledde arbetet, avled hastigt sommaren 2006. Peter blev 40 år och var mycket omtyckt och uppskattad, både som kollega och person. Rapporten har slutredigerats av Jenny Norrman på SGI och kontaktperson för Hållbar Sanering har varit Bo Svensson på Linköpings Universitet

Författarna svarar ensamma för innehåll, slutsatser och eventuella rekommendationer.

Naturvårdsverket februari 2008



# Innehåll

<b>SAMMANFATTNING</b>	<b>7</b>
<b>SUMMARY</b>	<b>9</b>
<b>1 INTRODUKTION</b>	<b>11</b>
<b>2 OSÄKERHETER I RISKBEDÖMNINGEN</b>	<b>13</b>
2.1 Bakgrund och syfte med osäkerhetsstudier	13
2.2 Typ av osäkerheter	14
2.2.1 Konceptuella osäkerheter	14
2.2.2 Modellosäkerheter	15
2.2.3 Osäkerheter i skattning av modellparametrar	15
2.2.4 Osäkerheter i karaktärisering av källan	15
<b>3 FRÅN RISKBEDÖMNING TILL BESLUTSUNDERLAG</b>	<b>17</b>
3.1 Relation mellan riskkriterium och observerade halter	17
3.2 Beslutsfel	18
3.3 Hypotesprövning: teori	20
3.4 Formulering av nollhypotes	22
<b>4 RUMSLIGA ASPEKTER</b>	<b>23</b>
<b>5 OSÄKERHETER I KLASSIFICERING AV FÖRORENINGSGRADEN</b>	<b>25</b>
5.1 Problemställning och förutsättningar	25
5.2 Objektbeskrivning	31
5.3 Insamlade data	31
5.4 Kvantifiering av osäkerhet	31
5.4.1 Zinkförorening	31
5.4.2 Kadmiumförorening	37
5.5 Säkerhet i beslutsunderlag	38
5.6 Slutsatser	39
<b>6 RUMSLIG OSÄKERHET I RISKBEDÖMNING</b>	<b>41</b>
6.1 Problemställning och förutsättningar	41
6.2 Objektbeskrivning	41
6.3 Insamlade data	42
6.4 Interpolering och riskbedömning	43
6.4.1 Interpolering - överblick	43
6.4.2 Riskmodell	44
6.5 Resultat	44
6.5.1 Riskkartor	45
6.5.2 Area of Concern	47
6.6 Slutsatser	54

<b>7 OSÄKERHETER I EKOLOGISK RISKBEDÖMNING</b>	<b>57</b>
7.1 Problemformulering	57
7.2 Tillämpning inom riskbedömning i Sverige	58
7.3 Metodik	59
7.3.1 Kvalitetsfaktor	60
7.3.2 Exponeringsmodell och vandringsmönster	60
7.3.3 Fluorförorening korrelerad med kvalitetsindex	62
7.4 Resultat	64
7.5 Slutsatser	66
<b>8 REKOMMENDATIONER</b>	<b>68</b>
<b>9 REFERENSER</b>	<b>69</b>

# Sammanfattning

Grundreglerna för en kvantitativ riskbedömning av förorenad mark i Sverige innebär att toxikologiskt baserade referensvärden för en given recipient först räknas om till motsvarande koncentration förorening i jord och eventuellt modifieras med hänsyn till bakgrundskoncentration, akuttoxicitet, säkerhetsfaktor m.m. Detta resulterar i ett riktvärde som sedan jämförs med den observerade föroreningshalten i den aktuella jorden.

Det finns flera källor till osäkerhet i bedömningen om risken för påverkan på en recipient är signifikant eller inte. I denna studie har osäkerhetsproblematiken angripits utifrån tre utgångspunkter:

1. *Osäkerheter i förhållandet mellan föroreningsgraden (källan) och uppställt riskkriterium (riktvärde) och dess effekter på efterföljande beslut.*  
Denna typ av osäkerheter anses vara mest betydelsefulla i situationer där förorening förväntas vara heterogent fördelad utan tendens till att bilda sammanhängande och avgränsningsbara *hotspots*.
2. *Osäkerheter i rumslig utbredning av förorening.* Rumslig osäkerhet är mest signifikant vid avgränsning av *hotspots* och vid kvantifiering av förorenade volymer med halt över uppställt riskkriterium.
3. *Osäkerheter i ekotoxikologisk exponeringsmodell med avseende på recipientens vandringsmönster i förhållande till rumslig fördelning av förorening.*

Studien belyser att valet av metod för att karakterisera föroreningssituationen i relation till uppställt riskkriterium styrs av den konceptuella uppfattningen av ämnets rumsliga fördelning och är beroende av strukturen i insamlad data (observationer).

Kvantifiering av osäkerheten i förhållandet mellan föroreningsgraden och riskkriteriet har demonstrerats genom att använda hypotesprövning och konfidensintervall under olika antaganden om datastrukturen. Resultaten betonar vikten av att tillämpa statistiska metoder och illustrerar hur valet av felaktig utvärderingsmetodik påverkar kvaliteten i riskbedömningen och beslutsunderlaget.

Osäkerheten i mängdskattningar av jord med halter som överskrider acceptabel risknivå visar sig vara beroende av vald interpolationsteknik i relativt låg grad för det fall som analyserats i den här studien. Dock förväntas osäkerheten öka signifikant, när ett mer restriktivt riskkriterium (dvs. lägre riktvärde) används. Utöver osäkerheter i mängdskattningen kan osäkerheterna i den exakta positionen av den förorenade volymen vara väsentliga. Demonstrationsexemplet visar att ett riskbaserat underlag till saneringsåtgärder kan variera beroende av antaganden i interpolationsprocessen.

Osäkerheter i ekotoxikologisk exponeringsmodell, med avseende på recipientens vandringsmönster i förhållande till rumslig fördelning av förorening, demonstreras med studier på gaffelantilop (*Antilocapra americana*). Resultaten, som bygger på numeriska simuleringar, visar att bristande kunskap i hur recipienten



förflyttar sig kan resultera i mycket stor osäkerhet vid bestämning av exponeringsdosen. Det presenterade konceptet är mest lämplig för ”mobila” individer som rör sig på stora områden.

I rapporten presenteras statistiska begrepp och metoder. En del får mindre utrymme i texten och läsaren hänvisas i stället till lämplig statistisk litteratur. Det skall dock betonas att syftet med studien och rapporten är att lyfta fram frågeställningar kring osäkerheter vid riskbedömning och demonstrera möjliga konsekvenser av felaktigt tillvägagångssätt snarare än att presentera handfasta råd och rekommendationer. Slutsatserna från projektet kan förhoppningsvis bidra till fortsatt arbete med att etablera nya angreppssätt i vägledning för säkrare riskbedömningar.

# Summary

The general concept for risk assessment in accordance with the Swedish model for contaminated soil implies that the toxicological reference value for a given receptor is first back-calculated to a corresponding concentration of a compound in soil and (if applicable) then modified with respect to e.g. background levels, acute toxicity, and factor of safety. This result in a guideline value that is subsequently compared to the observed concentration levels.

Many sources of uncertainty exist when assessing whether the risk for a receptor is significant or not. In this study, the uncertainty aspects have been addressed from three standpoints:

1. *Uncertainty in the comparison between the level of contamination (source) and a given risk criterion (e.g. a guideline value) and possible implications on subsequent decisions.* This type of uncertainty is considered to be most important in situations where a contaminant is expected to be spatially heterogeneous without any tendency to form isolated clusters (hotspots) that can be easily delineated, i.e. where mean values are appropriate to compare to the risk criterion.
2. *Uncertainty in spatial distribution of a contaminant.* Spatial uncertainty should be accounted for when hotspots are to be delineated and the volume of soil contaminated with levels above a stated decision criterion has to be assessed (quantified).
3. *Uncertainty in an ecological exposure model with regard to the moving pattern of a receptor in relation to spatial distribution of contaminant in question.*

The study points out that the choice of methodology to characterize the relation between contaminant concentration and a pre-defined risk criterion is governed by a conceptual perception of the contaminant's spatial distribution and also depends on the structure of collected data (observations).

How uncertainty in transition from contaminant concentration into risk criterion can be quantified was demonstrated by applying hypothesis tests and the concept of confidence interval under different assumptions regarding the data structure. The results stress the importance to invoke statistical methods and also illustrate how the choice of a wrong methodology may affect the quality of risk assessment and foundations for decision making.

The uncertainty in assessing the volume of contaminated soil was shown to be dependant only to a low extent on the interpolation technique used for the specific case study analyzed. It is, however, expected that the uncertainty may increase significantly, if more restrictive risk criteria (lower guideline value) are applied. Despite a possible low uncertainty in assessing the contaminated soil volume, the uncertainty in its localization can be substantial. Based on the demo example presented, it comes out that the risk-based input for decision on soil treatment may vary depending on what assumptions were adopted during interpolation process.

Uncertainty in an ecological exposure model with regard to the moving pattern of a receptor in relation till spatial distribution of contaminant has been demonstrated by studies on pronghorn (*Antilocapra americana*). The results from numerical simulations show that a lack in knowledge on the receptor moving routes may bring about substantial uncertainty in exposure assessment. The presented concept is mainly applicable for “mobile” receptors on relatively large areas.

A number of statistical definitions/methods/concepts are presented in the report of which some are not elaborated on in detail, while readers are referred to proper literature. The main goal with the study has been rather to shed more light on aspects related to uncertainty in risk assessment and to demonstrate potential consequences of wrong approach than to provide readers with formal guideline and recommendations. However, the outcome from the study will hopefully contribute to the further work on novel approaches towards more reliable risk assessments.

# 1 Introduktion

Riskbedömning är en komplex process som innefattar konceptualisering av riskkedjan och dess representation i matematiska uttryck. Eftersom vår konceptuella uppfattning om styrande transportmekanismer och föroreningens effekter på recipienter ofta är bristande pga. otillräcklig kunskap om föroreningssituationen och transportförhållandena, blir den slutliga riskbedömningen och de framtagna riktvärdena behäftade med betydande osäkerheter. Det är därför angeläget att identifiera de faktorer som har störst inverkan på den totala osäkerheten i riskbedömningen, kommunicera dem till beslutsfattare och ta fram rekommendationer hur osäkra men kvantifierbara resultat bör integreras i beslutsprocessen för beslut om eventuella saneringsåtgärder.

Efter flera års arbete, initierat av Naturvårdsverket, har det tagits fram ett beräkningsprogram för platsspecifika riktvärden och ett vägledningsdokument för riskbedömning (Naturvårdsverket 2005a; 2005c). Vägledningen uppmärksammar betydelsen av osäkerhetskvanter och ett demonstrationsexempel av probabilistisk (sannolikhetsbaserad) osäkerhets- och känslighetsanalys presenteras som bilaga i manualen till beräkningsverktyget.

Inom kunskapsuppbyggnadsprogrammet *Hållbar sanering* har det genomförts och rapporterats en första fas av ett projekt om probabilistisk riskbedömning (Naturvårdsverket 2005b). Studien bygger på inventering och kunskapssammansättning inom fältet men inkluderar också tillämpningsexempel, där effekten av osäkerheter i riskmodellparametrar på osäkerheten i skattning av riktvärde kvantifieras med hjälp av s.k. *Monte Carlo*-simuleringar. Aspekter kring osäkerhets- och känslighetsanalys belyses. Osäkerhetsanalys syftar till att kvantifiera totala osäkerheten i modellresultat som en respons på variationer i indata, medan känslighetsanalysen gör det möjligt att identifiera de parametrar som har störst inverkan på modelleringsresultatet och ger alltså ett mått på graden av inverkan. Därmed kan sådana parametrar ges mer utrymme och exempelvis bli föremål för extra mätningar, medan mindre signifikanta parametrar kan behandlas mer översiktligt.

Enligt det gällande konceptet för riskanalys av förorenad mark i Sverige (och i flera andra länder) är kvantitativ riskbedömning en komparativ process, där man bedömer om föroreningsnivån för ett berört område över- eller understiger framtaget riskkriterium (riktvärde). Ofta riskbedömer man genom att ta fram en så kallad *punktskattning* för observerade föroreningshalter och sedan jämföra den med en annan punktskattning för riktvärdet<sup>1</sup>. Principen är då rentav *deterministisk*, dvs. ingen hänsyn tas till *variabiliteten* i de två kvantiteter som jämförs med varandra. Ofta tillämpas en säkerhetsfaktor, där dock syftet snarare är att fullfölja kraven på försiktighetsprincipen än att beakta osäkerheter. Ett mer realistiskt förfarande är att utgå från att både föroreningshalt och riktvärde är *stokastiska* variabler och som sådana skall beskrivas med probabilistiska funktioner. Detta är dock inte ett

---

<sup>1</sup> Den vedertagna termen "risk assessment" innebär internationellt att utvärdera en sannolikhet för en oönskad händelse, och strikt sett skulle alltså inte en sådan jämförelse som beskrivs här kunna översättas till risk assessment, utan bör snarare översättas till "hazard assessment".

brukligt förfarande, men man bör vara medveten om att riktvärdet innehåller osäkerheter.

En utförlig dokumentation av probabilistiska riskbedömningar har utarbetats av amerikanska miljömyndigheter (US Environmental Protection Agency, US EPA 1997a; 2001) och som också har varit vägledande för strategisk planering och relevanta insatser inom riskbedömning vid motsvarande myndigheter i andra länder. I Sverige har forsknings- samt utbildningsinsatser inom ämnet initierats vid Kalmar högskola och under 2006 genomfördes en andra fas av ovan nämnda projekt om probabilistisk riskbedömning som del av *Hållbar saneringsprogrammet*. Projektet har till syfte att belysa osäkerhetsaspekter vid bestämning av riktvärde och presentera, testa och validera lämpliga verktyg för osäkerhetsanalyser.

Föreliggande rapport fokuserar på hur osäkerheterna i beskrivningen av föroreningskällor skall inkluderas i riskbedömningar. Detta steg kallas *compliance rules* enligt US EPA, dvs. regler för uppfyllande av uppställda miljökrav eller åtgärdskrav, och är avgörande för vilka beslut som är mest rimliga givet underlaget från genomförda undersökningar. I Sverige finns inga tydliga *compliance*-regler, vilket i praktiken kan leda till att olika beslut fattas baserade på ett och samma resultat.

Studien tar upp tre olika frågeställningar, där var och en berör kvantitativ analys av osäkerheter och dess konsekvenser för resultatet vid riskbedömning och kvalitet på underlagsmaterial för beslutsfattande:

1. Hur osäkerheter i beskrivning av föroreningskällan påverkar riskbedömningen och beslutsunderlaget när föroreningen förväntas vara heterogent fördelad inom ett område och observationsdata visar stor variabilitet och svag eller ingen rumslig korrelation. Problemställningen är alltså aktuell för fall där förorening inte föreligger som distinkta *hotspots* och där medelvärdesbetraktelser är relevanta.
2. Hur osäkerheter i rumslig utbredning av en förorening kan påverka exponeringen och risken samt beslut om eventuella saneringsåtgärder. Denna problemställning gäller där endast delar av undersökningsområdena är förorenade och när det finns starka skäl att förutsätta att föroreningsutbredningen är sammanhängande och rumsligt begränsad, dvs. lämpligt för *hotspots*.
3. Hur osäkerheter i ekotoxikologiska exponeringsmodeller beror av recipientens vandringsmönster i förhållande till den rumsliga fördelningen av förorening.

Projektet har genomförts av Statens geotekniska institut (SGI) i samarbete med The Institute for Environmental Modeling vid The University of Tennessee, Knoxville USA.

## 2 Osäkerheter i riskbedömningen

### 2.1 Bakgrund och syfte med osäkerhetsstudier

En riskbedömningsprocess innebär dels insamling av olika typer av information i form av både uppmätta och beskrivande data och dels olika bedömningsmoment, som inte sällan baseras på subjektiva antaganden och hypoteser. Ofta är en stor del av insamlade data och gjorda antaganden behäftade med betydande mått av osäkerhet. Även sammanvägningen av alla dessa data i en modellbeskrivning av riskkedjan innehåller osäkerheter. Vår kunskap om den aktuella föroreningskällan, dvs. dess storlek och föroreningsgraden, transportprocesser mellan källan och recipient och potentiella effekter på recipienten är ofta mycket begränsade pga. otillräcklig mängd och kvalitet av insamlad information i relation till naturens komplexitet. I synnerhet de kvantitativa modeller som tillämpas för riskanalyser inom området förorenad mark bör ses som grova förenklingar av de verkliga processerna och förhållandena. Resultaten bör i de fall där det är motiverat kompletteras med osäkerhetsstudier (US EPA 1989).

Syftet med osäkerhetsstudier är att få förståelse för kritiska moment i riskbedömningen och baserat på denna vetenskap:

- att optimera undersökningsinsatserna (eventuellt komplettera insamlad information) för en säkrare riskbedömning, och
- att ta fram en realistisk bedömning av variationsgraden och dess mönster av risknivå (t.ex. riktvärde) som ett underlag till den efterföljande beslutsprocessen (beslut om eventuella åtgärder).

Innan ett beslut kan tas om lämpliga åtgärder, ställs observerade föroreningshalter som anses vara representativa för hela eller delar av det förorenade området (beroende på förutsättningarna) mot ett uppsatt riskkriterium. Ett av de vanligaste stegen i den riskbedömningsprocess som tillämpas i Sverige, är att jämföra generella eller platsspecifika riktvärden med uppmätta föroreningshalter. I andra riskmodeller (t.ex. US EPA 1989; 1991) utgår man ifrån en minsta uppskattad risknivå (sannolikhet) vid vilken en individ utvecklar cancer under sin livslängd. Risknivån relateras till så kallade *Preliminary Remediation Goals* (PRG), som i princip motsvarar ett gränsvärde utöver vilket negativa hälsoeffekter kan uppstå. Alternativt (för icke cancerogena ämnen) används *hazard quotient*, som är kvoten mellan det kroniska intaget och en kronisk referensdos.

Generellt uttryckt skall resultaten från riskbedömningen ge svar på följande två frågor:

- Om föroreningshalten överskrider det aktuella beslutskriteriet (det kan vara t.ex. risknivå eller riktvärde) eller ej, och hur signifikant både statistiskt och i andra avseenden överskridandet är?

- Om de observerade halterna överskrider det aktuella beslutskriteriet, var inom det undersökta området och hur stora volymer jord handlar det om, samt hur stor är osäkerheten i volymuppskattningen och i avgränsningen av de förorenade områdena?

Problemställningen kring första frågan innebär en kvantitativ analys av förhållandet mellan en eller flera statistiska parametrar, som anses vara representativa för hela haltpopulationen inom undersökningsplatsen, och det framtagna riktvärdet eller risknivån. Haltpopulation är ett hypotetiskt begrepp – dess egenskaper återspeglas av en datapopulation, dvs. ett stickprov. Att använda sig av statistik för att ta fram representativa statistiska mått på föroreningsgraden och jämföra det med ett riktvärde är aktuellt när föroreningen är heterogent fördelad inom området med flera olika punktutsläpp och inget rumsligt beroende mellan uppmätta halter förväntas. I sådana fall ligger den största osäkerheten i en korrekt avspiegling av föroreningsgraden och det som skattats från insamlade data (stickprovet).

Den andra frågan berör riskbedömning med avseende på rumslig fördelning av jord, som bedöms vara förorenad med hänsyn till det uppsatta kriteriet. Här förväntar man till skillnad från första problemställningen att föroreningen bildar sammanhängande avgränsbara kluster (*hotspots*) varvid rumsliga aspekter har mycket större betydelse.

## 2.2 Typ av osäkerheter

Riskbedömning utgör en komplex process. För att hantera all information som ett sådant beräkningsförlopp ska bearbeta, ställs en modell upp där alla dess beståndsdelar, dvs. föroreningskälla(or), transportmedium och recipient, samt samband mellan dem, dvs. fysikaliska, kemiska och biologiska processer, först konceptualiseras. En konceptuell modell beskriver egenskaper av alla ingående komponenter samt förhållanden mellan olika delar i modellen med hjälp av naturlagar. Den konceptuella modellen utvecklas sedan vidare med hjälp av matematiska uttryck och formuleringar. Det finns olika typer av osäkerheter som är kopplade både till den konceptuella och den matematiska modellen beroende på vilken komponent av riskkedjan man beskriver.

### 2.2.1 Konceptuella osäkerheter

Denna typ av osäkerheter avser problemdefinitionen. Att ställa upp en beskrivning av riskkedjan kräver kunskap om vilka föroreningar som förekommer, hur de samverkar, spridningsvägarna mellan föroreningskällan och skyddsobjektet, markanvändningen och vilka aktiviteter som kan förväntas förekomma på området. Alla komponenter i denna kunskap kan vara behäftad med osäkerhet och beror bl.a. på vilken information som finns att utnyttja och hur bra den är. Osäkerheten ökar också när framtida förhållanden ska bedömas. Konceptuell osäkerhet kan både överskatta och underskatta de framtagna riktvärden eller risknivåer och leda till felaktiga beslut. Konceptuella osäkerheter kan inte kvantifieras, men man kan få en

uppfattning om denna genom att jämföra hur slutresultatet påverkas genom att göra parallella beräkningar med alternativa, rimliga och motiverade, konceptuella antaganden. Om betydande skillnader fås med olika antaganden bör en kompletterande undersökning övervägas.

### **2.2.2 Modellosäkerheter**

Kvantitativ riskbedömning kräver att fysikaliska, kemiska och biologiska tillstånd och processer beskrivs med matematiska formler. Formlerna återspeglar sällan den verkliga komplexiteten utan är förenklingar. För att undersöka påverkan från modellosäkerheter kan flera beräkningar göras med alternativa modellsystem med samma antaganden och indata. Det kan gälla hela modellsystemet, då t.ex. utländska riktvärdesmodeller används som alternativ eller delar av modellsystemet, där alternativa transport- och exponeringsmodeller med ökad komplexitet i beskrivningen av processerna används. Resultaten av komplexa modeller är dock inte alltid mer tillförlitliga, eftersom en ökad komplexitet medför nya osäkerheter i skattningen av de ingående parametrarna. Undersökningar av modellosäkerheter är tidsödande och resultaten kan ibland vara svåra att jämföra. I flera situationer är det svårt eller t.o.m. omöjligt att kalibrera modellen, vilket innebär att man inte har någon möjlighet att utvärdera tillförlitligheten i resultatet.

### **2.2.3 Osäkerheter i skattning av modellparametrar**

Den indata som ingår i en beräkningsmodell är ofta osäker på grund av mätfel, informationsbrist, naturlig variation och heterogenitet. Till detta kommer otillräcklig kunskap om styrande processer och mekanismer. Många parametrar uppvisar en stor variation och har därigenom en stor effekt på osäkerheten vid beräkningen av ett riktvärde eller risknivå. Till skillnad från konceptuella osäkerheter och modell-osäkerheter kan dock osäkerheter i modellparametrar utvärderas kvantitativt med hjälp av statistiska metoder.

### **2.2.4 Osäkerheter i karaktärisering av källan**

Karaktärisering av föroreningskälla har till syfte att ta fram ett representativt underlag för uppskattning av halter, rumslig utbredning och volym av den aktuella föroreningen. Om platsundersökningen fallerar kommer efterföljande riskbedömningar att leda till felaktiga beslut beträffande eventuella efterbehandlingsåtgärder. En korrekt genomförd provtagning och en korrekt utvärdering av insamlade data är förutsättningar för att beskrivningen av föroreningssituationen blir representativ. Representativitet innebär ett mått på överensstämmelsen mellan uppmätta resultat och det verkliga miljötillståndet i det undersökta objektet. Detta inkluderar ett mått på graden av noggrannhet och precision, med vilken uppmätta data representerar egenskaper hos populationen och dess variation i rum och tid. Representativiteten innebär att man kan dra slutsatser om föroreningssituationen för hela det undersökta området med osäkerheten skattad från gjorda observationer.

Möjligheten att fatta ett felaktigt beslut med underlag från insamlade data finns alltid eftersom ett stickprov (dvs. observationerna) aldrig helt felfritt återspeglar *target* population (målpopulation eller en hypotetisk population som är



representativ för hela den undersökta volymen). Denna diskrepans orsakas t.ex. av inbördes variabilitet och heterogenitet i tid och rum, provtagningsmönster, antalet prover, mätfel, provhantering och provberedning.

Variabiliteten är ett naturligt tillstånd som inte kan reduceras även om kvalitén i insamlad information ökar. Osäkerheten är ett mått på hur bra variabiliteten kan beskrivas. Osäkerheten kan minskas med mer och bättre data. Ju mindre osäker beskrivning av en föroreningskälla kan göras desto korrektare blir jämförelsen mellan föroreningsgraden (inverkan från källan) och riskkriterium.

## 3 Från riskbedömning till beslutsunderlag

### 3.1 Relation mellan riskkriterium och observerade halter

Resultaten från en riskbedömning bildar ett underlag för beslut om platsens föroreningsstatus i relation till aktuella miljökriterium. Kvalitén på riskanalysen, dvs. dess tydlighet, noggrannhet, precision i resultat och redovisning av osäkerheter påverkar direkt kvalitén på beslutsfattandet och avgör om och vilka åtgärder som skall genomföras. Huvudsakligen leder det till tre beslutsmöjligheter:

- Området (eller delar av det) är förorenat med avseende på det aktuella riskkriteriet. Möjligt beslut: sanering.
- Området är inte förorenat med avseende på det aktuella riskkriteriet. Möjligt beslut: inga åtgärder.
- Signifikansgraden (tillförlitligheten) i resultatet av riskbedömningen är inte tillräcklig för att kunna fatta ett beslut med tillfredställande konfidens. Möjliga beslut: a) mer kunskap krävs, t.ex. i form av kompletterande undersökningar, eller b) det ursprungliga konfidenskravet behöver revideras (sänkas).

I praktiken innebär riskbedömning att man för en given situation, förutsättningar och rådande förhållanden på plats tar fram plats-specifika riktvärden (eller generella riktvärden om så motiverat) och jämför dem med uppmätta halter av föroreningsarna i fråga. Det sättet att riskbedöma är vanligt bl.a. i Sverige, medan man t.ex. i USA i stället utgår ifrån en sannolikhet för risken att överskrida en referensdos. Denna till synes enkla operation ofta är en komplex process eftersom:

- markens heterogenitet och provtagnings- och mätosäkerheter kan medföra att uppmätta halter av en förorening uppvisar stor rumslig (och även tidsrelaterad) variation, vilket gör att den aktuella föroreningsnivån inte kan beskrivas med acceptabel konfidens; samt att
- det framtagna riktvärdet också kan ses som en variabel, eftersom indata till riktvärdesberäkning är behäftade med osäkerheter.

Om två kvantiteter som uppvisar en variation skall jämföras, är det i praktiken omöjligt att avgöra om den ena överstiger den andra utan att utvärdera data med hjälp av statistik slutledning (*inferens*). Det innebär att skillnader mellan observerat och acceptabelt tillstånd kvantifieras som skillnad mellan två sannolikhetsfördelningar.

Enligt ett datakvalitetsmålsdokument sanktionerat av US EPA (Data Quality Objective Process, DQOP: US EPA 1997b; US EPA 2000a) ska grundläggande koncept för utvärdering av observerade halter i relation till uppställda miljökriterier

(t.ex. riskbaserade) läggas fram redan under planeringen av en provtagningsstrategi, se nedan.

- Steg 5 i DQOP: ta fram en beslutsregel. Definiera en lämplig populationsparameter. Bekräfta att det framtagna miljökriteriet (vanligtvis relaterat till haltnivå) överstiger detektionsnivån. Definiera en beslutsregel som beskriver förhållanden för vilka olika åtgärder som skall vidtas om kriteriet överskrids.
- Steg 6 i DQOP: definiera tolerabla beslutsfel. Identifiera potentiella data-osäkerheter och förväntat variationsspann för data, ställ upp en nollhypotes, analysera potentiella konsekvenser av felaktiga beslut; tilldela sannolikheter till värden över och under antaget miljökriterium, dvs. sannolikheter som skall återspegla tolerabla nivåer för potentiella beslutsfel.

Risken att fatta felaktigt beslut med underlag från insamlad data och annan information samt tolkning av resultatet finns alltid eftersom ett stickprov aldrig fullt ut kommer att kunna återspegla den ”sanna” haltvariationen inom hela föroreningsområdet som undersökts.

## 3.2 Beslutsfel

På grund av osäkerheter i insamlad data och vid bestämning av riktvärdet är det ofta otydligt om observerade halter, eller mer exakt de parametrar som beskriver haltpopulationen, överstiger riktvärdet eller inte.

För att kunna göra en riskbedömning måste exponeringen bestämmas. För detta förutsätts att man först kan fastställa om genomsnittshalten av en aktuell förorening är representativ för det undersökta området. För att kunna hantera osäkerheten i bestämningen av genomsnittshalten är det allmänt vedertagna tillvägagångssättet i USA, att ange osäkerheten som det övre enkelsidiga 95%-iga konfidensintervallet ( $UCL_{95}$  - *Upper Confidence Limit*; (US EPA 1996; 2002). Även i Sverige har man uppmärksammat användning av konfidensintervall för medelhalter (Naturvårdsverket 1997). Detta tillämpas dock sällan i riskbedömningsstudier.

Konceptet med UCL tycks vara bättre lämpat för riskmodeller som räknar exponeringen ”framlänges”, som t.ex. i US EPA:s modell, och som därefter jämför den med toxikologiska gränsvärden. Vid ”baklänges” beräkning, som implementerats i den svenska modellen (Naturvårdsverket 2005a), utgår man från ett riktvärde, som motsvarar en referensdos, varefter man räknar sig ”bakåt” till motsvarande halt i marken. I Sverige finns det inga fasta rekommendationer vad som gäller när föroreningsgraden på plats skall jämföras med ett riktvärde: är det  $UCL_{95}$  eller någon annan parameter t.ex. medelhalt, median eller en viss percentil, som skall användas.

Om man utgår från att det är lämpligt att representera den observerade föroreningshalten med t.ex. medel- eller medianvärde för att relatera till ett riskkriterium kan osäkerheter i haltvariation tas med i beräkningen genom att

populationsparametern testas statistiskt för dess över- eller underskridandet av det uppställda kriteriet. Detta kan göras med statistisk hypotesprövning.

Det är enklare att använda UCL i riskbedömning jämfört med hypotesprövning även om källor till osäkerheter ingår vid båda tillvägagångssätten. Det finns inga formella krav på att använda antingen UCL eller hypotesprövning utan vilket tillvägagångssätt som är lämpligast bör avgöras från fall till fall. Generellt kan sägas att en hypotesprövning ger en djupare osäkerhetsanalys och en mer realistisk bild av osäkerheterna i de uppmätta halterna. Detta ger i sin tur större flexibilitet i tolkningen av testresultaten. Att istället beräkna ett UCL för medelvärdet och jämföra med riktvärdet är enklare att genomföra, men ger oftast med ett mer konservativt beslutsunderlag.

I denna rapport har vi medvetet prioriterat att illustrera osäkerheter i beskrivning av föroreningskällan och dess effekt på riskbedömning med hjälp av hypotesprövning och därmed lagt mindre fokus på beräkning av UCL. Det kan emellertid uppstå situationer där en hypotesprövning ger oklara beslutsunderlag och där i stället en tillämpning av UCL bör övervägas.

Vid en hypotesprövning där halter på en plats ställs mot ett beslutskriterium (t.ex. ett riktvärde), är ett av tre utfall möjligt:

- Utifrån det tillgängliga dataunderlaget och med en antagen testkonfidens bedömer man området som rent, medan halterna i verkligheten överstiger det uppställda miljökriteriet. Beslutsfelet (risken) kallas då *alfafel* eller *typ I-fel*.
- Utifrån det tillgängliga dataunderlaget och med en antagen testkonfidens bedömer man området förorenat, medan halterna i verkligheten understiger halterna i det uppställda miljökriteriet. Beslutsfelet (risken) kallas då *betafel* eller *typ II-fel*.
- Om konfidensnivån i testet inte är tillfredställande med hänsyn till acceptansgraden kan inte slutsatser dras med mindre än att dataunderlaget utökas för att uppnå tillfredställande konfidensnivå. Alternativt kan mindre restriktiva *alfa-* eller *betafel* övervägas om konsekvenserna av beslutsfelet bedöms som acceptabla för en given situation.

När en beslutsregel formuleras definierar man en s.k. baspremiss (*baseline condition*) som utgör en praktisk referensnivå för en något abstrakt statistisk hypotesprövning. Baspremissen är en 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 baspremissen är tillräckligt starka blir det alternativet (motsatsen till baspremissen) till en gällande slutsats. Exempel på baspremiss kan vara:

- ”Medelhalten av arsenik inom området är lika med eller överstiger riktvärdet”, eller
- ”Medelhalten av arsenik inom området understiger riktvärdet”

I statistisk terminologi kallas baspremisen för nollhypotes. Om nollhypotesen **förkastas, medan den *de facto* är sann**, begår man ett *alfafel* (typ I-fel). Alternativt, om nollhypotesen **inte förkastas, medan den *de facto* är falsk**, begås ett *betafel* (typ II-fel). I nästföljande kapitel förklaras principer för hypotesprövning.

### 3.3 Hypotesprövning: teori

I detta kapitel återanvänds delar av opublicerat material av Grandin (2003). Vid statistisk hypotesprövning arbetar man med två typer av hypoteser: dels **forskningshypoteser** (baspremiss) och dels **statistiska hypoteser**. En forskningshypotes är ett antagande man vill testa med rigorösa vetenskapliga experiment. Statistiska hypoteser följer, till skillnad från forskningshypoteser, ett strikt mönster. Först ställer man upp en nollhypotes, som alltid formuleras så att den beskriver att det inte finns någon skillnad eller effekt. Sedan formuleras en eller flera mothypoteser eller alternativhypoteser. Nollhypotesen skrivs ofta  $H_0$  och mothypotesen skrivs  $H_1$ . I testet kan nollhypotesen antingen förkastas eller inte förkastas, men aldrig accepteras!

Följande förutsättningar gäller för statistiska tester:

- De data man vill pröva skall vara ett slumpmässigt urval ur den bakomliggande populationen.
- Naturen hos de data man vill pröva avgör vilka tester man kan använda. Normalfördelade data kan analyseras med *parametriska* tester (t.ex. *t-test* (se Davis 1986)), medan data med andra fördelningar måste prövas med *icke-parametriska* (t.ex. *Mann-Whitney test*, se Davis 1986; Levine et al. 2001).
- Den risk man är villig att ta för att förkasta en sann nollhypotes definieras av *signifikansnivån*, eller  $\alpha$  (det maximalt accepterade typ I-felet). Det är den risk man tar att det stickprov man tagit på grund av slumpen visar ett annat mönster än hela populationen och att man därigenom tar ett felaktigt beslut i frågan om att behålla eller förkasta nollhypotesen.
- Vid all hypotesprövning kan man hamna i fyra olika situationer (se Tabell 3.1).

**Tabell 3.1. Möjliga utfall från statistisk hypotesprövning**

		Verkligt förhållande (okänt)	
		$H_0$ sann	$H_0$ falsk
Resultat av statistiskt test	$H_0$ förkastas	Typ I-fel, $\alpha$	Korrekt beslut, $1-\beta$
	$H_0$ förkastas ej	Korrekt beslut, $1-\alpha$	Typ II-fel, $\beta$

- Utifrån stickprovet beräknas en så kallad *testfunktion*. Testfunktionen är en egenskap som fungerar som beslutsunderlag för att bestämma om nollhypotesen skall förkastas eller ej. För varje statistiskt test finns en specifik formel för hur man beräknar testfunktionen. Beroende på vilket

test man utför kommer testfunktionen att ha olika namn: t.ex.  $t$ -,  $\chi^2$ - (chi-två) eller *Mann-Whitney*-test.

- Med utgångspunkt från testfunktionens fördelning formuleras en *beslutsregel*. Beslutsregeln ger det kritiska värde som anger om nollhypotesen ska förkastas eller ej. I statistiska tabeller (se t.ex. Gilbert 1987; Swan and Sandilands 1995) finns olika testfunktioners fördelning för olika signifikansnivåer och för olika stora stickprov. Med hjälp av dessa kan man fastlägga vid vilket värde på testvariabeln nollhypotesen ska förkastas.

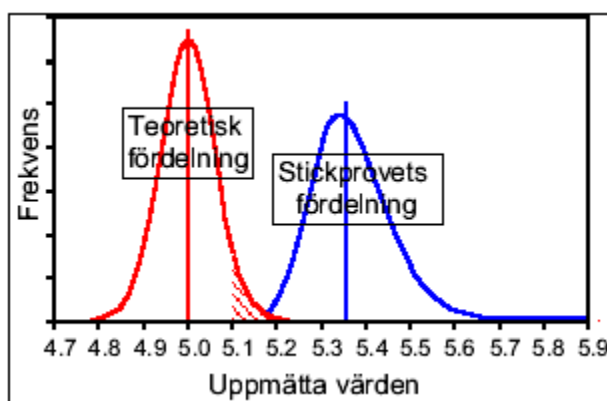
### Exempel 3.1

Hypotesen man oftast ställer upp, när man samlat in ett stickprov (en provtagningsgenomgång), är att stickprovet avviker från ett antaget värde. Det antagna värdet representerar exempelvis en föroreningsnivå motsvarande ett besluts-kriterium (t.ex. ett riktvärde).

Om man har normalfördelad eller approximativt normalfördelad data kan man använda ett *t-test* (Davis 1986). I de fall då stickprovet avviker grovt från normalfördelningen skall så kallade *icke-parametriska* tester användas.

I det här exemplet vill man testa om halter av ämnet A på en nedlagd industri-tomt signifikant avviker från sin naturliga bakgrunds-nivå. Om detta kan bevisas innebär det att de uppmätta halterna kommer från en annan population än de halterna som utgör populationen för naturlig bakgrund, och således kan konstateras att tomten troligen är förorenad.

Som resultat av provtagningen får man av 20 mätningar ett medelvärde på 5,36 mg/kg och en standardavvikelse på 0,27 mg/kg. Detta illustreras av den högra kurvan i figur 3.1.



Figur 3.1. Visuell presentation av principer för hypotesprövning med *t*-test. Sannolikheten att ett slumpmässigt urval från den teoretiska fördelningen skulle ge den fördelning som bygger på stickprov är  $10^{-4}$ . Figuren är hämtad från Grandin (2003).

Bakgrundshalten för A ligger runt 5 mg/kg (medelhalt). Anta vidare att insamlad data är normalfördelad. I ett *t*-test för ett stickprov utnyttjar man stickprovets variation och konstruerar med hjälp av denna variation en fördelningskurva kring det medelvärde som motsvarar en bakgrundshalt på 5 mg/kg (den vänstra kurvan i figur 3.1). Sedan testar man hur stor sannolikheten är att dra ett lika extremt stickprov från bakgrundsfördelningen (den teoretiska fördelningen till vänster) som det stickprov man har samlat in från den nedlagda tomten. Figur 3.1 visar att det är mycket osannolikt att ett slumpmässigt urval från den vänstra kurvan skulle få den fördelning som visas av den högra kurvan men att det dock inte är omöjligt att detta skulle kunna inträffa.

Det skuggade området i den teoretiska fördelningen är 5% av hela arean. I figuren ser det ut som att kurvan slutar vid ca 5,25. I verkligheten fortsätter kurvan oändligt långt åt båda sidor, men samtidigt oändligt nära noll. Om man väljer  $\alpha = 0,05$  anser man att stickprovet är signifikant skilt från 5,00 om stickprovets medelvärde hamnar inom den streckade delen. I detta fall ligger stickprovets medel och fördelning långt ut i ena svansen på den teoretiska fördelningen. Enligt *p*-värdet för ett *t*-test på dessa data kommer ett stickprov från den vänstra kurvan i ett fall av ca 10 000 upprepade dragningar ge ett medelvärde på 5,36, som i den högra kurvan. Vi kan således konstatera att det är ytterst osannolikt att stickprovet kommer från en population som har medelvärdet 5,00. Med andra ord, vi har starka skäl att anse att halterna A överstiger den naturliga bakgrunds-nivån och därmed att tomten är förorenad.

### 3.4 Formulering av nollhypotes

En nollhypotes bör definieras med hänsyn till möjliga konsekvenser av ett felaktigt beslut. En beslutsfattare bör därför vara medveten om vilka konsekvenser som uppstår om t.ex. en hypotes förkastas då den i verkligheten är sann och tvärtom. Man bör sträva efter att formulera nollhypotesen på ett sådant sätt att inga allvarliga konsekvenser uppstår vid ett felaktigt beslut. Som exempel kan risken att dra slutsatsen att ett område inte är förorenat, medan det i verkligheten är det, medföra mycket allvarliga konsekvenser, eftersom potentiella följder av detta felaktiga beslut kan innebära hälsofara, skador på ekosystem, böter, dålig PR m.m. Ett motbeslut, dvs. om man konstaterar att området är förorenat medan det i verkligheten är rent, innebär onödiga saneringskostnader, men samtidigt minimeras eller rentav utesluts eventuella hälso- och miljörisker.

Hur nollhypotesen skall formuleras och hur mycket konfidens i data (signifikansen i resultat från hypotesprövning) som anses vara tillräckligt för ett beslut i en given situation beror på syftet med provtagningen och förutsättningar på plats. Om området är förorenat t.ex. med cancerogena ämnen och platsen avses för bostäder skall nollhypotesen vara att observerade halter är lika med eller överstiger riktvärdet. För att minimera potentiella risker att klassa området som lämpligt för bostäder, medan det i verkligheten kan visa sig olämpligt för det ändamålet, bör *alfafelet* vara så litet som möjligt. Det finns dock inga fasta rekommendationer för tolerabel nivå. Intervallet varierar dock oftast mellan 0,05 och 0,01 (US EPA 1996; Pulsipher et al. 2003).

## 4 Rumsliga aspekter

I många fall blir osäkerheterna i en beskrivning och kvantifiering av föroreningsgraden för ett undersökt område minst lika stora som de osäkerheter som kan relateras explicit till ”riskmekanismer”, dvs. exponeringsparametrar, transport och effekter på recipienten. Det är därför angeläget att välja lämpligt tillvägagångssätt, när uppmätta halter skall jämföras med ett miljökriterium för att fastställa risknivå. Det finns inga rekommendationer rapporterade om hur valet av tillvägagångssätt för att jämföra föroreningshalter med gräns- eller riktvärde beror på observerad datastruktur och variabilitet.

I de fall där ämnen uppvisar ett genomgående heterogent variationsmönster inom hela undersökningsområdet, t.ex. många små punktkällor, eller där föroreningen föreligger i massor som har flyttats uppreparade gånger, är det mindre sannolikt att identifiera en sammanhängande förorenad jordvolym som tydligt kan avgränsas och betraktas som en ”isolerad” *hotspot*. I sådana situationer kan det vara befogat att antingen dela in området i ett antal block och studera varje block för sig, eller att utgå från att hela området är påverkat och kvantifiera hur stor påverkan är i relation till uppställda beslutskriterier. Om det, som i ovanstående fall, inte finns någon indikation på rumsligt beroende (rumslig korrelation eller *spatial correlation*) är detta troligtvis orsakat av hög variabilitet i uppmätt data. En interpolation för att ta fram en karta över föroreningsutbredning kommer i detta fall sannolikt inte att resultera i någon tillförlitlig bild över föroreningskoncentrationen. I sådana situationer är hypotesprövning att föredra. Föroreningsgraden representeras då utifrån stickprovet med en statistisk fördelning eller annan lämplig parameter som jämförs med ett riskkriterium representerat antingen av ett fast riktvärde eller med en sannolikhetsfördelning.

Hypotesprövning utesluter i praktiken att hänsyn tas till rumsliga aspekter eftersom någon fysisk avgränsning av förorenade massor baserad på rumsligt beteende i observerade data ej kan göras. För att undersöka om det finns ett rumsligt beroende kan flera olika metoder användas, bland annat Mantel test, Geary’s C eller Moran’s I kartor (Purucker and Stewart 2004). Ett annat sätt att studera rumsligt beroende är variogramanalys - *pure nugget effect* variogram (se t.ex. Isaaks and Srivastava 1989). Det bör poängteras att avsaknad på bevis för ett rumsligt beroende inte alltid utesluter möjligheten att interpolera data. Under vissa omständigheter kan det finnas starka skäl att avgränsa föroreningen och betrakta den som en sammanhängande ”klump” även om observationerna inte ger någon tydlig indikation på att data är rumsligt korrelerade.

Ett flertal studerade förorenade objekt uppvisar förekomst av *hotspots*, där deras rumsliga utbredning kunnat karteras genom att halter interpolerats utifrån tillgängliga observationer. Detta fenomen är ”typiskt” för t.ex. petroleumprodukter, men det är inte ovanligt att andra ämnen kan forma sammanhängande strukturer som kan beskrivas som en föroreningsplym.

I denna studie har vi gjort en konceptuell uppdelning i två olika förorenings-situationer som används för att belysa osäkerheter vid riskbedömning:



1. Heterogen förorenings-spridning.
2. Sammanhängande *hotspots*.

Kapitel 5 tar upp osäkerheter och konsekvenser vid riskbedömning och efterföljande beslut som kan vara aktuella för den första situationen, medan andra situationen behandlas i kapitel 6. För båda situationerna har en avgränsning gjorts i den mening att begreppet osäkerhet tillämpas utifrån karaktäriseringen av föroreningskällan och den kvantitativa jämförelsen mellan föroreningskällan och det uppställda riskkriteriet (vanligtvis i form av ett riktvärde).

## 5 Osäkerheter i klassificering av föroreningsgraden

I föregående kapitel uppmärksammades hur den konceptuella uppfattningen om en föroreningssituation påverkar möjliga tillvägagångssätt i fråga om utvärderingen av osäkerheterna i beskrivningen av källan och i bedömning av risker. I detta kapitel föreslås ett tillvägagångssätt för att förbereda underlag till beslut, när en förorening förväntas vara heterogent fördelad och observationsdata visar stor variabilitet och svag eller ingen rumslig korrelation. Problemställningen är alltså aktuell när föroreningen inte bildar några *hotspots*. I syfte att underlätta förståelsen av de statistiska metoderna och koncepten i rapporten bygger demonstrationsexemplen på praktiska tillämpningar i en befintlig miljöutredning – en nedlagd industrifastighet, FA1, i centrala Göteborg.

### 5.1 Problemställning och förutsättningar

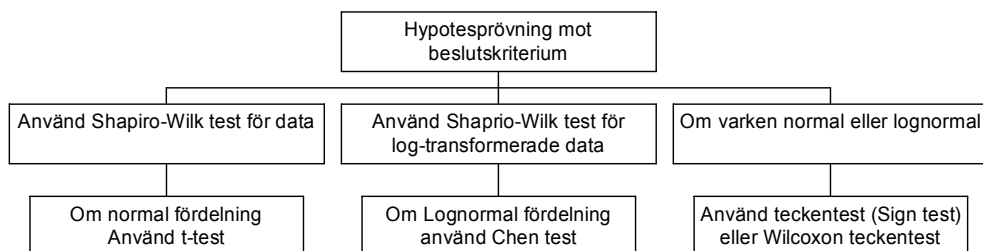
När ämnen uppvisar ett genomgående heterogent variationsmönster inom hela undersökningsområdet, är det mindre sannolikt att hitta en större sammanhängande förorenad jordvolym, som tydligt kan avgränsas och betraktas som en *hotspot*. Detta gäller t.ex. vid förekomst av många små punktkällor eller där föroreningen ligger i massor som upprepande gånger har flyttats om från plats till plats. I sådana situationer kan det vara befogat att antingen dela in område i ett antal delområden och studera varje delområde för sig eller, om området är relativt litet, betrakta hela området som en enhet och kvantifiera föroreningsgraden inom hela enheten. Om en indelning sker i delområden bör man sträva efter att dela in efter förväntade föroreningspopulationer, dvs. en konceptuell modell bör ligga till grund för en indelning.

Oavsett om man delar in platsen i delområden eller ej, är det lämpligt att genomföra hypotesprövning, där man testar data statistiskt, för varje delområde (om så är motiverat) eller inom hela enheten, för överskridande av en uppställd risknivå (riktvärde eller annat gränsvärde). Valet av metod för hypotesprövningen beror på observationernas egenskaper. Med stöd i Box et al. (1978), Davis (1986) och US EPA (2002) konstateras att tre möjliga scenarier kan förekomma och att följande då är att rekommendera:

- Om observerade halter följer en normalfördelning bör s.k. parametriska tester tillämpas, t.ex. *t-test* eller *F-test*.
- Om observerade halter följer en lognormal fördelning används *Chen-test*.
- Om observerade halter varken är normal- eller lognormalfördelade, bör *icke-parametriska* tester användas, t.ex. *Mann-Whitney*, *teckentest* eller *Wilcox-test*.

För att undersöka om data är normalfördelade kan olika test tillämpas. US EPA rekommenderar att *Shapiro-Wilk* test används, vilken även kan tillämpas för att

testa om data följer en lognormal fördelning (US EPA 2000a). Ett enklare sätt att undersöka om data följer en normalfördelning är att studera skevheten i ett plottat histogram och varianskoefficienten (kvoten mellan standardavvikelse och medelvärde). Om data är osymmetriska och variationskoefficienten större än 1 är det troligt att data inte följer en normalfördelning. Figur 5.1 sammanfattar kort vilka steg som bör tas för att identifiera lämpligt tillvägagångssätt för analys av data.



Figur 5.1. Flödesdiagram för tillämpning av lämpliga statistiska procedurer för analys av data för att jämföra en stickprovspopulation med ett fixt värde.

I föreliggande rapport ges inga fasta steg för steg rekommendationer för hur en korrekt hypotesprövning bör genomföras, eller vilka relevanta antagande som görs. Syftet med arbetet är snarare att demonstrera och väcka uppmärksamhet kring möjliga källor till osäkerheter som kan påverka hypotesprövningen och vad detta kan innebära i praktiska sammanhang. Ett hypotetiskt exempel nedan illustrerar att trots att man följer stegen i figur 5.1, bör man ha en djupare förståelse för de analyser man utför för att korrekt dra slutsatser.

### Exempel 5.1

Inom ett område har 25 prover tagits i ett systematiskt randomiserat mönster och analyserats för ämnet  $X$  (se halterna nedan).

Halt mg/kg	120,0	1,6	1,7	5,0	1,0	1,0	1,3	1,1	
	3,1	21,0	23,0	12,8	10,8	2,8	33,0	15,0	24,0
	2,5	1,2	9,1	1,9	2,1	1,8	45,0	2,2	

Med underlag från existerande information kommer  $X$  från luftnedfall och man har konstaterat att det inte fanns några förutsättningar för bildning av distinkta *hotspots*, utan  $X$  förväntas vara spritt över hela området. En visuell bedömning bekräftar att höga halter inte hänger samman utan verkar ligga slumpartat och är separerade från varandra av områden med låga halter. Det finns alltså inga direkta tecken på en rumslik korrelation även om inget formellt test är utfört. Anta att det generella riktvärdet för  $X$  ligger på 15 mg/kg och att frågeställningen är att utifrån en förenklad riskbedömning fastställa om området skall klassas som förorenat. Med hjälp av hypotesprövning vill man testa om uppmätta halter över- eller understiger det uppställda riskkriteriet (15 mg/kg).

#### Utvärdering A

Anta att utvärderingen sker utan att datastrukturen först studerats och att man förutsätter att halterna är normalfördelade och att man för att göra hypotesprövning använder sig av  $t$ -test med signifikansnivå 5%. Detta resulterar i följande:

$H_0$ : medelhalten = 15 mg/kg och  $H_1$ : medelhalten  $\neq$  15 mg/kg ger  $p$ -värde = 0,8

$H_0$ : medelhalten  $\geq$  15 mg/kg och  $H_1$ : medelhalten  $<$  15 mg/kg ger  $p$ -värde = 0,4

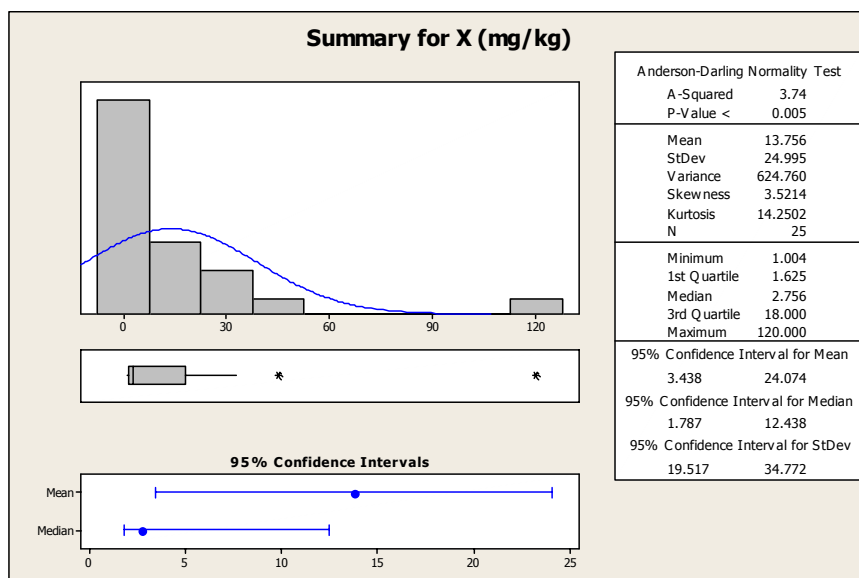
$H_0$ : medelhalten  $\leq$  15 mg/kg och  $H_1$ : medelhalten  $>$  15 mg/kg ger  $p$ -värde = 0,6

Här är  $p$ -värdet risken att ha fel (typ I-fel) om nollhypotesen förkastas.

I inget av de tre fallen är  $p$ -värdet tillräckligt lågt (dvs. lägre än 0,05) för att kunna förkasta nollhypotesen och *de facto* acceptera den alternativa hypotesen. Det lägsta  $p$ -värdet = 0,4 innebär att om vi på basis av data påstår att området är rent så är risken att vi har fel så stor som 40%, vilket är oacceptabelt i de flesta sammanhang. Kontentan är att mer data behövs för att öka konfidensen i hypotesprövningen för att ge ett mer tillförlitligt underlag till beslut.

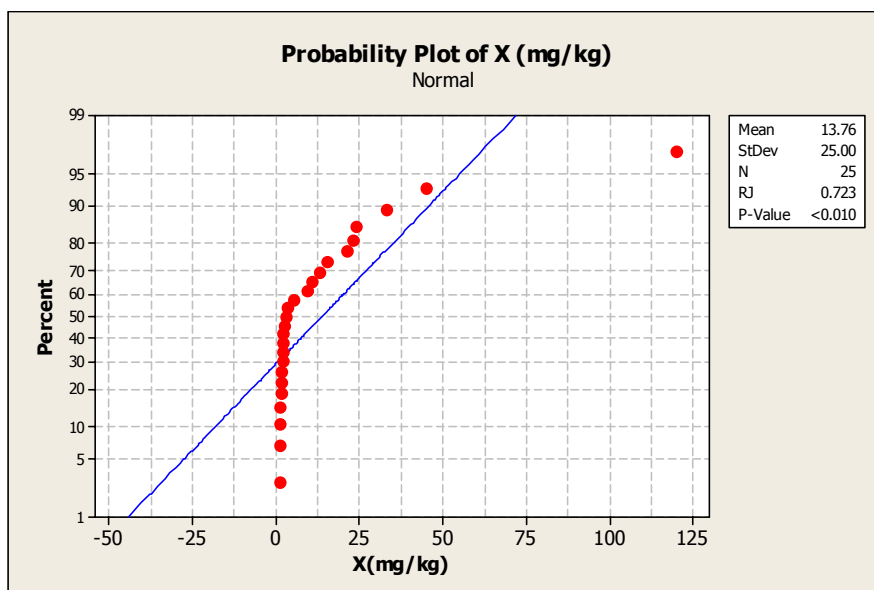
#### Utvärdering B

Anta att utvärdering genomförs enligt stegen specificerade i figur 5.1. En preliminär datautvärdering indikerar hög varians och heterogen fördelning av data (se figur 5.2). Histogrammet i figuren uppvisar en positiv skevhet, dessutom är variationskoefficienten 180% (CV = standardavvikelse/medelhalt  $\approx$  1,8), vilket tyder på att data med stor sannolikhet inte följer normalfördelning.

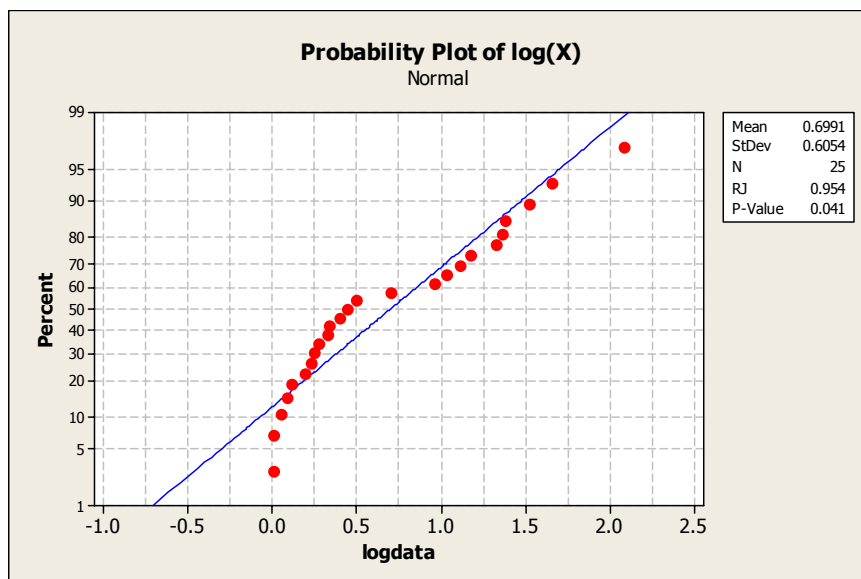


Figur 5.2. Histogram och beskrivande statistik för 25 stickprover av ett hypotetiskt ämne X (halt X i mg/kg).

Resultatet av ett normalitetstest med Rayan-Joiner metoden (Minitab 2004), som är en modifiering av den rekommenderade Shapiro-Wilks metoden, presenteras i figur 5.3. Eftersom ett stort antal datapunkter i figuren avviker från den räta linjen och p-värdet < 0,01, konstateras att data inte är normalfördelade. Med andra ord, risken att vi har fel när vi förkastar nollhypotesen (nollhypotesen är att data är normalfördelade) ligger under 1% - en osäkerhet som i flesta sammanhang kan accepteras. I nästa steg undersöks om data är lognormalfördelade, där testet samt formuleringen av nollhypotesen är samma som för normalitetstest men för logaritmerade data (se figur 5.4).



Figur 5.3. Normalitetstest med Rayan-Joiner för stickprovsfördelning av ämne X. Data visar sig vara icke-normalfördelade (p-värdet < 0,01).



Figur 5.4. Lognormalitetstest för stickprov fördelning av ämne X. Data visar sig icke-lognormala (under en förutsättning att vi accepterar en risk på 4% att ha fel om vi förkastar nollhypotesen).

Lognormalitetstestet resulterar i att risken att vi har fel (typ I-fel) om vi förkastar nollhypotesen är ca 4,1%, d.v.s.  $p$ -värdet = 0,041 vilket är en osäkerhet som kan accepteras i de flesta sammanhang. Om det däremot är så att man har starka skäl att tro att data är lognormalfördelade, så är frågan om man verkligen vågar sig på att förkasta nollhypotesen att data är lognormalfördelade.

Vi vågar inte förkasta nollhypotesen att data är lognormalfördelade utan går vidare och utför ett hypotestest med Chentestet, som rekommenderats i Figur 5.1 för lognormalfördelade data. Chentestet är framtaget för att testa data som har en tydlig skevhet, vilket är fallet här. Chentestet presenteras t ex av US EPA (2006). Detta resulterar i följande om testet utförs med en signifikansnivå på 5%:

$H_0$ : medelhalten = 15 mg/kg och  $H_1$ : medelhalten  $\neq$  15 mg/kg ger  $p$ -värde = 0,90

$H_0$ : medelhalten  $\geq$  15 mg/kg och  $H_1$ : medelhalten  $<$  15 mg/kg ger  $p$ -värde = 0,45

$H_0$ : medelhalten  $\leq$  15 mg/kg och  $H_1$ : medelhalten  $>$  15 mg/kg ger  $p$ -värde = 0,55

Resultatet är liknande det som erhöles för t-testet när vi antog att data var normalfördelade. Här är risken att ha fel ännu högre, 45%, om vi på basis av data påstår att området är rent. Samma resultat här alltså: Mer data behövs för att kunna konstatera huruvida området är rent eller förorenat.

#### *Utvärdering C*

Om vi istället drar slutsatsen att uppmätta halter av X inte är lognormalfördelade heller, bör man enligt Figur 5.1 använda sig av icke-parametriska metoder för vidare hypotesprövning. Både teckentest och Wilcoxon-test kan tillämpas. För principerna bakom dessa tester hänvisas läsaren till statistisk litteratur eller till US EPA (2000a). I korthet kan sägas att i teckentest räknas antalet observationer som överstiger ett jämförelsevärde (riktvärdet i detta specifika fall) och en binomialfördelning används för att räkna fram hur sannolikt det är att antalet överskridanden i en hypotetisk population, som här återspeglas av uppmätta halter X (stickproven), är lika med det antal som man observerat eller större. I Wilcoxon-test utnyttjar man stickprovets avvikelse från ett hypotetisk medianvärde (15 mg/kg).

För teckentest är det lägsta p-värdet man får 0,011 för följande hypotesformulering:

$H_0$ : medianhalten  $\geq$  15 mg/kg och  $H_1$ : medianhalten  $<$  15 mg/kg

För Wilcoxon test är lägsta p-värdet = 0,031 för samma hypotesformulering:

$H_0$ : medianhalten  $\geq$  15 mg/kg och  $H_1$ : medianhalten  $<$  15 mg/kg

Detta indikerar att om vi förkastar nollhypotesen så är risken att vi har fel (typ I-fel) högst 3,1%. Följaktligen finns det starka grunder till att acceptera den alternativa hypotesen. Sålunda kan konstateras att det funnits mer övertygande bevis på att medianhalten av X på området understiger riktvärdet än det motsatta.

#### *Slutsats från utvärdering A, B och C*

Från utvärderingarna A och B verkar vi inte kunna säga så mycket. Vi drar slutsatsen att mer data behövs för att konstatera om området är rent eller förorenat. Hur kan det då komma sig att i fallet med icke-parametrisk hypotestestning (Utvärdering C) så verkar vi kunna säga ganska säkert att området är rent? Här måste man gå tillbaka till vad som egentligen testas. I fallen A och B, jämför vi *medel*halten av ämnet X med riktvärdet. Men i fallet C, jämför vi *median*halten med riktvärdet. I en skev fördelning som denna ligger typiskt medianvärdet (här 2,80 för stickprovet) långt under medelvärdet (här 13,76 för stickprovet), vilket också gäller för den skattade medianhalten och medelhalten. Enligt Gilbert (1987) och US EPA (2000a) kan Wilcoxon och teckentest ge missvisande resultat om data uppvisar en tydligt skev fördelning, vilket ju är fallet med X. Det missvisande resultatet beror alltså på, att för skeva fördelningar så är median och medelvärden helt skilda från varandra, ju skevare fördelning, desto större skillnad. Detta är alltså något man bör se upp med vid användandet av icke-parametriska metoder. Man bör vara uppmärksam på att det i det här fallet är medianhalten som med hög säkerhet understiger riktvärdet och inte medelhalten.

Man bör istället gå vidare och beräkna  $UCL_{95}$  för medelhalten.  $UCL_{95}$  skall också beräknas på basis av datas fördelning. Om data vore normalfördelad blir  $UCL_{95} = 22,3$  mg/kg (Student's t). Om data istället antas vara lognormalfördelade ger Land's metod  $UCL_{95} = 31,6$  mg/kg (se US EPA 2006 för mer information om metoderna, uträkningarna gjorda med freeware ProUCL:

<http://www.epa.gov/esd/tsc/form.htm>). Båda dessa skattningar av  $UCL_{95}$  för medelhalten i området ligger över riktvärdet.

Den övergripande problemformuleringen, som bearbetas vidare i det här kapitlet kan kort sammanfattas som:

*Att demonstrera vilka osäkerheter som kan uppstå, när föroreningshalten inom ett område jämförs med ett antaget riskkriterium (riktvärde), där undersökningsobjektet begränsas till ett relativt litet område och där föroreningen förväntas vara*

*utspridd över hela området och inga sammanhängande och tydligt avgränsade hotspots föreligger.*

Följande sektioner innehåller objektbeskrivning, vilka data och annan information man har tagit hänsyn till, förutsättningar för riskbedömning, osäkerhetsanalys och potentiella effekter på beslutsunderlag.

## 5.2 Objektbeskrivning

Platsen utgör en nedlagd industrifastighet om ca 20 000 m<sup>2</sup>. Olika typer av verksamheter pågick på området från 1950-talet fram till början på 1990-talet. Bland annat finns uppgifter om plåtmanufaktur, skrotgård, blysmältverk och gjutjärns-kross. Genomförda miljöundersökningar visar att olika föroreningar återfinns i marken: metaller (Pb, Cd, Cu, Cr), olja, lösningsmedel och PCB. Markprofilen är huvudsakligen uppbyggd av två jordlager:

- Fyllnadsmassor med inslag av naturliga jordar, byggnadsrester, skrotrester, trä, slagg, avfall m.m. (ca 1-2 m)
- Lera (5-10 m)

Området är plant och gränsar till en väg, två andra industrifastigheter och Göta älv. Grundvattennivån ligger ca 1,5 – 2 m under marknivå. Den framtida markanvändningen bedöms bygga på att marken inte kommer att användas för odlingsändamål. Däremot kan läget göra det aktuellt att i framtid använda marken för bostadsbebyggelse.

## 5.3 Insamlade data

För att utveckla och belysa den givna problemställningen, som formulerades i kapitel 5.1, har man valt att utgå från föroreningarna zink (n = 30) och kadmium (n = 30), för vilka data har insamlats under 1988 – 2005. Ingen hänsyn togs till eventuella osäkerheter i olika provtagnings- och analysförfarande vid olika undersökningskampanjer, dock är majoriteten av provtagningar och analyser gjorda enligt exakt samma procedur. Prover togs enligt ett slumpmässigt systematiskt provtagningsmönster, dvs. området delades in i ett antal lika stora rutor varpå en provgrop med slumpartade koordinater inom varje ruta grävdes, där ett jordprov togs för analys. Ett flertal metaller analyserades och databasen, som användes i denna studie, kommer till största delen från undersökningar gjorda av SGI och Chalmers inom ramen för mastersutbildningen ”Applied Environmental Measurement Techniques” under perioden 2003 – 2005.

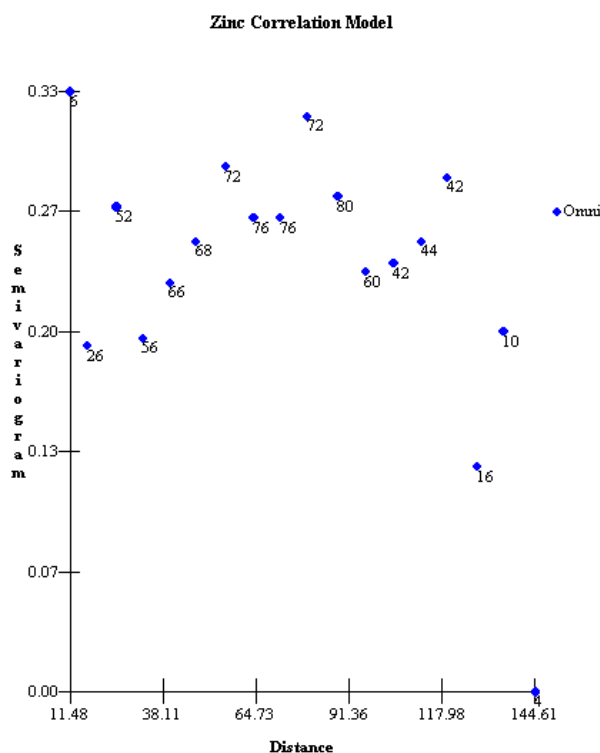
## 5.4 Kvantifiering av osäkerhet

### 5.4.1 Zinkförorening

Granskning av historisk bakgrundsinformation, inklusive information om tidigare aktiviteter på platsen samt en preliminär utvärdering av insamlade data, gav



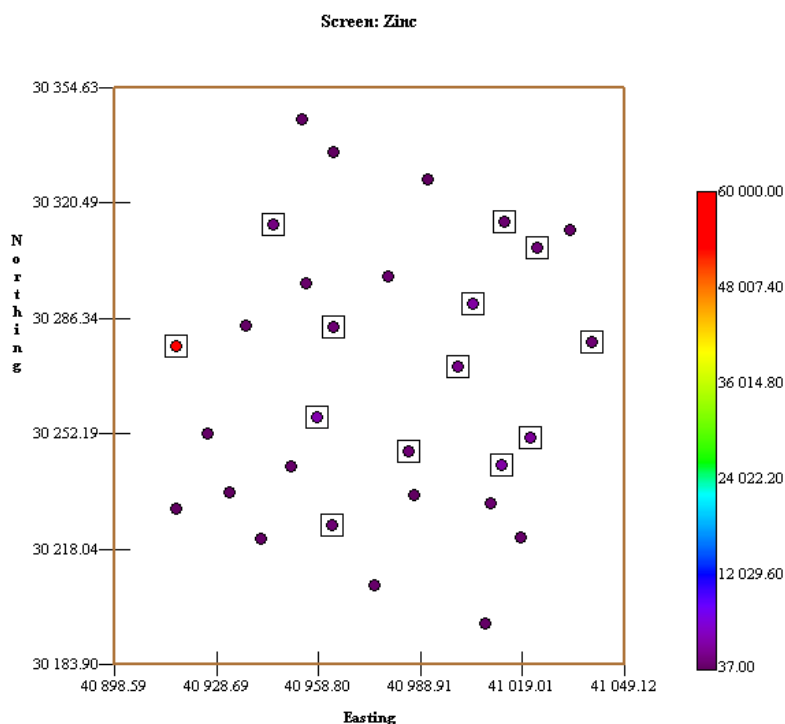
indikation på att Zn är ojämnt fördelat i jordmassorna snarare än att det finns större sammanhängande och avgränsbara *hotspots*. För att verifiera detta gjordes en korrelationsstudie som resulterade i ett *pure nugget effect* variogram, dvs. inget rumsligt beroende kunde konstateras. Variogrammodellering är ingen specifik metod för korrelationsstudier, det kan dock användas som substitut i kombination med annan information. Figur 5.5 visar ett indikatorvariogram för Zn, där indikatorvärdet satts till 700 mg/kg, vilket är det svenska riktvärdet för Zn för MKM, Mindre Känslig Markanvändning (Naturvårdsverket 1996). (Variogrammodellering beskrivs bland annat i Isaaks and Srivastava (1989) och tas inte upp i denna rapport som ett separat ämne.)



Figur 5.5 Indikatorvariogram för Zn. Indikatorvärde har satts till 700 mg/kg, vilket motsvarar svenskt riktvärde för Zn för MKM (Mindre Känslig Markanvändning).

I vissa situationer kan det vara problematiskt att avgöra om och hur föroreningen är rumsligt fördelad och istället är det mest optimala sättet att kvantifiera föroreningshalten i syfte att relatera den till uppställda riskkriterier. Figur 5.6 är ett exempel på ett otydligt budskap, när endast en visuell analys av datastrukturen görs. Figuren är en så kallad *screening map* och visar var inom området uppmätta halter överskrider det generella riktvärdet för zink för MKM. Att direkt från kartan avgöra om det handlar om några sammanhängande *hotspots* eller ett varierande spridningsmönster utan någon tydlig distinkt struktur är svårt. Variogrammodellering, som kan användas för att studera rumsligt beroende, kräver kunskaper inom geostatistik och används sällan. De metoder som kan tillämpas mer explicit för att studera rumslig korrelation och som nämnts i kapitel 4 tillämpas inte heller i någon större

utsträckning. Att lyfta fram betydelsen av en bredare användning av metoder för studier av rumsligt beroende bedöms bidra till mer trovärdiga riskbedömningar och bättre beslutsunderlag.

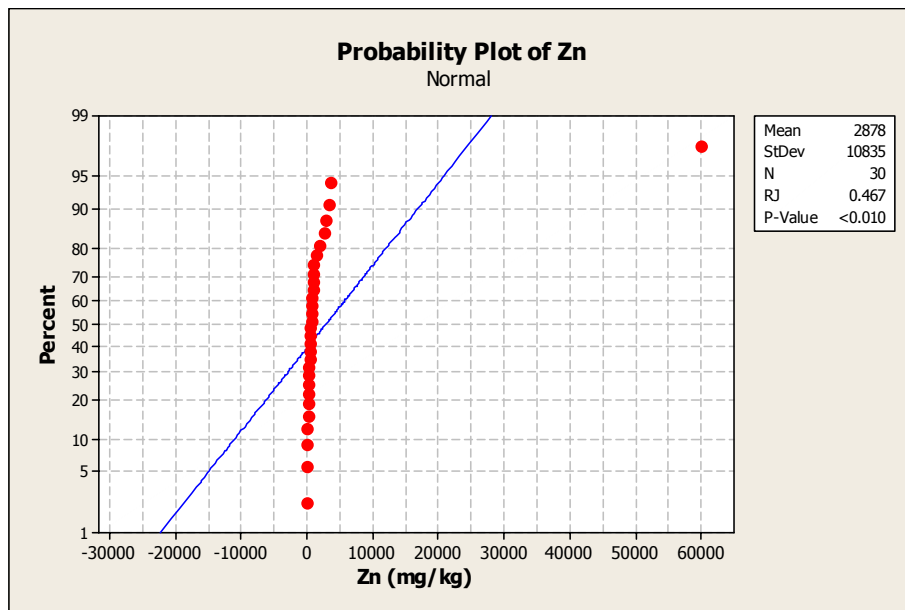


Figur 5.6. *Screening*karta för Zn inom det undersökta området. Inrutade provpunkter avser halter över 700 mg/kg (MKM). Kartan visar att det inte går att dra någon konkret slutsats om Zn bildar sammanhängande *hotspots* eller är utspridd över hela området.

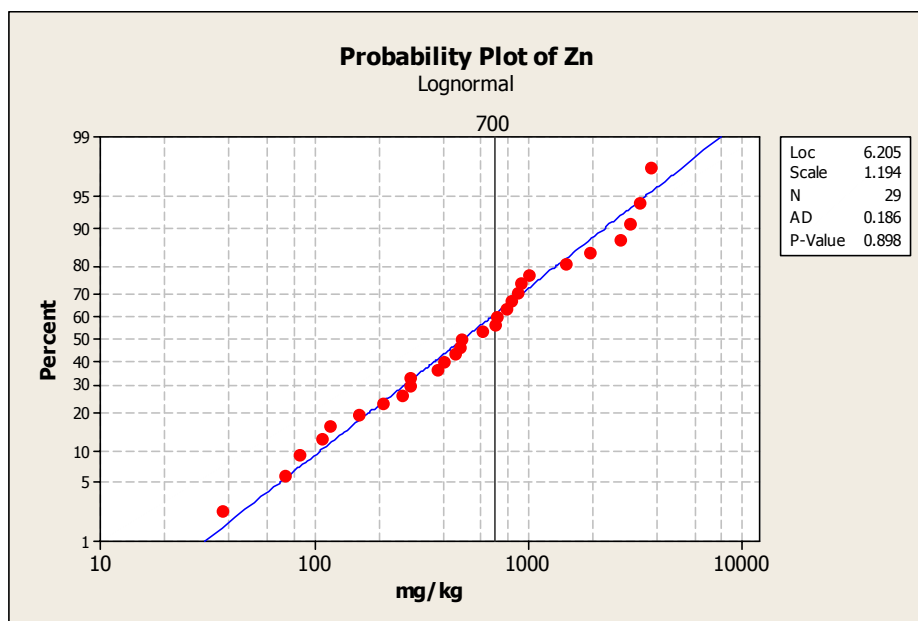
För zinkdata i detta exempel dras slutsatsen att det är svårt att påvisa något rumsligt beroende (*pure nugget effect* observerad i figur 5.5) och därmed rekommenderas istället en hypotesprövning för en kvantitativ analys av föroreningsgraden i relation till uppställt riskkriterium. Eftersom valet av lämplig metod för hypotesprövning beror på hur data är statistiskt fördelad, undersöks först vilken fördelning data följer. Ett normalitetstest för Zn visar att data avviker tydligt från en normalfördelning, se figur 5.7. Dessutom avviker en datapunkt (halt 60 000 mg/kg) avsevärt från resten av data. Det är inte sällan man konfronteras med en fråga om avvikande data beror på mätfel och kan betraktas som så kallade *outlier*, eller om det man ser är ”verkliga” halter och man skall acceptera mycket stora variationsspann. Det finns statistiska tester för att undersöka om det kan vara en *outlier* (se t.ex. Gilbert 1987). I vissa fall kan det finnas skäl att plocka bort avvikande data inför fortsatt datautvärdering. Detta är dock **inte att rekommendera, så länge det inte finns starka bevis** för att avvikelse(er) beror på andra identifierbara orsaker.

I föreliggande exempel visade det sig att positionen på den avvikande punkten (den röda punkten längst till vänster i figur 5.6) sammanfaller med ett f.d. metallsmältverk och man har observerat att flera metaller uppvisar betydligt högre halter just i denna punkt jämfört med resten av området. Då man kunde spåra den

avvikande datapunkten till en liten lokal föroreningskälla rekommenderas att platsen i dess närmaste omgivning behandlas separat och att datapunkten tas bort från fortsatt utvärdering av resterande data. Även utan den borttagna datapunkten avviker data från en normalfördelning och vidare analys resulterade i att halterna ligger närmast en lognormalfördelning (se figur 5.8). Det skall dock påpekas att den punkten *inte* påverkar utslaget av de goodness-of-fit tester, som utfördes på data: vare sig med eller utan den punkten kan data konstateras vara normalfördelade, men både med och utan den punkten kan data antas vara lognormalfördelade.



Figur 5.7. Normalitetstest för zink. p-värdet ger att risken att ha fel om man förkastar nollhypotesen (dvs. att data är normalfördelade) är mindre än 1%. Därav drar vi slutsatsen att nollhypotesen kan förkastas, dvs. data är inte normalfördelade.



Figur 5.8. Lognormalitetstest för zink. Enligt testet går det inte att förkasta nollhypotesen om att data är lognormalfördelade. p-värdet ger att risken att ha fel om man förkastar nollhypotesen (som är att data är lognormalfördelade) är så stor som 90%, dvs. data kan antas vara lognormalfördelade.

Om en riskbedömare eller beslutsfattare konfronteras med frågan hur osäkerheter i Zn-data skall inkluderas när föroreningsgraden skall jämföras med t.ex. ett riktvärde, kan flera olika tillvägagångssätt vara aktuella. Detta beror på att det i Sverige praktiskt taget inte finns några vägledningar och rekommendationer inom ämnet med myndighets sanktion. Det kan därför uppstå situationer där olika sätt att behandla samma problemställning resulterar i olika beslutsunderlag.

Nedan presenteras tre olika sätt att hantera osäkerheten och vilka konsekvenser det kan ha för beslutsunderlaget.

### Utvärdering 1

Ingen hänsyn tas till att Zn-data inte är normalfördelade. En aritmetisk medelvärdehalt räknas fram och  $UCL_{95}$  jämförs med riktvärdet. Medelvärdet för stickprovet är 908 mg/kg och  $UCL_{95}$  för medelhalten beräknas till 1232 mg/kg med antagandet att data följer en normalfördelning (med freeware ProUCL). Detta överstiger klart 700 mg/kg, och man konstaterar att området är förorenat.

### Utvärdering 2

Data testas för normalitet och med stöd i Figur 5.8 dras slutsatsen att data är lognormalfördelade. *Chen-test* tillämpas därför med en signifikansnivå på 5% och följande hypotesformulering:

$$H_0: \text{medelhalten} \geq 700 \text{ mg/kg}, H_1: \text{medelhalten} < 700 \text{ mg/kg}$$

ger  $p\text{-värde} = 0.905$

Resultatet av testet säger att vi *inte* kan förkasta nollhypotesen att medelhalten i området är lika med eller över riktvärdet på 700 mg/kg. Risken att ha fel *om* nollhypotesen skulle förkastas är mycket hög, 90,5%, dvs. om vi på basis av data konstaterar att området är rent är risken att vi har fel 90,5%. Därför konstaterar vi att medelhalten i området troligtvis ligger över riktvärdet.

Ett annat sätt att ta hänsyn till osäkerheten i föroreningshalten om man har starka skäl att påstå att data är lognormalfördelad är att använda sig av  $UCL_{95}$  beräknat med t.ex. *Land's* metod eller *Chebyshev Inequality*-metoden (US EPA 2002). Enligt *Land's* metod blir  $UCL_{95} = 1856$  mg/kg (beräknat med ProUCL) och detta värde överskrider markant det aktuella riktvärdet.

### Utvärdering 3

Om man har starka skäl att tro att data ej följer en lognormalfördelning och drar slutsatsen att data inte följer en lognormalfördelning trots allt, utför man istället en icke-parametrisk hypotesprövning eller en beräkning av  $UCL_{95}$  för medelhalten baserad på så kallad *bootstrap sampling* (US EPA 2002; ADEC 2003).

En *icke-parametrisk* hypotesprövning resulterar i följande:

- För teckentest blev det lägsta *p*-värde 0.23 och det för formuleringen:  
 $H_0$ : medianhalten  $\geq 700$  mg/kg och  $H_1$ : medianhalten  $< 700$  mg/kg
- För Wilcoxon teckentest det lägsta *p*-värde blev 0.32 och också det för  
 $H_0$ : medianhalten  $\geq 700$  mg/kg och  $H_1$ : medianhalten  $< 700$  mg/kg

Slutsatsen från hypotesprövningen är att nollhypotesen *inte* kan förkastas och att risken att ha fel *om* man förkastar nollhypotesen (som var att medianhalten är större eller lika med riktvärdet) är lägst 23%, dvs. en stor risk. Här ska man dock komma ihåg att testet utförs på basis av *medianhalten*, som för en skev fördelning är betydligt lägre än medelhalten. Jämfört med utvärdering 2, är det alltså inte så konstigt att man får en större felrisk, när man testar mot medianen eftersom medianen troligtvis ligger närmre riktvärdet än vad medelhalten gör.

Istället går man vidare med att beräkna  $UCL_{95}$  för medelhalten med antagandet att data är icke-parametriska. Bootstrapping är en statistisk metodik för att skatta fördelningar av mätdata. I motsats till traditionella metoder så ligger inga teoretiska fördelningar (som t.ex. normal- eller lognormalfördelning) till grund för beräkningarna. I stället skapas referensfördelningen från insamlade mätdata genom att göra ett stort antal slumpmässiga provurval från originaldata. Praktiskt går bootstrapping till enligt följande:

- Ett prov dras slumpmässigt från de aktuella mätdata och proceduren upprepas tills antalet prov är lika med det ursprungliga. Det valda provet "läggs tillbaka" varje gång. Ett enstaka prov kommer därför att ibland representeras flera gånger i det nya provurvalet och ibland inte alls.
- Medelvärdet (och vad man övrigt kan önska) beräknas för det nya provurvalet.

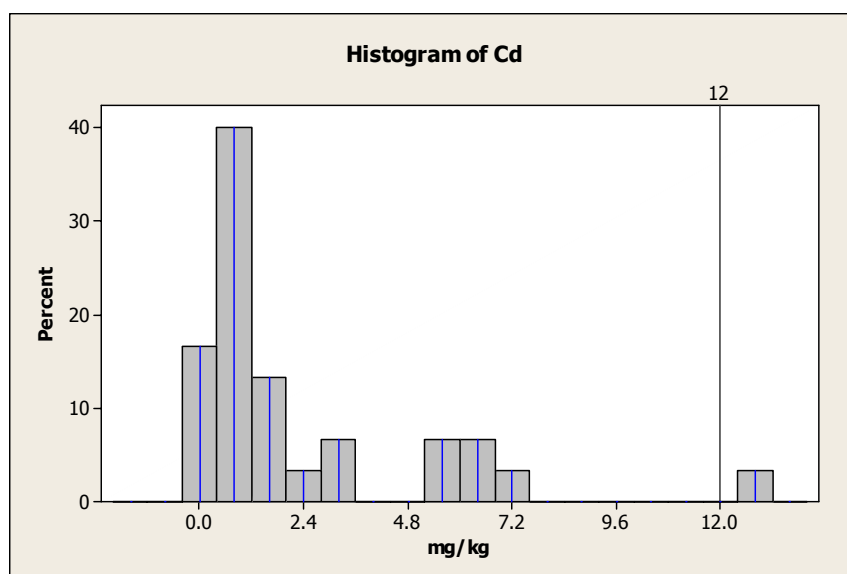
- Steg 1-2 upprepas ett stort antal gånger (1000-10000). Dessa 1000-tals skattningar av medelvärdet ger den nya referensfördelningen och den 95%-iga konfidensgränsen är helt enkelt 95%-percentilen.

Bootstrapping kan hantera situationer, när traditionell statistik kommer till korta eller blir extremt komplicerad, t.ex. när många prov är under detektionsgränsen.  $UCL_{95}$  beräknat med bootstrappade Zn-data, enligt tillvägagångssättet från (US EPA 2002) är 1289 mg/kg, vilket ligger klart över riktvärdet på 700 mg/kg.

#### 5.4.2 Kadmiumförorening

Kadmium valdes för att ytterligare exemplifiera osäkerheter i beskrivning av föroreningskällan i relation till riskkriteriet. Dess rumsliga utbredning liknar mycket det mönstret som observerats för zink, dvs. inget rumsligt beroende kunde påvisas från variogrammodellering och det fanns inga tecken på sammanhängande *hot-spots*.

Uppmätta halter ( $n = 30$ ) visade sig vara icke-parametriska, dvs. varken normal- eller lognormalfördelning kunde påvisas. I figur 5.9 presenteras histogram för Cd med angivet riktvärde 12 mg/kg (Naturvårdsverket 1996).



Figur 5.9. Histogram för kadmiumdata. Riktvärde för MKM på 12 mg/kg angivet.

Både *teckentest* och *Wilcoxon-test* resulterade i att nollhypotesen:

$$H_0: \text{medianhalten} = 12 \text{ mg/kg} \text{ och } H_1: \text{medianhalten} < 12 \text{ mg/kg}$$

kunde förkastas med risken att ha fel om man förkastar nollhypotesen lika med p-värdet som beräknas till 0(!). Vidare, om  $UCL_{95}$  beräknas med Bootstrapping-metoden, erhålls 3,7 mg/kg vilket också ligger långt under riktvärdet.

Med underlag från hypotesprövning och  $UCL_{95}$  konstaterar man att det finns starka skäl för att anse att Cd halter (representerade antingen med median eller  $UCL_{95}$ ) inte överstiger riktvärdet. Situationen med Cd är relativt lätt att hantera eftersom *typ I-felet (alfafelet)* är mycket litet (i princip noll) och nollhypotesen kan därför förkastas utan någon större risk att man förkastar en sann hypotes. Medianen för stickprovet är 1,1 mg/kg och även om Cd data uppvisar en hög varians (variationskoefficient  $\approx 1,3$ ) är det inte skäl för att inte förkasta nollhypotesen. Slutsatsen kan dras att varken medel- eller medianhalten av kadmium i området överskrider riktvärdet.

## 5.5 Säkerhet i beslutsunderlag

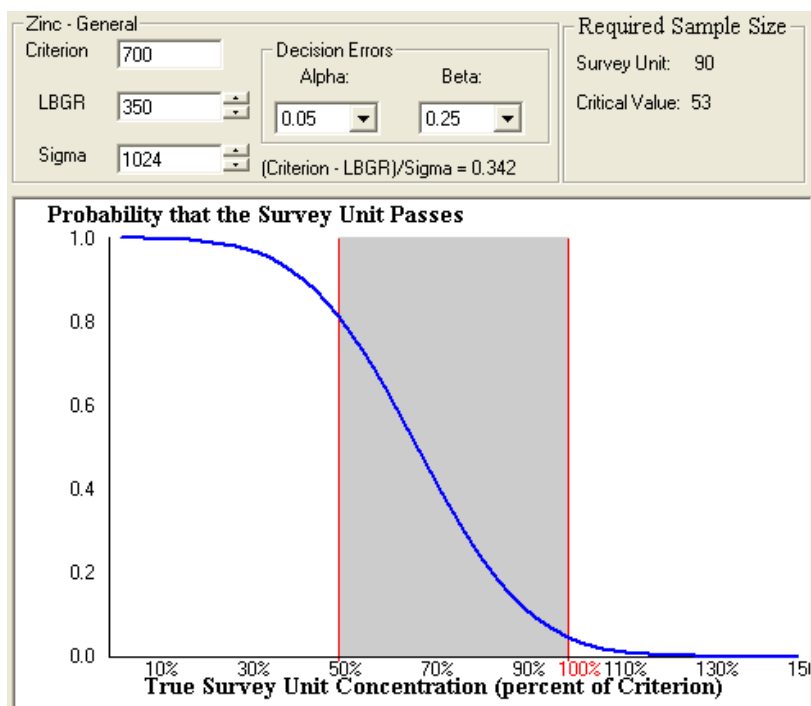
Eftersom kvaliteten på data påverkar kvaliteten på riskbedömningen kommer en beslutsfattare att vara intresserad av om dataunderlaget för att beskriva föroreningsgraden på området är tillräckligt för att gå vidare med förkanalys och eventuella åtgärder. Framtagna riktvärden förväntas ha sina ”egna” osäkerheter och det kan vara berättigat att utgå ifrån ett variationsspann och inte basera det enbart på ett fast värde (punktskattning), särskilt när man befinner sig i en så kallad *risk-screening* fas. Säkerheten i beslutsunderlaget beror av både mängden data och datas variationsmönster. Då få datapunkter är tillgängliga och om dessutom haltena varierar mycket kommer det att vara svårt att dra slutsatser om populationens parametrar och dess relation till riskkriterier med en rimlig konfidens. US EPA har tagit fram ett koncept, kallat *gray region*, för att kunna kvantifiera den mängd information (data) som krävs för att säkerställa tolerabla osäkerhetsnivåer i relation till variansen i data och uppställda beslutskriterier (US EPA 2000b). *Gray region* kan betraktas som ett osäkerhetsintervall, som räknas fram för medel eller medianhalten utifrån bestämda gränsvärden och ”önskade” nivåer på *alfa-* och *beta-fel*. Storleken på *gray region* definieras av beslutsfattaren och beräkningen av beslutsfel görs med så kallad *Decision Power Curve* (DCP) (US EPA 2000b; Purucker and Stewart 2004).

Figur 5.10 illustrerar DCP för Zn-data, där *gray region* har satts till mellan 350 och 700 mg/kg, där regionens nedre gräns (*Lower Boundary Gray Region – LBGR*) är lika med halva riktvärdet och övre gräns är lika med riktvärdet (*Criterion* i figuren). *Sigma* är stickprovets standardavvikelse och är lika med 1024 mg/kg. Tolkningen av DPC för det givna exemplet är följande:

- Om det visar sig att den observerade medianhalten är lika eller överstiger *Criterion* (riktvärde) drar vi slutsatsen att området är förorenat. Denna slutsats är dock ”belastad” med 5% risk (*alf fel*) att den är felaktig, eftersom den ju bygger på en stickprovspopulation. För att kunna dra denna slutsats krävs det dock total 90 prover (se *Survey Unit* i figuren), alltså  $90-29 = 61$  extra prover. Vi förutsätter dessutom att det är just medianhalten som är den representativa parametern för hela populationen och lämplig att jämföra med riskkriterium och inte t.ex. medelhalten.

- Om det visar sig att den observerade medianhalten är lika eller lägre än *LBGR* (halva riktvärdet) drar vi en slutsats att området är rent med en konfidens på  $1 - \beta$  fel, dvs.,  $1 - 0,75$ . Också i det här fallet krävs det 61 nya prover.
- Om medianhalten däremot skulle hamna inom *gray region* kan praktiskt taget inget beslutsunderlag tas fram inom den specificerade säkerhetsnivån.

Om en beslutsfattare vill vara säker (med 5% *alfa* fel) på att inte felaktigt konstatera att halten Zn är under riktvärdet, men samtidigt vill smala in *gray region* från andra hållet kan DCP användas för att räkna fram hur mycket data behövs, när *LBGR* höjs till t.ex. 400 mg/kg med oförändrat *beta* fel. För detta skulle det behövas 94 nya datapunkter. Vinsten blir att *LBGR* höjs med 50 mg/kg och med det beslutsunderlaget kommer man följaktligen inte behöva sanera området om medianhalten blir lägre än 400 mg/kg.



Figur 5.10. *Decision Power Curve* för zink data. För *alfafel* = 5% och *betafel* = 25% krävs totalt 90 provpunkter för att *gray region* skall sträcka sig mellan 350 och 700 mg/kg. Gray region anger osäkerhetsspannet för medianhalten. *DPC* är beräknad med verktyget SADA - Spatial Analysis and Decision Assistance, en gratisprogramvara sanktionerad av US EPA. (Se vidare <http://www.tiem.utk.edu/~sada/index.shtml> (Purucker and Stewart, 2004)).

## 5.6 Slutsatser

Valet av metod att karaktärisera en föroreningsituation i relation till uppställt risk-kriterium styrs av en konceptuell uppfattning om ämnets rumsliga fördelning. I



detta kapitel presenterades betydelsen av att välja rätt metodik, när man inte förväntar sig distinkta *hotspots* utan heterogen spridning av föroreningar.

Nedan sammanfattas de mest relevanta slutsatserna:

- Uppmätta halter uppvisar oftast stor variabilitet och för att ta hänsyn till variabiliteten, när halter jämförs med ett riskkriterium, krävs att ett lämpligt mått på osäkerheten inkluderas.
- Om riskkriterium bygger på *punktskattning* (ett *deterministiskt* riktvärde) ska denna skattning utvärderas med hänsyn till det framtagna osäkerhetsspannet som gäller för halter representativa för beskrivning av föroreningskällan.
- För att föra in osäkerheten i beskrivningen av källan i riskbedömningen kan antingen hypotesprövning eller konfidensintervall (UCL) användas. Valet är inte alltid självklart och det optimala är att använda sig av båda.
- Hypotesprövning utförs med hjälp av statistiska tester. Testerna bör väljas i relation till den statistiska fördelning insamlad data uppvisar och kan även utföras för andra skattningar än medelvärde eller median, t.ex. högre percentiler. Samma princip gäller vid valet av lämplig metod för skattning av UCL.
- Hypotesprövning kan användas för att utvärdera om kompletterande undersökningar (mer data) medför ett mer trovärdigt underlag för riskbedömning med hjälp av *Decision Power Curve*.
- Tolerabla osäkerhetsnivåer, som kvantifieras med *alfa*- och *beta*fel bestäms av beslutsfattare. Beslut om att tillåta större osäkerhet skall relateras till potentiella konsekvenser för riskbedömningen. På samma sätt, om mindre osäkerhet efterfrågas, bör en kostnadsnyttoanalys göras.

## 6 Rumslig osäkerhet i riskbedömning

I detta kapitel presenteras hur osäkerheterna i den rumsliga utbredningen av föroreningar kan påverka beslut om eventuella saneringsåtgärder. Studierna bygger på ett verkligt fall, en industritomt i Göteborg, som undersökts vid flera olika tillfällen och med tillfredställande dataunderlag för att användas demonstrationsobjekt. Detta är samma objekt som används i illustrativt syfte i kapitel 5. Formulerade slutsatser och rekommendationer bör endast ses som en generell pedagogisk beskrivning av problemställningen, som angrips inom ramen för det här projektet. Demonstrationsexemplet har till syfte att sprida kunskapen och skall alltså inte användas som beslutsunderlag för eventuella verksamheter rörande det konkreta objektet.

### 6.1 Problemställning och förutsättningar

I detta kapitel problematiseras ett fall där endast delar av undersökningsområdet är påverkat och där det finns starka skäl att förutsätta att föroreningsutbredningen är rumsligt begränsad. Sådana situationer är vanliga vid undersökningar av t.ex. nedlagda bensinstationer, där man ofta identifierar en punktkälla (t.ex. utsläpp från en korroderad diesel- eller bensintank och en geografiskt begränsad föroreningsplym kring denna.

Under förutsättning att provtagnings- och mättstrategierna har genomförts korrekt och att det finns en tillräcklig datamängd i relation till uppställda kvalitetskrav tas föroreningskartor fram (antigen som 2D eller 3D) med hjälp av interpolationsteknik. Ur datautvärderingssynpunkt innebär detta en avgränsning av föroreningen. I interpolationsprocessen skattar man olika parametrar och väger in olika bedömningar. Interpolerade kartor kan lätt bli influerade av subjektiva antaganden. De ”personliga” skattningarna och antagandena behöver inte nödvändigtvis minska trovärdigheten i framtagna interpoleringsmodeller. Å andra sidan testar man sällan alternativa antaganden och därmed har man ofta dåliga kunskaper om det subjektiva inslaget i interpoleringen.

För att demonstrera hur både valet av interpolationsmetod och specifika parametrar kan påverka riskbedömningen och beslutsprocessen har man med olika antaganden tagit fram föroreningskartor för ett och samma förorenade område med fyra olika interpolationstekniker. Samtidigt testades effekten av hur två olika riskmodeller: US EPA:s (US EPA 1989; US EPA 1991) och Naturvårdsverkets modell (Naturvårdsverket 1996) i kombination med olika interpolationstekniken inverkar på beslutsunderlaget.

### 6.2 Objektbeskrivning

Se beskrivning under avsnitt 5.2.

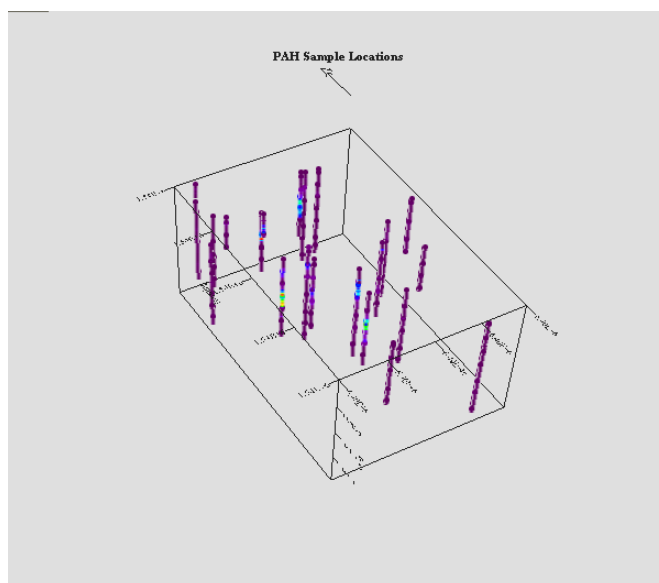
## 6.3 Insamlade data

Kunskapen om vilka verksamheter som har utövats på den aktuella tomten har använts för att ta fram en konceptuell modell för området. Modellen bygger på riskkedjan (källan-transport-recipient) och ingår i ett första steg av datakvalitetsmålprocessen (DQOP; se US EPA 2000b). Denna kunskap tillsammans med resultaten från tidigare platsundersökningar resulterade i att man identifierade tre möjliga utsläpp (*hotspots*) av kolvätebaserade petroleumprodukter.

Som undersökningsmetod användes FFD-miljösond (Fuel Fluorescence Detector, (Apitz et al. 1992; Nilsson-Påledal et al. 2003) integrerad med ett CPT system (Cone Penetration Technology). I korthet ger FFD en relativ föroreningshalt av aromatiska kolväten med indelning i lätta och tunga aromater, medan CPT ger information om jordens mekaniska/geotekniska egenskaper. FFD är en skanningmetod, vilket betyder att den ger en direkt påplatsinformation om det finns tecken på petroleumprodukter. Sonden trycks ner i marken från en borrhandsvagn och mätningar utförs kontinuerligt längs djupprofilen.

För den aktuella platsen består FFD-databasen av 2290 mätpunkter, varav 1465 är över detektionsnivå. Totalt har man utfört 20 sonderingar ner till ca 4 m djup inom ramen för ett samarbetsprojekt mellan SGI och Applied Research Associates Inc. hösten 2004. Med undantag från första sonderingen, som placerades utifrån information från tidigare studier och antogs ligga inom föroreningsplymen, styrdes placeringen av varje sonderingspunkt av resultaten från föregående sondering. I möjligaste mån placerades sonderingarna i profiler längs och tvärs över den konceptualiserade föroreningsplymen. Om det visade sig finnas förhöjda halter petroleumprodukter i en sondering, fortsatte man ca 5 m bort längs samma profil tills man hittade en ”ren” sondering. På det sättet fick man en avgränsning av plymen både vertikalt och horisontellt.

Eftersom FFD-sonden inte analyserar enstaka ämnen utan ett brett spektrum av olika komponenter, har man gjort en omräkning av FFD-resultat till ett ämne, som bäst ”matchar” sondens utslag. Det ”ämne” som bedömdes ligga närmast FFD-värdena, med avseende på kemisk sammansättning, var den totala mängden petroleumbaserade kolväten (*Total Petroleum Hydrocarbon – medium range*), TPH. Omräkningen från FFD till mg/kg TPH bygger på ett korrelationssamband mellan FFD-utslaget och laboratorieanalysen av ett jordprov taget nära intill ett av sonderingshålen. På så sätt kunde alla FFD-mätningar räknas om till mg/kg TPH. Det ska påpekas att denna omräkning är grovt förenklad och i en ”vanlig” utredning bör den kvalitetssäkras med kompletterande analyser. Omräkningen innebär alltså ytterligare en källa till osäkerhet. Den behandlas dock inte i denna studie. Figur 6.1 presenterar hur sonderingarna placerades i förhållande till varandra.



Figur 6.1. Placering av sonderingshål inom det undersökta området.

## 6.4 Interpolering och riskbedömning

### 6.4.1 Interpolering - överblick

Uppskattning av volym förorenad jord bygger på interpolation mellan observationspunkter (här FFD-mätningar). För att demonstrera effekten, som olika metoder och antaganden kan ha på skattning av mängden förorenade jordmassor, visas fyra olika interpolationsmetoder:

- Nearest neighbour (NN)
- Inverse distance (ID)
- Kriging med korrekt variogrammodell ( korrekt K)
- Kriging med fel variogrammodell ( fel K)

För teori bakom varje metod hänvisas läsaren till t.ex. Isaaks and Srivastava (1989) och Purucker and Stewart (2004), här sammanfattas endast metoderna helt kort. Metoden NN går ut på att den punkt, som värdet skall interpoleras för, får samma värde som närmaste observationspunkt. I ID-metoden läggs en influensarea (*search neighbourhood*), en ellips för 2D och en ellipsoid för 3D med centrum i interpolationspunkten, och ett slags medelvärde för alla observationen inom influensarean beräknas. De observationer som ligger närmast den interpolerade punkten, får mer vikt (mer prioritet). Vikten beror således på avståndet mellan den interpolerade punkten och observationspunkter inom influensarean. Kriging är en betydligt mer komplex metod, som bygger på geostatistik där observationerna viktas beroende på hur en s.k. korrelationsstruktur (variogram) ser ut för det aktuella fallet.

Interpolationsosäkerheter återspeglar de fel och osäkerheter som kan uppstå under själva interpolationsproceduren. Antalet observationspunkter, samt deras placering i förhållande till utbredning av föroreningen, påverkar direkt kvalitet på

interpolationen. Kartor över föroreningshalter ger ett kvantitativt underlag för vidare beräkning av storleken (volymen) av massor som överskrider aktuellt riskkriterium och kan vara ett utmärkt pedagogiskt redskap. Ofta föredras dock formen framför innehållet. Få interpolerade kartor tas fram under rätta antaganden och i många fall kan osäkerheter bli så betydande att resultatet i högsta grad bör ifrågasättas.

Kriging möjliggör kvantitativ analys av osäkerheter vid interpolering. Kartor över varians och sannolikhetskarter (sannolikheter att hitta en specifik föroreningshalt inom undersökt område) kan tas fram. Dessa aspekter presenteras dock inte i denna studie.

Genom att jämföra resultat för NN, ID och K, illustreras skillnaden i volym-skattning och rumslig fördelning av förorenade massor som en effekt av de olika interpoleringskoncepten. Vidare, som komplettering visas med en kriginginterpolation hur känslig variogrammodellering kan vara för efterföljande beräkningar av förorenade volymer. En uppskattning av volymerna görs för ett ”korrekt” modellerat variogram och en för ett felaktigt. I den felaktiga modellen antar man dubbelt så stor *range* och ingen *nugget*-effekt (för förklaring och fördjupning inom variogram och geostatistik hänvisas läsaren till Journel and Huijbregts (1978) och Isaaks and Srivastava (1989). Variogrammodellering görs ofta utifrån förutbestämda *default*-värden, som modelleringsprogrammet tar fram och sällan med ett aktivt bidrag från modellören. Detta leder till att krigingkartor ofta bygger på en felaktigt skattad korrelationsstruktur.

All interpolering i denna studie gjordes med hjälp av verktyget SADA (Spatial Analysis and Decision Assistance, se bl.a. Purucker and Stewart (2004)).

#### 6.4.2 Riskmodell

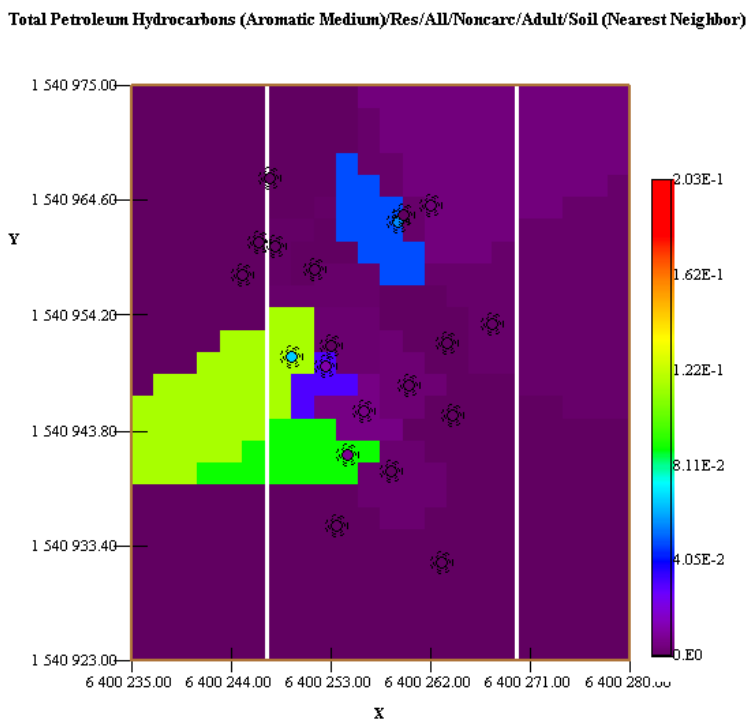
Parallellt med utvärderingen av interpolationsrelaterade osäkerheter undersöktes också effekten av olika risknivåer på den rumsliga fördelningen av föroreningen. Beräkningarna gjordes under antagandet att platsen i framtiden kommer att nyttjas för bostadsbebyggelse. Både US EPAs riskmodell för hälsoeffekter (US EPA 1989) och Naturvårdsverkets generella riktvärdesmodell (Naturvårdsverket 1996) användes. Enligt US EPA-modellen blev det beräknade riktvärdet (betraktas som *Preliminary Remediation Goals*, PRG) för *residential landuse scenario* 70 mg/kg och enligt den svenska modellen blev det 20 mg/kg.

## 6.5 Resultat

Rumsliga osäkerheter presenteras som en jämförelse av risk- och AOC (area of concern)-kartor för olika interpoleringsmetoder (och antaganden) och olika riskkriterier. Riskkartorna visar hur sannolikheten för att överskrida en framtagen referensdos varierar inom området. AOC-kartor bygger direkt på risk och demonstrerar vilka delar av området som inte uppfyller det uppställda miljökriteriet och därför kan bli föremål för sanering.

### 6.5.1 Riskkartor

Varje interpoleringsmetod resulterade i en ”egen” riskkarta (se figurerna 6.2 – 6.5). Varje figur presenterar samma skikt (djupintervall mellan 1,2 och 1,5 m), då det skiktet visade sig vara mest förorenat och lämpligast som demonstrationsexempel. Resultaten bygger på US EPA:s riskmodell och beräknat PRG = 70 mg/kg.

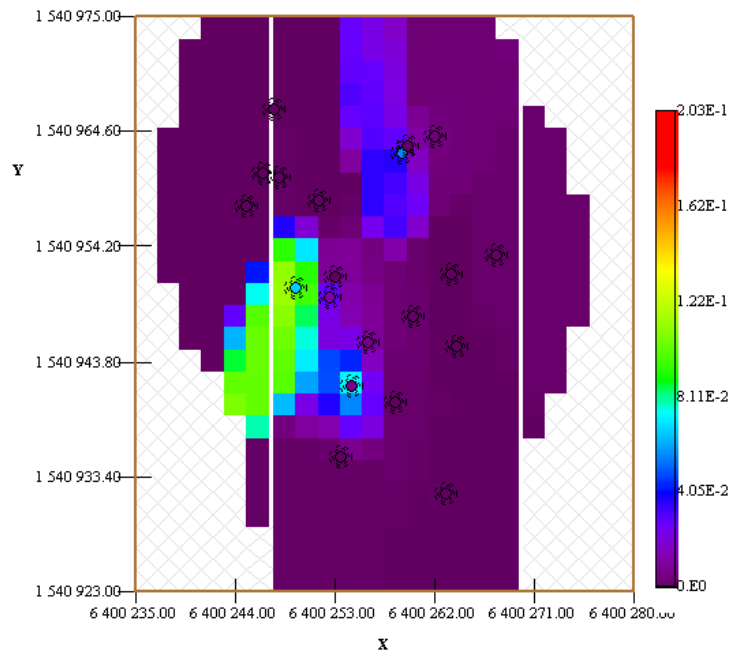


Figur 6.2 Riskkarta för Nearest Neighbour interpolation. Färgerna motsvarar olika risknivåer, som överstiger PRG-nivå 70 mg/kg (Preliminary Remediation Goals). PRG har räknats fram med riskmodellen från US EPA.

Enligt US EPAs modell är värdena i de presenterade riskkartorna den troliga risken i varje cell. Den troliga risken beräknas som kvoten mellan den dagliga dosen för en individ (baserad på den troliga halten i varje cell och exponeringsan-taganden) och en referensdos vid vilken hälsoeffekter kan uppstå (dvs. en hazard quotient).

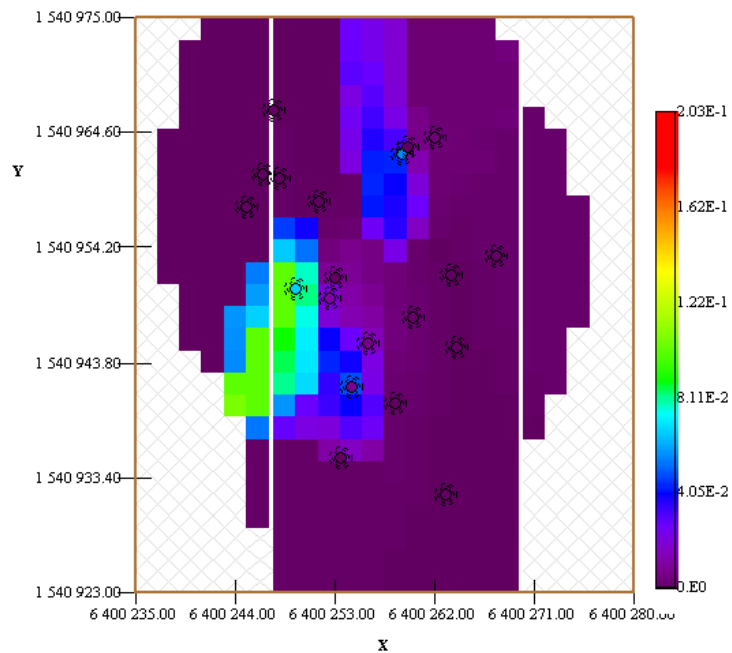
Utifrån en visuell bedömning är riskkartan baserad på NN interpolering (figur 6.2) den som avviker mest från de andra riskkartorna. Skillnaden mellan riskkartorna för ID och kriging är inte stor: tydligast är att med ID verkar området öster om den mest förorenade delen vara mer förorenat jämfört med korrekt K. För kartan baserad på fel K är risken inom samma område (öster om den mest förorenade delen) något underskattad jämfört med korrekt K och ID.

Total Petroleum Hydrocarbons (Aromatic Medium)/Res/All/Noncarc/Adult/Soil (Inverse Distance)



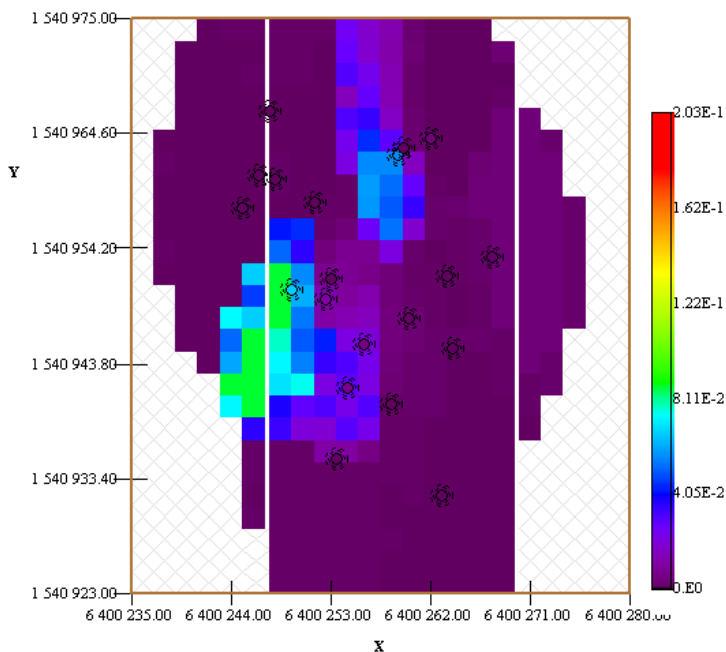
Figur 6.3 Riskkarta för Inverse Distance interpolation. Färgerna motsvarar olika risknivåer som överstiger PRG-nivå 70 mg/kg (Preliminary Remediation Goals). PRG har räknats fram med riskmodellen från US EPA.

Total Petroleum Hydrocarbons (Aromatic Medium)/Res/All/Noncarc/Adult/Soil (Ordinary Kriging)



Figur 6.4 Riskkarta för Kriging interpolation med korrekt variogram. Färgerna motsvarar olika risknivåer som överstiger PRG-nivå 70 mg/kg (Preliminary Remediation Goals). PRG har räknats fram med riskmodellen från US EPA.

Total Petroleum Hydrocarbons (Aromatic Medium)Res/All/Noncarc/Adult/Soil (Ordinary Kriging)



Figur 6.5. Riskkarta för Kriging interpolation med felaktigt variogram. Färgerna motsvarar olika risknivåer som överstiger PRG-nivå 70 mg/kg (Preliminary Remediation Goals). PRG har räknats fram med riskmodellen från US EPA.

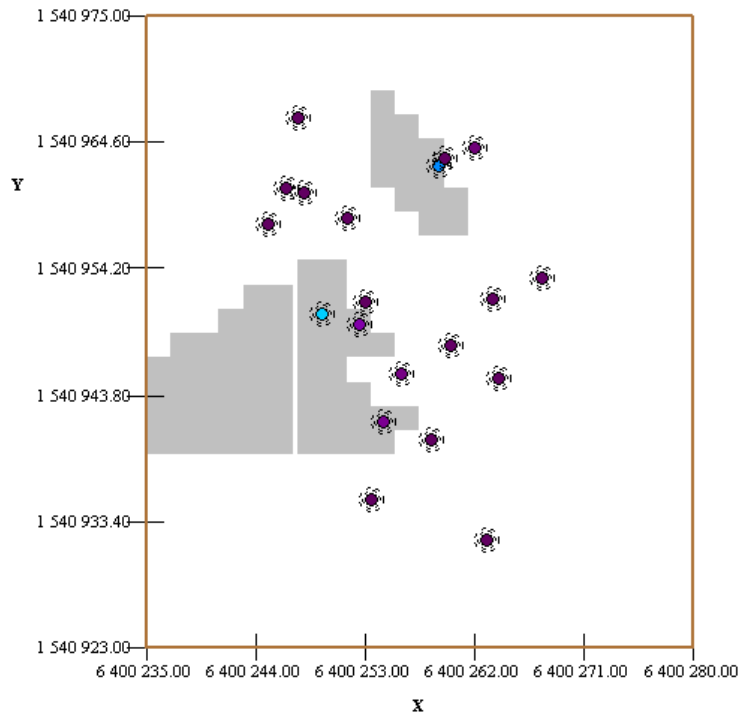
Riskkartorna kan inte användas för att direkt utvärdera kvantitativa skillnader mellan olika interpoleringstekniker utan ger endast en överblick över den rumsliga fördelning av celler/block/rutor, som uppvisar en specifik risknivå. Riskkartorna är ett steg före framtagandet av AOC-kartor. De kan dock ge viktig information till beslutsfattare om att föroreningen grupperas till vissa mindre delar av området och att en selektiv saneringsåtgärd eventuellt kan vara aktuell. Riskkartor bör också användas för en initial optimering av interpolationsproceduren.

### 6.5.2 Area of Concern

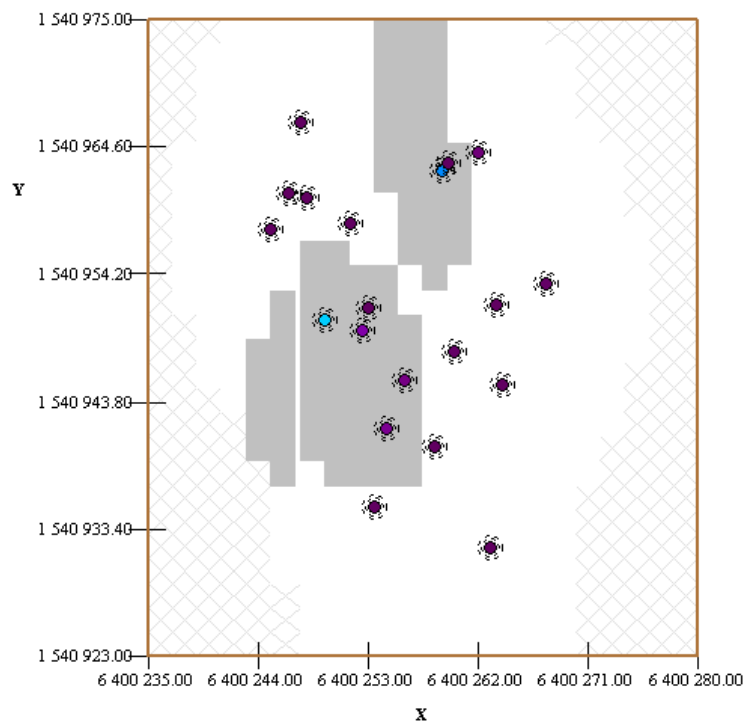
Koncept av AOC kan sammanfattas kort som en yta (volym), där halter överskrider aktuellt riskkriterium och potentiellt utgör föremål för sanering. I figurerna 6.6 - 6.9 demonstreras hur storleken och den rumsliga utbredningen av AOC beror av olika interpolationsmetoder.

Med undantag för skattningen med NN skiljer sig inte metoderna i det här *specifika fallet* (ID och felaktigt K) i någon betydande grad från K.

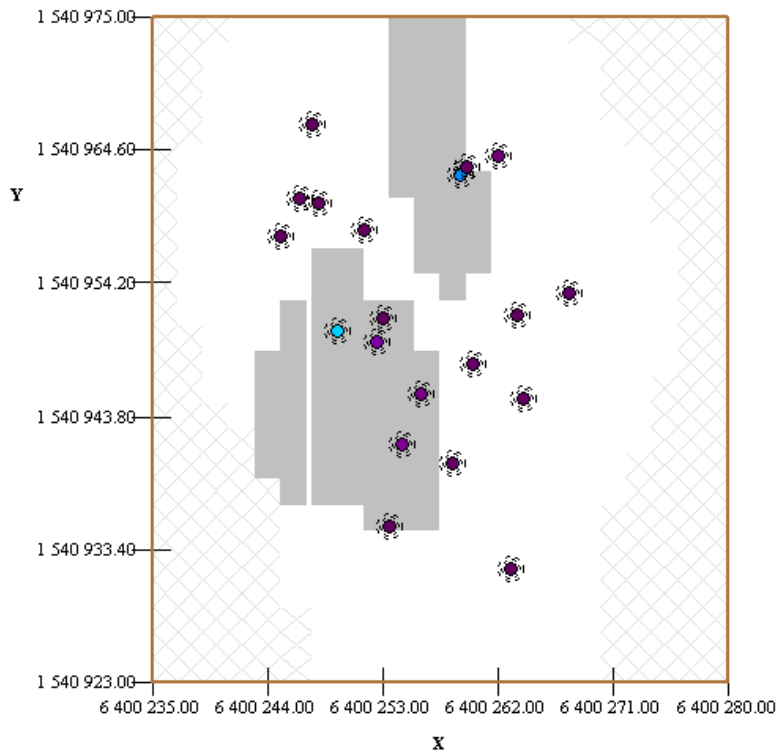




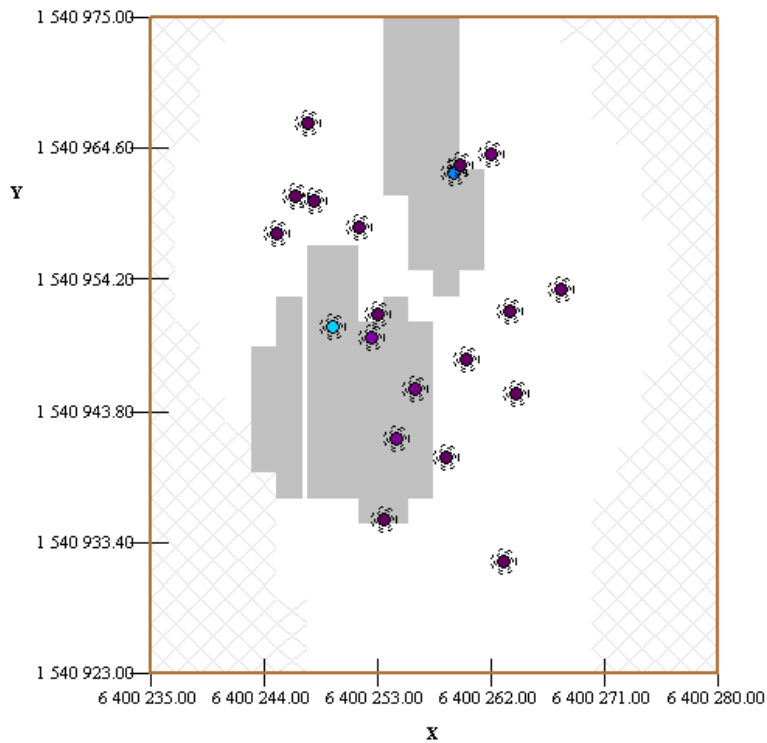
Figur 6.6 Area of Concern för Nearest Neighbour interpolation.  $PRG = 70$  mg/kg. Lägg märke till att NN-interpolationen för delar av området, där de andra metoderna inte klarar av att estimerera ett värde (se kryssade celler i Figurerna 6.7, 6.8 och 6.9 nedan). Den metoden uppskattar alltså arean på det rena området som större än de andra metoderna.



Figur 6.7 Area of Concern för Inverse Distance interpolation.  $PRG = 70$  mg/kg



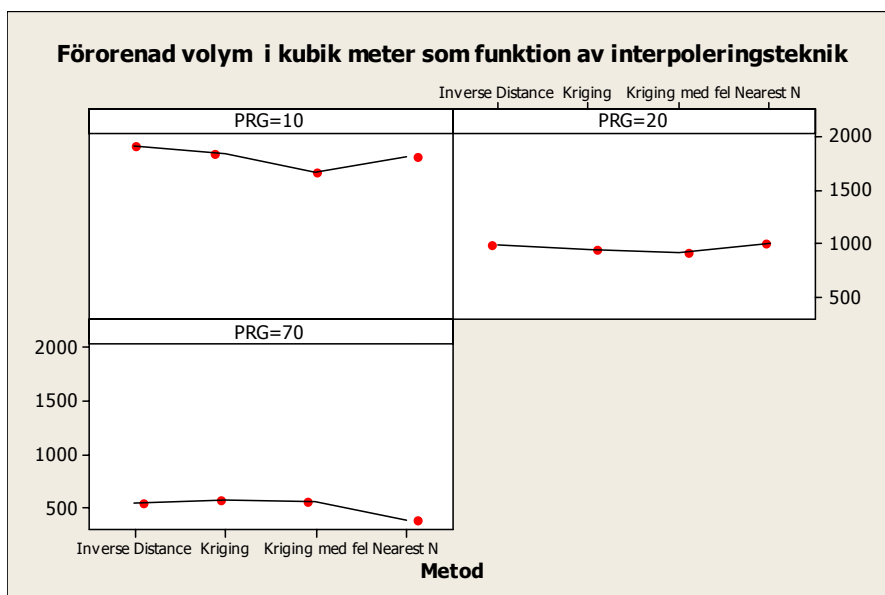
Figur 6.8 Area of Concern för Kriging interpolation med korrekt variogram. *PRG* = 70 mg/kg



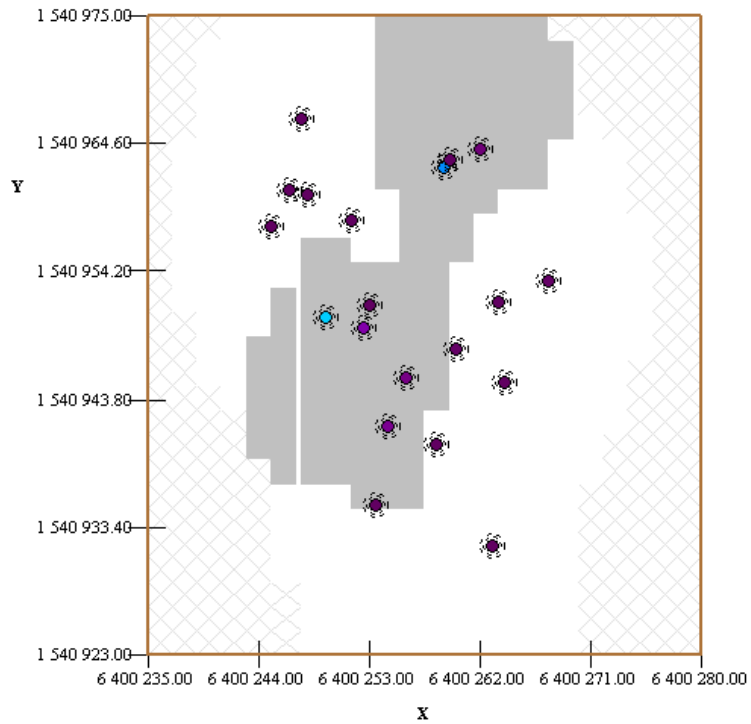
Figur 6.9 Area of Concern för Kriging interpolation med felaktigt variogram. *PRG* = 70 mg/kg

Eftersom PRG = 70 mg/kg, baserat på den amerikanska modellen, är ett värde som gäller enbart för hälsoeffekter, har vi också studerat utfallet från den svenska modellen, baserat på generella riktvärden men med hänsyn tagen till miljöeffekter. På så sätt kan vi jämföra hur modellosäkerheter (och indirekt också konceptuella osäkerheter), som beskrivs i avsnitt 2.2, i kombination med olika interpoleringsmetoder inverkar på AOC. För TPH finns det inga riktvärden fastställda för Sverige. Det närmast ämnesmässiga är aromater C10-C35. För dessa har man 20 mg/kg som ett föreslaget riktvärde. För att ytterligare undersöka hur osäkerheter i en bedömning av ett ”rätt” ämne (det finns ju en stor osäkerhet i att 20 mg/kg antogs gälla som ett riktvärde för TPH och samtidigt har man ju inte mätt halter TPH direkt) resulterar i skattningen av AOC, har vi gjort identiska beräkningar för hälften av det föreslagna riktvärdet, dvs. 10 mg/kg. Figur 6.10 illustrerar sambandet mellan AOC (förorenade jordvolym i m<sup>3</sup>) och interpolationsmetoder för tre olika PRG (riktvärden): 70, 20 och 10 mg/kg.

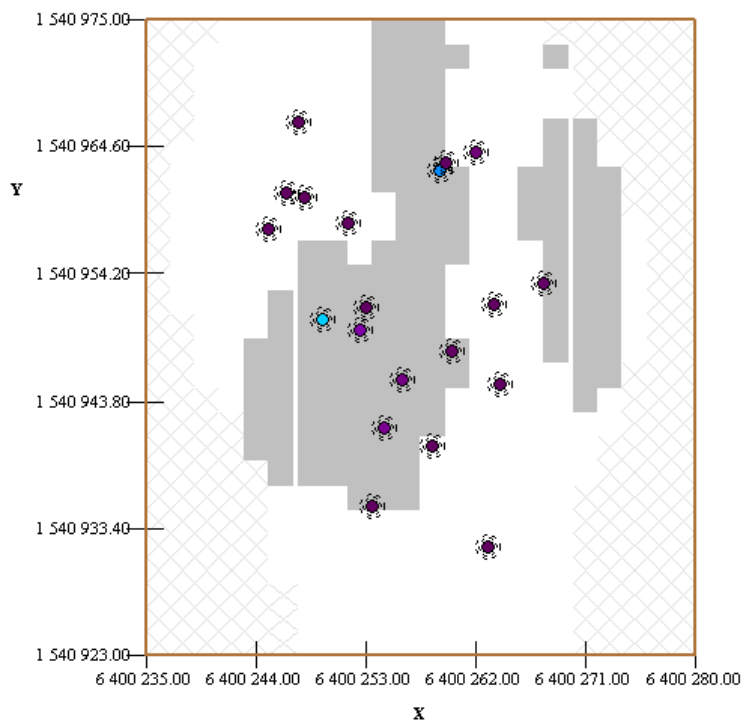
Som det framgår av figur 6.10, erhöles inga betydande skillnader i de beräknade förorenade mängderna (volym) mellan de olika interpoleringsmetoderna. Man har observerat att även om skattade volymer inte skiljer mycket mellan olika interpoleringstekniker, så kan AOC-kartorna se olika ut. Till exempel, för PRG = 20 mg/kg är differensen i volymkattning mellan rätt och fel kriging endast 37 m<sup>3</sup>, medan den rumsliga positionen av förorenade block inom området uppvisar tydliga skillnader, se figurer 6.11 för korrekt kriging och 6.12 för fel kriging.



Figur 6.10. Area of Concern omräknad till volym förorenad jord som funktion av interpolationsmetod och olika nivåer på PRG. Interpolation har utförts med en gridstorlek på 2 x 2 m.



Figur 6.11 Area of Concern för Kriging med korrekt variogram, PRG = 20 mg/kg. Gridstorlek för interpolation 2 x 2 m.

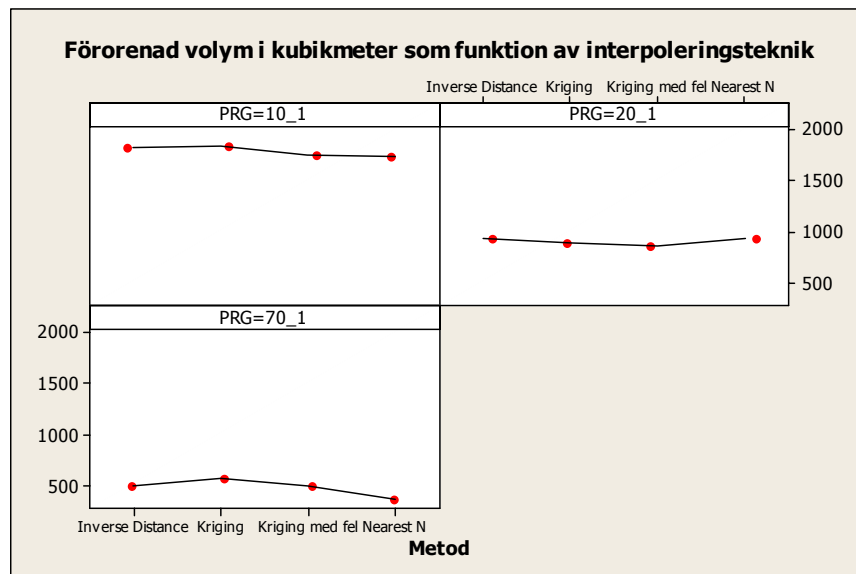


Figur 6.12 Area of Concern för Kriging med felaktigt variogram, PRG = 20 mg/kg. Gridstorlek för interpolation 2 x 2 m.

Alla beräkningar/skattningar presenterade i rapporten hittills gjordes med en förutbestämmd gridstorlek om 2 x 2 m. Eftersom det förorenade området ryms inom en

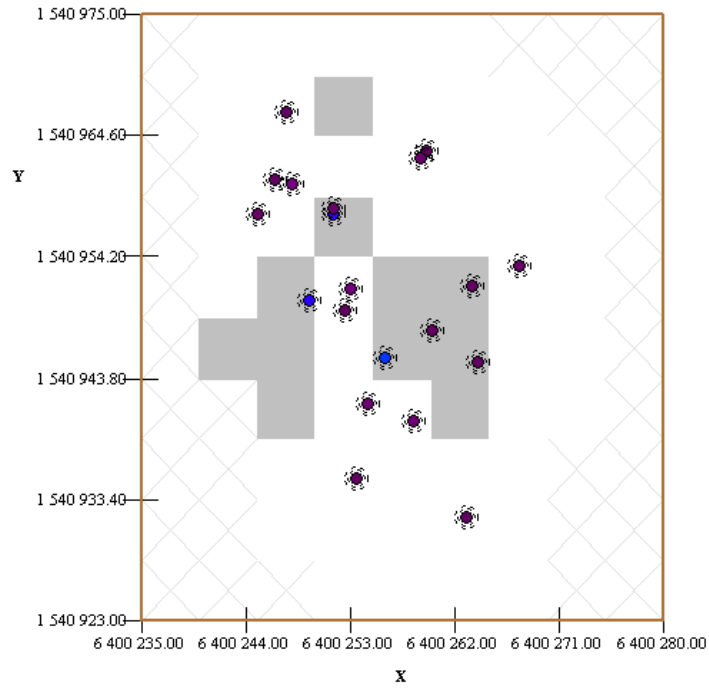
rektangeln om ca 45 x 50 m, är en indelning i 2 x 2 m rutor inte orimlig med avseende på kartprecision i förhållande till den mängd data som fanns tillgänglig. En indelning i större rutor kan dock vara mer berättigad om medelavståndet mellan provtagnings- eller mätpunkter i det horisontella planet är större än 2 m. En annan viktig aspekt med en indirekt, men stark, återkoppling till gridstorleken är dess relation till en *selektiv efterbehandlingsvolym* (SEV; Naturvårdsverket 1997). SEV är den minsta volym jord som under saneringsarbetet avses separeras och klassificeras. SEV styrs av riskerna för hälsa och miljö, men samtidigt måste hänsyn tas till det praktiska genomförandet. Om haltvariationen inom SEV är stor bör den minskas och vice versa. Det kan alltså finnas olika anledningar till vad man väljer för gridstorlek: (1) liten – avgränsningarna ser mer realistiska ut istället för att vara ”blockiga” och resultatet kan vara lättare att kommunicera, (2) baserad på exponeringsenheter, t.ex. tomtvis, och (3) efterbehandlingsenheter (SEV) som nämnts ovan.

För att studera resultatet av interpolering med en större grid har nya AOC-kartor räknats fram för en rutstorlek om 5x5 m. Som framgår av figur 6.13, kan man inte heller för den större griden observera några signifikanta skillnader i mängd förorenad jord mellan de olika interpoleringsmetoderna (jämför med figur 6.10).

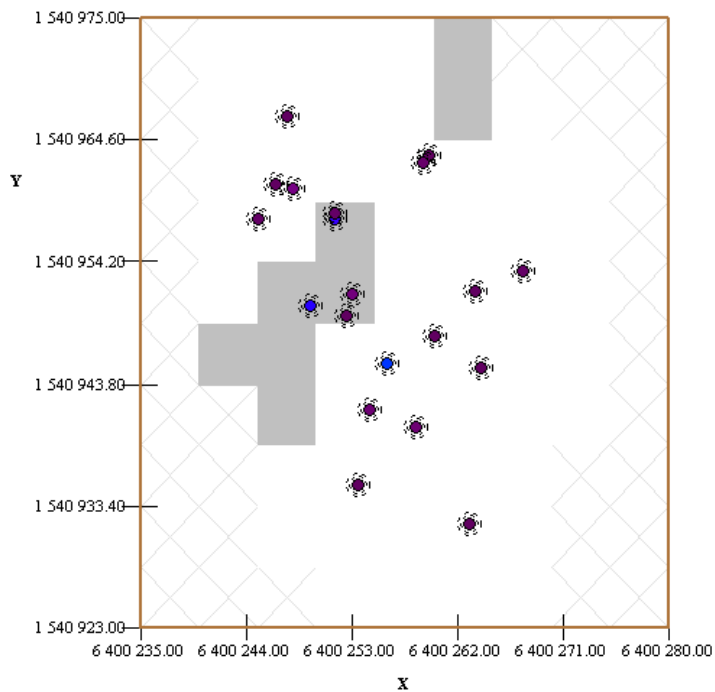


Figur 6.13 Area of Concern omräknad till volym förorenad jord som funktion av interpolationsmetod och olika nivåer på PRG. Interpolation har utförts med en gridstorlek på 5 x 5 m.

För denna gridstorlek och för PRG = 20 mg/kg visade sig dock den rumsliga fördelningen av AOC vara mer känslig än för 2 x 2 m. Detta illustreras i figurer 6.14 och 6.15. Den uppskattade AOC för korrekt K var 894 m<sup>3</sup> (949 m<sup>3</sup> för 2 x 2 m) och för fel K 863 m<sup>3</sup> (912 m<sup>3</sup> för 2 x 2 m). Volymskillnaden är alltså relativt liten: 30 m<sup>3</sup>, men däremot, som det framgår av figurerna nedan, skiljer sig det rumsliga mönstret en hel del.



Figur 6.14 Area of Concern för Kriging interpolation med korrekt variogram, PRG = 20 mg/kg. Gridstorlek för interpolation 5 x 5 m.



Figur 6.15 Area of Concern för Kriging interpolation med felaktigt variogram, PRG = 20 mg/kg. Gridstorlek för interpolation 5 x 5 m.

## 6.6 Slutsatser

Det finns flera möjliga källor till rumslig osäkerhet i riskbedömning. I detta kapitel demonstrerades effekten av att använda olika interpolationsmetoder och av valet av metodspecifika parametrar (och antaganden) på uppskattning av volym förorenad jord och dess rumsliga utbredning. Dessutom undersöktes inverkan av olika riskkriterier och riskkoncept.

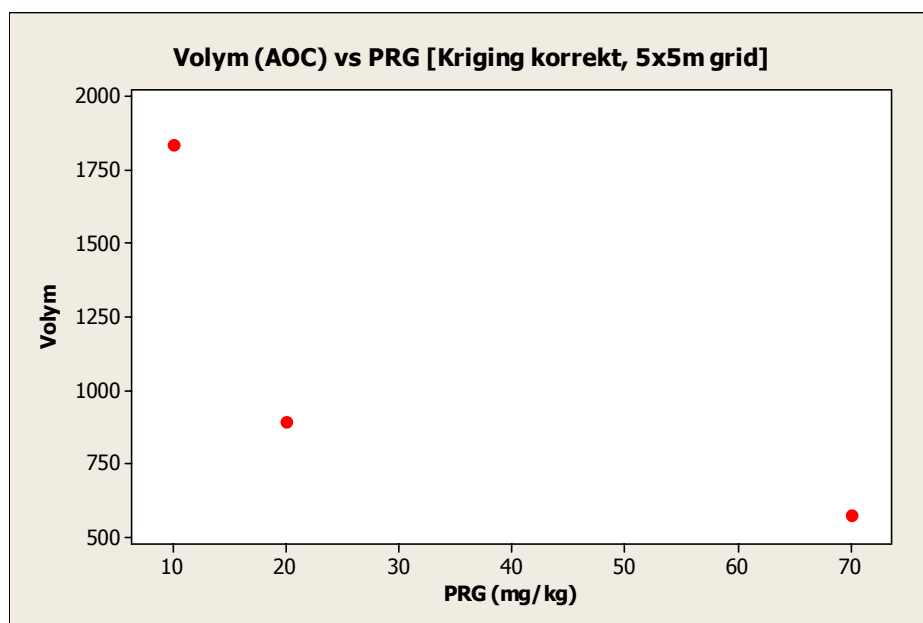
Studien visade att **i somliga situationer** kan valet av interpolationsmetod ha stora konsekvenser för riskbedömningen och efterföljande beslut. De metoder och metodparametrar som är mest lämpliga kan variera från fall till fall. I detta arbete ges inga fasta rekommendationer, utan istället fokuseras på om det överhuvudtaget är av betydelse att välja en metod framför andra.

Följande har konstaterats för *det studerade exemplet*:

- Skattning av mängder förorenad jordvolym visade sig tämligen oberoende av interpolationsteknik (figurer 6.10 och 6.13). Oavsett metodval och gridstorlek blev storleken på Area of Concern nästan densamma.
- Som förväntat, blev volymen (AOC) känslig för risknivå (uttryckt med *Preliminary Remediation Goals* i mg/kg). Sambandet är dock inte linjärt utan snarare exponentiellt (se exempel i figur 6.16). Detta faktum kan ha stora implikationer när saneringskostnader uppskattas i relation till en aktuell risknivå (riktvärde) – ju strängare riskkriterier desto snabbare ökning i åtgärdskostnad. Den här typen av osäkerhet är i grunden av modell- och konceptuell karaktär. PRG, som här användes som riskkriterium beror på vilket riskkoncept (modell) man använder sig av (PRG = 70 mg/kg baserade på US EPAs modell eller PRG = 20 mg/kg baserade på SNVs modell). Också osäkerheten i korrekt detektering av det aktuella föroreningsämnet som riskbedömning gjordes för är av betydelse. Denna osäkerhet är i sin tur delvis både konceptuell och delvis relaterad till felet vid uppskattningen av halter utifrån utslaget från mätinstrumentet och (i detta specifika fall) en omräkningsfaktor från skanningsdata till motsvarande laboratorieanalys.
- För ett ”milt” riskkriterium (PRG=70mg/kg) och relativt liten gridstorlek (2x2m) hade varken valet av interpoleringsmetod eller metodparametrar någon effekt på den rumsliga utbredning av förorenade jordmassor. Den mest sannolika förklaringen är att variationsmönstret för höga halter är något mer ”stationärt” och rumsligt begränsat jämfört med lägre halter.
- Ju strängare risknivå (lägre PRG) desto större olikheter i den rumsliga utbredningen av AOC som funktion av interpolationsmetod observerades (jämför figurerna 6.11 och 6.12).
- Skillnaderna i rumslig utbredning vid strängare riskkriterium mellan olika interpolationstekniker blev ännu mer påtagliga när gridstorleken ökades från 2x2m till 5x5m (jämför figur 6.14 och 6.15).
- För samma gridstorlek men olika interpolationsmetoder (figur 6.11 och 6.12) är storleken på AOC approximativt densamma och därmed kostnaden

för eventuella åtgärder lika. På grund av olika rumslig utbredning m.a.p. AOC kan dock genomförandet av saneringsarbetet resultera i **olika sluteffekter**: Man kanske sanerar en korrekt uppskattad mängd förorenad jord, men risken är stor att man delvis åtgärdar icke-förorenade massor och samtidigt lämnar kvar massor där halten överstiger risknivån.

- När en kombinerad inverkan från interpoleringsmetod och gridstorlek studeras kan både volym-skattning och rumslig fördelning skilja mycket (jämför figur 6.11 och 6.15).



Figur 6.16. Förorenad jordvolym som funktion av risknivå (PRG) för kriging interpolation med korrekt variogram och gridstorlek 5 x 5 m.

Av uppenbara skäl blir den praktiska konsekvensen (kostnads-kalkyl av eventuell sanering) av rumsliga osäkerheter mer märkbar för större platser än för små. För småskaliga problem är en närmare analys av känsligheten av AOC för valet av interpoleringsmetod inte alltid motiverad. I takt med att större volymer och större geografisk spridning förväntas kan en känslighetsanalys utnyttjas som ett verktyg först för att identifiera potentiella skillnader i uppskattad AOC och efteråt för att optimera lämpligt tillvägagångssätt för utvärdering och göra en kvantitativ skattning av osäkerheter.

Generellt kan sägas att geostatistiska interpolationsmetoder möjligtvis är att föredraga framför deterministiska metoder som ID och NN, eftersom man i de förra undersöker sin datastruktur genom en variogramanalys och sedan har den analysen som stöd för att välja parametrar för interpolationen. Med kriging kan man också analysera osäkerheten i skattningen av halt i varje cell, vilket inte är möjligt med deterministiska metoder. Att göra jämförelser mellan modeller är svårt, eftersom man inte har facit på hand, men ett stöd kan vara att korsvalidera data. Vid en korsvalidering plockar man bort en datapunkt i sänder och gör om



interpolationen utan den valda punkten och beräknar felet i uppmätt halt och framinterpolerad halt. Proceduren upprepas för alla datapunkter, varvid ett totalfel kan beräknas.

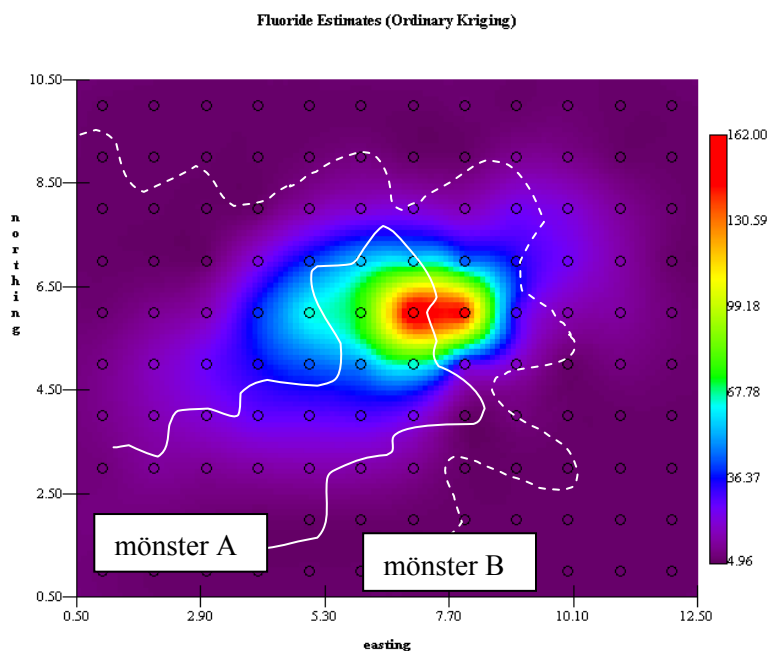
Rumsliga osäkerheter bör också tas med i beräkning i situationer när föroreningen uppvisar ett starkt heterogent spridningsmönster. Det kan dock vara svårt, eller t.o.m. omöjligt, att ta fram en verklighetstrogen haltkarta med hjälp av interpolering. I sådana fall är det mer optimalt att dela in området i rutor (motsvarande SEV – selektiv efterbehandlingsvolym) och bedöma om halten överstiger risknivån för varje ruta åt gången. Man kan använda antingen s.k. *block scale approach* eller *site scale approach* – se Purucker and Stewart (2004) eller utgå från att ingen selektiv bedömning är möjlig och att riskkriteriet skall relateras till en statistisk parameter representativ för haltpopulation inom hela området.

## 7 Osäkerheter i ekologisk riskbedömning

### 7.1 Problemformulering

Principerna för ekologisk riskbedömning i Sverige bygger till stor del på nederländska ekotoxikologiska riktvärden (Verbruggen et al. 2001) och är utformade för att uppskatta effekter i mark- och ytvattenmiljön. Generella riktvärden för KM (Känslig Markanvändning) och MKM (Mindre Känslig Markanvändning) motsvarar ca 75% resp. 50% skydd av arter. Riktvärdena är behäftade med stora osäkerheter, som bland annat beror på att lite hänsyn tas till det faktum att olika arter reagerar olika på samma ämne och att sambandet mellan halten förorening och dödlighetskvoten (*mortality-index*) sällan är linjärt och svårt att kvantifiera.

I en traditionell ekotoxikologisk riskbedömning antar man att en recipient (t.ex. djur, fisk, maskar eller mikroorganismer) inte följer några specifika vandringsmönster utan att de är stationära och mer eller mindre jämnt fördelade över området, som misstänks vara förorenat. Ett sådant antagande förenklar avsevärt riskbedömningsprocessen, men kan samtidigt resultera i en signifikant över- eller underskattning av den faktiska risknivån. Om man föreställer sig att det kan finnas vissa platser, där en recipient föredrar att vistas, och att samtidigt just de platserna råkar vara kraftigt förorenade, blir den faktiska exponeringen klart mer allvarlig än om man antar att recipienten befinner sig lika lång tid inom förorenade och rena zoner. Figur 7.1 exemplifierar två hypotetiska vandringsmönster, A och B, för en djurart inom ett område med en föroreningsplym i mitten.



Figur 7.1. Två hypotetiska vandringsmönster, A och B, för en djurart inom ett område med en föroreningsplym. Individerna som följer mönster A kommer att utsättas för mer exponering från föroreningen jämfört med den som följer mönster B.

Under förutsättning att två likartade recipienter har samma konstanta förflyttningshastighet kommer uppenbarligen recipienten som förflyttar sig efter mönster A att bli mer exponerad för föroreningen jämfört med den som följer mönster B.

Forskning om vandringsmodeller för däggdjur har vunnit ökande uppmärksamhet inom ekologiska vetenskaper på senare tid främst i USA mycket tack vare den snabba utvecklingen av avancerade GIS-verktyg. Möjligheten att simulera migrationsmönster i relation till identifierade föroreningskällor har resulterat i en mer fullständig skattning av exponering och risker för känsliga populationer.

Studier av korrelationseffekter mellan rumslig fördelning av föroreningar och djurs förflyttningmönster (*foraging activity*) på exponering har bland annat rapporterats av Suter (1993), Freshman and Menzie (1996) och Engen et al. (2002).

I föreliggande rapport presenteras effekten av osäkerheter i gaffelantilopers vandringsmönster (*Antilocapra americana*; se figur 7.2) på dess exponering för en fluoridförorening. Studien bygger på numeriska simuleringar.



Figur 7.2. Gaffelantilop (*Antilocapra americana*).

## 7.2 Tillämpning inom riskbedömning i Sverige

Konceptet, där en exponeringsmodell tar hänsyn till hur recipientens vandringsmönster förhåller sig till rumslig fördelning av förorening, är mest lämpat för ”mobila” recipienter och i situationer, där föroreningsspridningen är relativt stor. Från de studier som hittills har gjorts, främst i USA, är det svårt att dra några konkreta slutsatser om hur denna typ av riskbedömning kan vara användbar för att kvantifiera risker för ytvattenrecipienter och marklevande organismer, som den svenska riskmodellen har som utgångspunkt.

Ekologisk riskbedömning är dock en relativt ny disciplin och kommer under den närmaste framtiden förmodligen att utvecklas i takt med att nya koncept prövas och ny kunskap tillförs. Det är därför viktigt att utforska alternativa idéer även om

de inte kan implementeras direkt i den modellen som används i Sverige – det kan bidra till en bredare förståelse av hur osäkerheter kan analyseras och vilka verktyg som kan användas för att kommunicera dessa i beslutsprocessen.

Arbetet som presenteras här bör betraktas som ett sätt att problematisera ekologisk riskbedömning utifrån ett nytt perspektiv och som idégivare till den framtida utvecklingen inom gebitet.

## 7.3 Metodik

I detta kapitel beskrivs hur osäkerheter i exponeringen av fluoridförening på gaffelantilop har simulerats med en sannolikhetsbaserad riskmodell, vilka indata som använts, vilka antaganden som gjorts och vilka verktyg som brukats.

Nyckelparametern, som har simulerats här, är den daglig dos fluorid som den mobila receptorn exponerats för. För att kunna ta hänsyn till och undersöka potentiell rumslig variation i exponeringen krävs indata/kunskap om:

- *Kvalitetsindex (habitat quality information)* som definierar areor/områden utifrån deras ”attraktivitet” för receptorn (gaffelantilopen), dvs. tillgång till föda, väderförhållanden, terräng, skydd mot vind etc.
- Exponeringsmodell byggd utifrån receptorns vandringsmönster.
- Rumslig utbredning av föreningen (fluorid).

Metodiken bygger på *stokastiska* simuleringar, som utförs enligt *Monte Carlo* principen, se t.ex. US EPA (2001). Den dagliga exponeringsdosen beräknas utifrån den mängd förorening (fluorid: F), som recipienten får i sig främst via intag av föda och direkt intag av jord. Mängden F gaffelantilopen exponeras för beror i sin tur på hur långa tidsperioder den uppehåller sig på platser, där halten F i föda och jord är höga. Områden med höga halter F kan råka sammanfalla med platser, där också *kvalitetsindex* är högt, t.ex. på grund av god tillgång till bete, bra skydd mot vind m.m., varför uppehållstiden för gaffelantilopen blir lång. Detta innebär hög exponering. På platser där F-halterna är låga och sammanfaller med lågt ett *kvalitetsindex* blir exponeringen avsevärt lägre. Sambandet i rumslig utbredning mellan *kvalitetsindex* och föroreningshalter är alltså av stor betydelse. Hur gaffelantilopen förflyttar sig, var den kommer att vistas längre eller kortare perioder, styrs av dess vandringsmönster som då är direkt beroende av den rumsliga fördelningen av platser med olika *kvalitetsindex*. I *Monte Carlo*-simuleringar skapas ett stort antal möjliga utfall, där varje utfall representerar en möjlig vandringsträcka under en simulerad tidsperiod. Resultatet blir en beräkning av den totala exponeringen för en individ. Genom att ställa samman alla individer (alla simuleringar) bygger man en empirisk fördelning som beskriver den genomsnittliga exponeringen för en hypotetisk gaffelantilopspopulation. Utifrån den beräknade empiriska fördelningen kan man studera osäkerheter i den prognostiserade exponeringen och ställa den i relation till ett aktuellt gräns-/referensvärde.

### 7.3.1 Kvalitetsfaktor

Kvalitetsfaktorn definieras utifrån recipientens preferenser med avseende på resurser nödvändiga för överlevnad och reproduktion. Djur dras till ställen med bra tillgång på föda och vatten men också till områden där de kan skydda sig mot rovdjur. Gaffelantiloper kommer att föredra bra betesmark framför ökenlandskap. Den mest dominerande exponeringsvägen blir direkt upptag via förorenade växter och jord. Upptag av förorenat ytvatten betraktas inte som en signifikant exponeringskälla, eftersom gaffelantiloper tillgodoser sitt vätskebehov framförallt via konsumtion av växter.

I denna studie används så kallad *HSI (Habitat Suitability Index)* som *kvalitetsindex*, vilket framtagits för flera mobila djurarter bl.a. för gaffelantilop (Allen et al. 1984). *HSI* varierar mellan 0 och 1, där noll innebär lågt kvalitetsindex och ett står för optimal kvalitet. *HSI* användes ursprungligen för att studera populationsdensitet och dynamik. Försiktighet anbefalls vid användandet av *HSI* för framtagandet av vandringsmodeller pga att variablerna i *HSI* ofta beskriver populationsdensitet väl utan att ange sambanden med förflyttningsmönstren.

Vanligtvis samlas indata för *HSI* med hjälp av GIS-teknik. Informationen verifieras sedan med observerade vandringsmönster för en given art. Det bästa ur verifieringssynpunkt är att förse ett antal individer från en population av intresse med sändare, som möjliggör att spåra deras rörelser inom ett geografiskt område och samtidigt relatera uppehållstiden till information om växtligheten, förekomst av vattendrag, topografi, temperaturfluktuationer m.m. På det viset kan man identifiera ställen som prefereras mer än andra och till slut undersöka vilka faktorer som gör att populationen kommer att dras till just dessa ställen.

För att studera hur stora osäkerheter det kan vara fråga om, när *kvalitetsindex inte inkluderas* i en riskbedömning, har här ett stort antal *HSI*-scenarier simulerats fram inom ett större område. Detta material har sedan använts vidare som indata för att ta fram en vandringsmodell för gaffelantiloper.

### 7.3.2 Exponeringsmodell och vandringsmönster

Intag av fluorid via upptag av växter och direkt intag av jord representeras med daglig dos enligt ekvationen nedan:

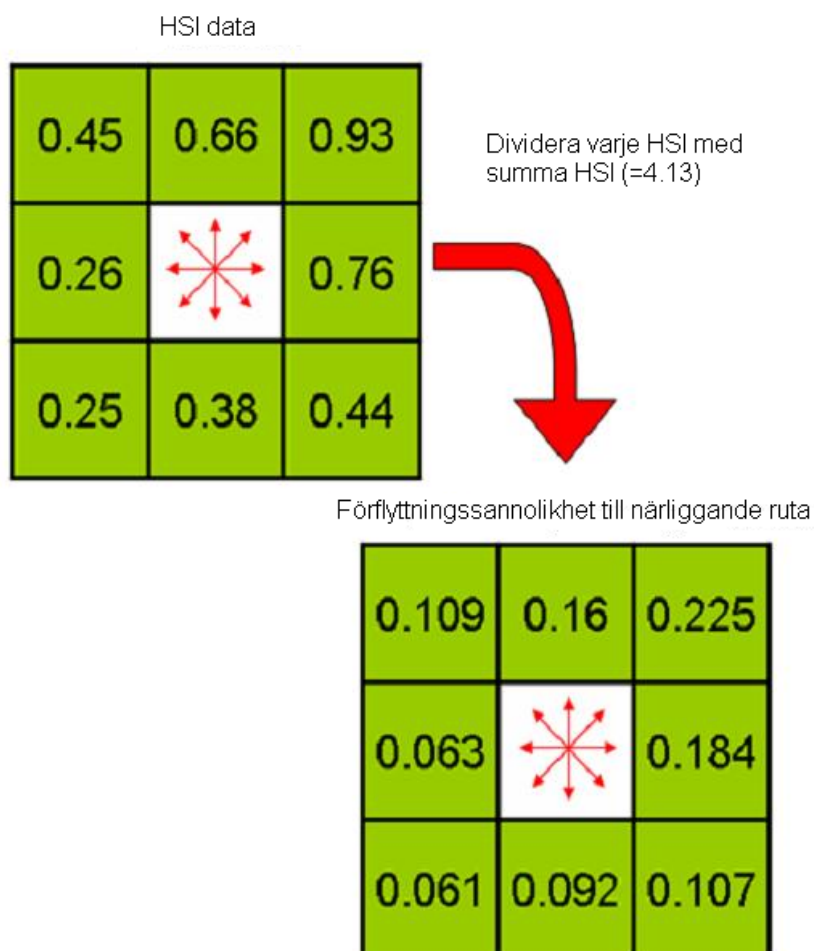
$$Dos = \sum_{t=1}^T \frac{(C_t I_f + \frac{C_t}{B_v} I_s) A}{T}$$

där  $C_t$  är halt fluorid i föda (växter),  $I_f$  är dagligt intag av växt,  $B_v$  är jord-växt-koncentrationsfaktor,  $A$  är absorptionsfaktor,  $I_s$  är dagligt intag av jord,  $T$  är antalet dagar.

All indata för beräkning av fluoriddosen är summerade hos Purucker et al. (2007) och presenteras inte i denna rapport. För numeriska simuleringar antas 40 kg vara en genomsnittsvikt för gaffelantiloper, intag av föda (växter) är

ca 0.85 kg/dag och det direkta intaget av jord är 49 g/dag. Det *TRV* (toxikologiska referensvärde) som användes i denna studie var 760 mg/kg/dag.

Gaffelantiloper lever i grupper om upp till 1000 individer under höst och vinter och bryter upp i könssegregerade grupper under vår-sommar. Dagliga förflyttningar varierar mellan 0.1 - 0.8 km. Simulering av vandringsmönster har i denna studie gjorts inom ett område av ca 310 km<sup>2</sup> indelat i rutor om 0.5 x 0.5 km. I vandringsmodellen antar man att gaffelantilopen betar i en ruta under en dag och nästa dag flyttar till en av 8 angränsande rutor. Vilken ruta som kommer att prefereras att flytta till beror på vilket det mest sannolika valet är, vilket bestäms av hur högt kvalitetsindex (*HSI*) de närliggande rutor har tilldelats. Figur 7.3 illustrerar principen för skattning av sannolikheter för valet av nästa ruta utifrån kvalitetsindex (*HSI*).



Figur 7.3. Schematisk presentation av omräkning från kvalitetsindexkarta (HSI index) till karta över förflyttningssannolikhet för gaffelantiloper.

Exponering för hela den hypotetiska gaffelantilopspopulationen bygger på simuleringar av hur 1000 individer flyttar mellan rutor under T = 90 dagar (sommarperioden), med en slumpad startposition inom området för varje individ.

Simuleringar avser honor och kalvar, som i motsats till hanar förflyttar sig fritt. Hanar uppvisar ett annat förflytningsmönster mer betingat av deras revirhärskande beteende.

I vandringsmodeller kan därför flera olika typer av migration förekomma:

- Icke-korrelerad så kallad *random walk*, där arter antas förflytta sig i skutt med konstant längd och tid. Efter varje skutt väljs nästa destination på ett slumpmässigt sätt.
- Korrelerade *random walk*, där det är mer sannolikt att individer kommer att välja nästa skutt med samma (eller nära) förflytningsriktning som ”skutten” före, dvs. det finns ett samband (korrelation) mellan närliggande skutt med avseende på förflytningsriktning.
- *Non-random walk*, där valet av nästa skutt är betingat av faktorer/förhållanden, som är icke-slumpmässiga och har potential att styra individers rörelser

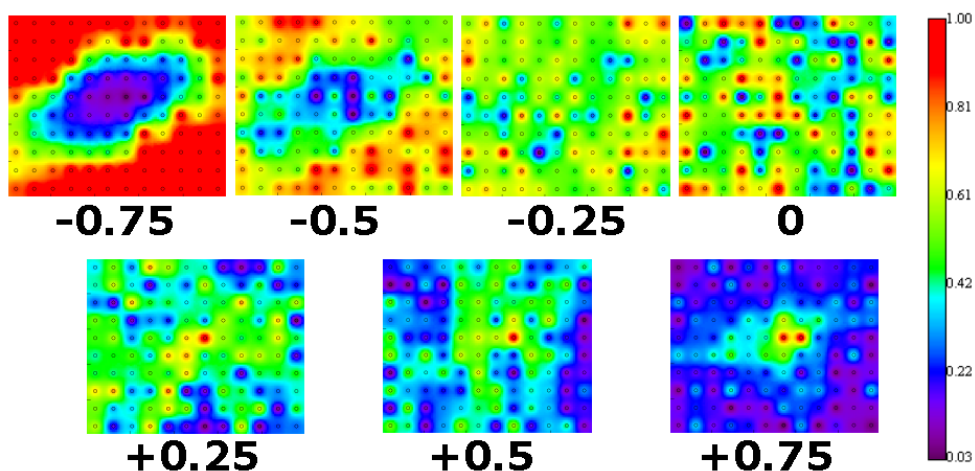
I detta arbete har man tillämpat *non-random walk* modell där valet av varje skutt styrs av sannolikheten att ”beträda” en specifik ruta räknad utifrån *HSI* data (figur 7.3).

### 7.3.3 Fluorförening korrelerad med kvalitetsindex

I kapitel 6 har vi illustrerat hur osäkerheter i en förorenings rumsliga utbredning påverkar kvaliteten på underlaget för beslutsfattande. De aspekterna upprepas inte i detta kapitel, utan fokus skiftas till att studera om det har någon betydelse att kombinera information om föroreningsutbredning och kunskapen om recipientens vandringsmönster i strävan att uppnå tillförlitligare exponeringsmodeller och i sin tur en mer trovärdig riskbedömning.

Kriginginterpolation har använts för att ta fram en karta över fluorid inom området. Fluoridföreningen kommer från luftburen deposition från en närliggande tillverkningsindustri. Observera att detta är samma område för vilket *HSI*-data simulerades med *Monte Carlo* och beskrevs i förra kapitlet. Indata för interpoleringen bestod av uppmätta halter i *desert sagebrush* (*Artemisia tridentata*) – en växt från solrosfamiljen, som är basföda för gaffelantiloper under sommarperioden. Prover togs i ett systematiskt ruttmönster över en yta på drygt 400 km<sup>2</sup> (Rope et al. 1988).

I syfte att undersöka den kombinerade effekten som kvalitetsindex och förorenings rumsliga utbredning har på exponeringsdosen, har man simulerat kartor med kvalitetsindex för olika korrelationsgrad mellan *HSI*-data och halten fluorid. Figur 7.4 visar kartor, där *HSI* är negativt, neutralt och positivt korrelerad med fluoridkoncentration i *sagebrush* inom området.

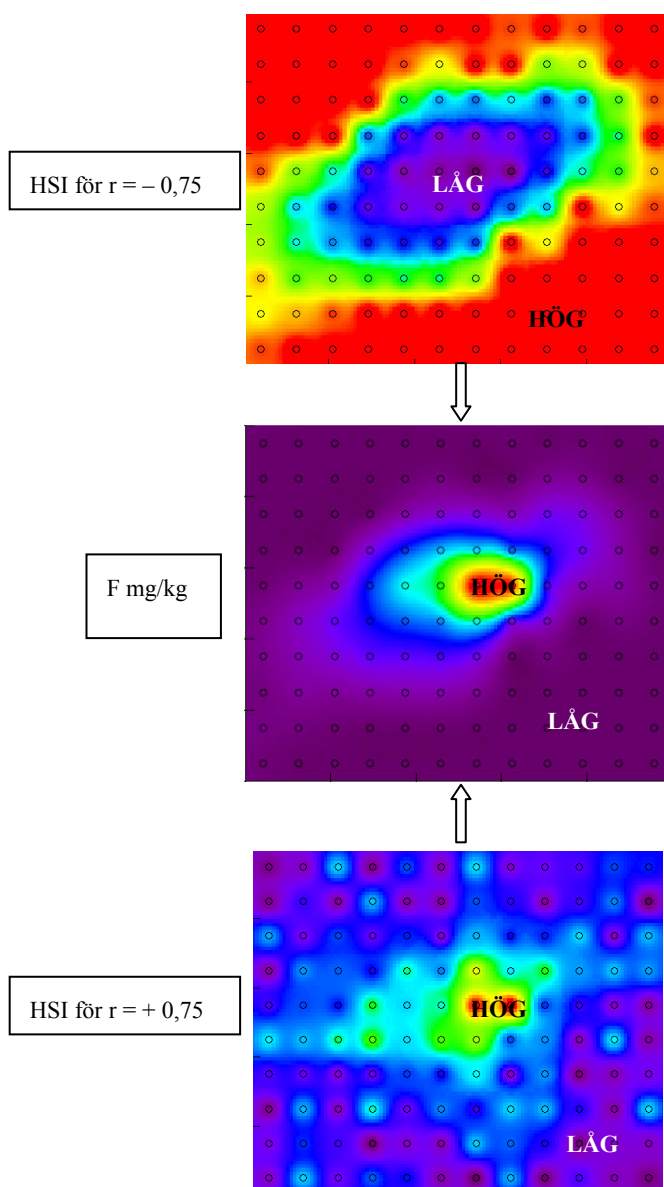


Figur 7.4. Kartor med simulerad kvalitetsindex (HSI) för olika korrelationsgrad mellan HSI-data och halten fluorid (dess rumsliga utbredningen). HSI varierar mellan 0 och 1.

För hög positiv korrelation (+0,75) är höga värden för interpolerade halter av fluorid tilldelade höga värden av *HSI*, medan hög negativ korrelation (-0,75) innebär att höga fluoridhalter ”matchas” med låga *HSI*. Kartorna i figuren är syntetiska (gjorda med datorsimuleringar) och har till syfte att illustrera vilka möjliga konstellationer som kan uppstå i verkligheten om man förfogar över kunskapen om rumslig utbredning av *HSI* och kan kombinera den med en karta över föroreningsutbredning.

Anta att *HSI*-data uppvisar en signifikant negativ korrelation (-0,75 i figur 7.5) med halter fluorid. I praktiken kommer det att innebära att när *HSI*-kartan kombineras med kartan över föroreningshalten och ett vandringsmönster, som styrs av *HSI*, tillämpas blir den genomsnittliga exponeringen inom hela området lägsta möjliga: hög *HSI*, där recipienten föredrar att vistas länge, kombineras med låg halt fluorid, samtidigt är platser med låg *HSI* förorenade med höga halter F – dvs. recipienten är exponerad för höga doser under kort tid och uppehåller sig längre på platser med låga halter fluorid. För situationen där *HSI* korrelerar positivt med fluoridhalterna (se +0,75 i figur 7.5) blir den genomsnittliga exponering högst av alla: Platser med hög *HSI* sammanfaller med höga halter förorening och låg *HSI* ”matchas” av låga fluorid halter – konsekvensen är en hög exponering under lång tid och liten exponering under kort tid.





Figur 7.5. Illustration av hur underlag till exponeringsmodellen tas fram. Kartan i mitten (representerar halter av fluorid i mg/kg; röd = höga halter, lila = låga halter) kombineras med kartan över kvalitetsindex (två HSI kartor med negativ och positiv korrelation mellan HSI och halter fluorid presenteras).

## 7.4 Resultat

För varje fall av korrelation (scenario) mellan *HSI* och fluoridkoncentration har 1000 *Monte Carlo*-simuleringar utförts. Varje simulering motsvarar ett möjligt vandringsmönster  $i$  ( $i = 1, \dots, 1000$ ) för en gaffelantilopindivid under  $T = 90$  dagar. Simuleringarna återspeglar alltså hur en population av 1000 honor och kalvar kommer att förflytta sig mellan rutor om  $0,5 \times 0,5$  km under sommarperioden och hur varje individ i sin tur kommer att exponeras för fluorid. Utöver simuleringar för åtta olika korrelations samband har samma antal simuleringar för vandringsmönstret *random walk* gjorts. Resultaten räknas om till en genomsnittlig exponeringsdos

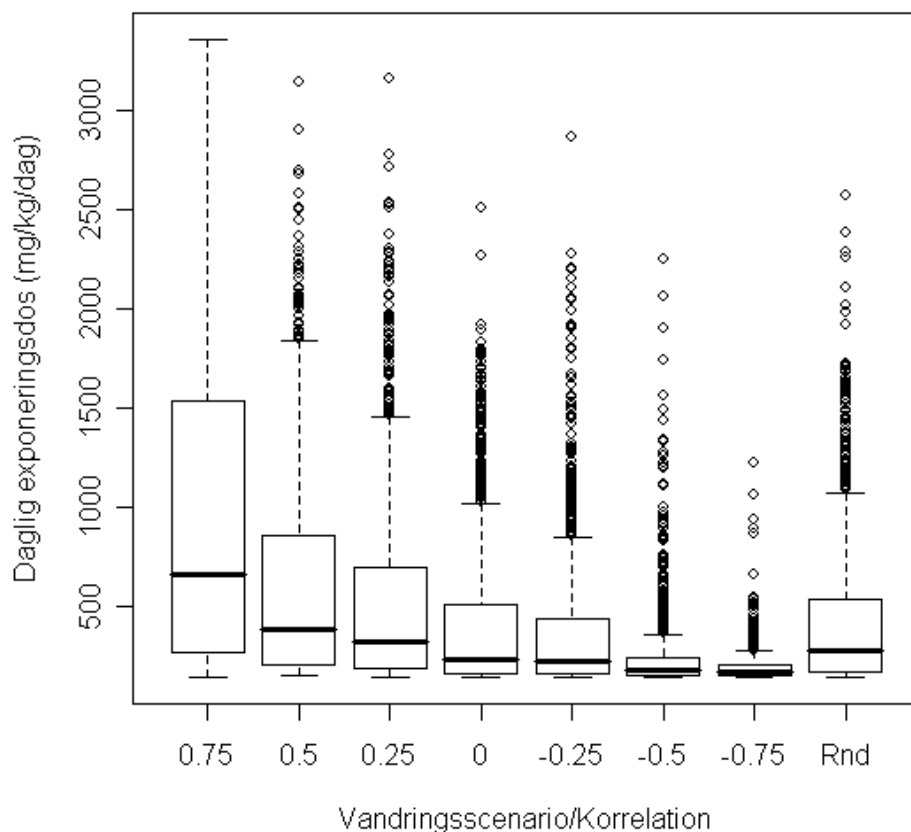
representativ för en hypotetisk gaffelantilopspopulation, som därefter jämförs med  $TRV = 760$  mg/kg kroppsvikt/dag. Det kompletta simuleringsutfallet sammanfattas i tabell 7.1.

**Tabell 7.1. Simulerad exponering av fluorid för gaffelantilop för åtta olika scenarier (sju olika korrelationer mellan HSI och rumslig utbredning av fluorid-halter, samt ett random walk-scenario). Referensdos = 760 mg/kg/dag**

Scenario	Median mg/kg/dag	Medel mg/kg/dag	Standard avvikelse mg/kg/dag	Variationsko- efficient	% över- skridande av referensdos
r = 0,75	664	958	811	0,85	46
r = 0,5	381	621	558	0,90	29
r = 0,25	319	543	521	0,96	23
r = 0	230	424	408	0,96	17
r = -0,25	221	379	371	0,98	12
r = -0,5	177	251	221	0,88	3,4
r = -0,75	166	194	86	0,44	0,5
Random Walk	275	439	389	0,89	18
Medelexpone- ring	-	417	-	-	under
Chebyshev UCL <sub>95</sub>	-	649	-	-	under

För jämförelsens skull kompletteras simuleringsresultaten i tabellen med exponeringsdos motsvarande den uppmätta medelhalten fluorid på området samt exponering, som motsvarar UCL<sub>95</sub> för medelhalt beräknad enligt *Chebyshev*-metoden. Med hänvisning till kapitel 5 används Chebyshev UCL här, eftersom fluoriddata visade sig vara icke-normalfördelade med en tydlig positiv skevhet. Exponeringsdos baserad på UCL<sub>95</sub> skall enligt US EPA (2002) tillämpas som en representativ risknivå i situationer, där ingen rumslig analys görs och osäkerheten justeras uppåt. Därigenom erhålls en *deterministisk* och konservativ riskuppskattning. Syftet med att i tabellen presentera både simulerad medelexponering och UCL är att jämföra deterministisk och probabilistisk riskuppskattning med hänsyn till ett referensvärde (*TRV*).

Både exponeringen beräknad utifrån medelhalten fluorid och *Chebyshev* UCL<sub>95</sub> understiger referensdosen (jämför 417 mg/kg/dag och 649 mg/kg/dag med 760 mg/kg/dag). Den simulerade exponeringsdosen för olika scenarier har utvärderats så att för varje scenario (åtta korrelationer samt *random walk*-modellen) har man byggt en kumulativ statistisk fördelning av simulerade doser och därefter beräknat hur stor procentsats av fördelningen som hamnar över referensdosen. Från sista kolumnen i tabellen ovan kan utläsas att det är positiva korrelationer mellan *HSI* och halten fluorid, som resulterar i det största överskridandet av referensdosen, medan låga korrelationer ger betydligt lägre exponering. Det skall också noteras att *random walk*-scenariet och scenariet med nollkorrelation inte avviker mycket från varandra. Figur 7.6 ger en visuell inblick i osäkerheterna i exponeringsdosen för varje scenario i sig och i relation till scenarierna emellan.



Figur 7.6. Simulerat osäkerhetsspann över exponeringsdosen för gaffelantiloper och för olika scenarier, dvs. olika korrelationer mellan HSI-index och den rumsliga fördelningen av fluoridhalten.

Simuleringarna visar att skillnaden i medel exponering mellan extrema negativa och positiva korrelationer kan vara betydande (ta t.ex. kvoten mellan medel exponering för  $r = +0,75$  och för  $r = -0,75$  som är 400%). Också osäkerheten mellan olika scenarier varierar: Största osäkerheten (storleken på boxar i figuren ovan åter speglar 95% konfidensintervall för medel exponering) har konstaterats för  $r = +0,75$  och minsta för  $r = -0,75$ .

## 7.5 Slutsatser

I likhet med hälsoriskbedömning kan osäkerheter i ekologisk riskbedömning vara betydande och ha flera olika källor. Som i föregående kapitel har fokus även här varit på osäkerheter kring exponering från en föroreningskälla.

Metodiken, som har presenterats och testats på gaffelantilop, är mest lämplig för icke-stationära recipienter och i de fall där föroeningen är av en storskalig karaktär. Den största begränsningen i att kombinera kunskapen om recipientens förflyttningsmönster och rumslig utbredning av föroeningar för framtagande av en realistisk exponeringsdos är att det kan vara resurskrävande att ta fram nödvändig information om vilka platser recipienten kommer att uppehålla sig på under längre

tid och vilka kommer den att undvika. *Kvalitetsindex (HSI)* kan finnas tillgängligt för vissa känsliga och prioriterade (utrotningshotade) populationer, men det är inga standarddata som samlas in rutinmässigt för ekologisk riskbedömning.

Spatial ekologisk riskbedömning har inte använts i Sverige mest pga. att konceptet inte har uppmärksammats tidigare, men också av det skälet att föroreningsproblem är av en mindre skala än exempelvis i USA, samt att i stort sett de riskmodeller som används utgår från risker för ytvatten- och markorganismer och inte från risker för t.ex. däggdjur. Det presenterade konceptet innehåller dock innovationer, som kan aktualiseras i fråga om implementeringsmöjligheter i framtida riskmodellering i takt med utvecklingen av riskanalyser utanför de ramar som inringar det nuvarande riskkonceptet i Sverige.

De slutsatser som kan dras från föregående studier om exponering av fluorid för gaffelantilop och som är av mer generell karaktär är:

- Osäkerheter i skattning av exponeringsdos pga. okunskap om recipiентовs vandringmönster i relation till rumslig utbredning av en förorening kan vara betydande (upp till 400% skillnad mellan genomsnittsexponering för olika scenarier i denna studie).
- Om det finns signifikant positiv eller negativ korrelation mellan kvalitetsindex och föroreningskoncentration är en bra idé att införa en exponeringsmodell som tar hänsyn till individers rörelsemönster.
- Att inkludera inslaget från *HSI* och rumslig förorenings-spridning kan ha viktiga implikationer, som även sträcker sig utanför riskbedömningsprocessen. Om saneringsåtgärder är aktuella så kan saneringsarbetet effektiviseras genom att man, i stället för att sanera hela det förorenade området eller sanera de mest förorenade platserna/blocken, sanerar de platser/block där exponeringen är som högst. Saneringen genomförs blockvis och efter varje sanerat block räknas en ”ny” medalexponering fram som jämförs med uppsatt kriterium. Ett problem kan dock vara att saneringen i sig förstör habitatet och således utgöra en annan typ av risk för det specifika däggdjuret.

## 8 Rekommendationer

Riskbedömning är ett levande koncept och i takt med att arbetet kring metodutveckling, effektivare markundersökningar, toxikologiska databaser m.m. fortskrider bör konceptet kontinuerligt uppdateras med nya rön och erfarenheter. Betydelsen av att ta hänsyn till osäkerhetsaspekter i riskbedömningar har uppmärksamats på senare tid och svenska Naturvårdsverket har via kunskapsprogrammet *Hållbar Sanering* tagit initiativ till en vetenskaplig analys av hur osäkerheter skall inkluderas i bedömningsprocessen.

För att arbete/studier och annan verksamhet kring vetenskaplig utveckling och granskning också skall kunna implementeras i praktiska sammanhang och bidra till förbättrad kvalitet i riskbedömningar krävs det en attitydförändring hos i huvudsak problemägare och beställare i fråga om hantering av osäkerheter. Karaktärisering och kvantifiering av osäkerheter betraktas ofta som en separat del i riskbedömningen och i uppdragsformuleringen inkluderas den snarare som ett ”extra” tillägg än som en kravspecifikation. Riskbedömningar, som bygger på punktskattningsprincipen, är ofta konservativa, men kan också resultera i underskattning av riskerna. Avsaknad av osäkerhetsanalyser gör det omöjligt för beslutsfattare att undersöka storleken på konfidensen i resultatet, vilka de största källorna till osäkerheter är och hur resurskrävande en eventuell minskning av osäkerheter kan vara.

Ett regulatoriskt incitament skulle innebära ett formellt åtagande för riskbedömare att kvantifiera risknivån och dess osäkerhetsspann och vad det praktiskt innebär. Utifrån den här studien kan man konstatera att det finns ett stort behov av att upprätta ramar för tillvägagångssättet för karaktärisering av föroreningskällan och den kvantitativa jämförelsen mellan föroreningsgraden och uppställt riskkriterium. Kunskapen om vilka tester och metoder (hypotesprövning eller UCL, parametriska eller icke-parametriska tester, interpolering eller ej, m.m.) som är mest lämpliga för att uppskatta en realistisk variation i observerade halter av föroreningen och hur den förhåller sig i relation till t.ex. riktvärde bör utvecklas och systematiseras.

Sammanfattningsvis bör fortsatt arbete kring osäkerheter i riskbedömning inrikta sig på följande:

- Ta fram underlag till vägledning för statistisk kvantifiering av osäkerheter i förhållandet mellan föroreningsgraden (i källan) och uppställt riskkriterium (riktvärde).
- Anordna kurser i riskanalys med fokus på osäkerheter för yrkesverksamma (både konsulter och miljöhandläggare på länsstyrelser och kommuner).
- Initiera arbete kring formalisering av rekommendationer för genomförande av uppdrag/studier, som inkluderar fördjupad riskbedömning.

## 9 Referenser

- ADEC (2003). Use of the Bootstrap Method in Calculating the Concentration Term for Estimating Risks at Contaminated Sites. Technical Memorandum-01-004. A. D. o. E. Conservation.
- Allen, A. W., J. G. Cook and M. J. Armbruster (1984). Habitat suitability index models: Pronghorn. FWS/OBS-82/10.65. U.S. Fish and Wildlife Service.
- Apitz, S., L. Borbridge, G. Theriault and S. Lieberman (1992). "Remote in situ determination of fuel products in soils: field results and laboratory investigations." *Analysis* **20**(461-474).
- Box, G. E. P., W. G. Hunter and J. S. Hunter (1978). *Statistics for Experimenters*. Wiley and Sons.
- Davis, J. C. (1986). *Statistics and Data Analysis in Geology*. 2nd. New York, John Wiley & Sons.
- Engen, S., R. Lande and B. Saether (2002). "Migration and spatiotemporal variation in population dynamics in a heterogenous environment." *Ecology* **83**(2): 570-579.
- Freshman, J. S. and C. A. Menzie (1996). "Two wildlife models to assess impacts at the individual and population levels and the efficacy of remedial actions." *Human and Ecological Risk Assessment* **2**(3): 481-498.
- Gilbert, R. O. (1987). *Statistical Methods for Environmental Pollution Monitoring*. New York, John Wiley & Sons, Inc.
- Grandin, U. (2003). Dataanalys och hypotesprövning för statistikanvändare. opublicerat PM. Naturvårdsverket.
- Isaaks, E. H. and R. M. Srivastava (1989). *Applied Geostatistics*. New York, Oxford University Press.
- Journel, A. G. and C. J. Huijbregts (1978). *Mining Geostatistics*. London, Academic Press.
- Levine, D. M., P. P. Ramsey and R. K. Smidt (2001). *Applied Statistics for Engineers and Scientists. Using Microsoft Excel and MINITAB*. Upper Saddle River, New Jersey, Prentice Hall.
- Minitab, I. (2004). Minitab 14.
- Naturvårdsverket (1996). Development of generic guideline values. 4639. Naturvårdsverket. Stockholm.
- Naturvårdsverket (1997). Åtgärdskrav vid efterbehandling. Vägledning för att acceptabla resthalter och restmängder uppnås - metoder och säkerhet. NV-4918. Stockholm.

- Naturvårdsverket (2005a). Beräkningsmodell för riktvärden för mark. REMISSVERSION. 2005-07-04. Stockholm.
- Naturvårdsverket (2005b). Probabilistisk riskbedömning fas 1. Rapport 5532. Stockholm.
- Naturvårdsverket (2005c). Vägledning för riskbedömning av förorenade områden. REMISSVERSION. 2005-07-04. Stockholm.
- Nilsson-Påledal, S., P. Starzec and P. Karlsson (2003). Vidare utveckling av FFD-sonden. Laboratorieundersökning av mätosäkerheter och detekterbara ämnen. Varia 525. Statens geotekniska institut. Linköping.
- Pulsipher, B. A., R. O. Gilbert and J. R. Wilson (2003). Measurement uncertainty in Visual Sample Plan (VSP). PNNL-SA-38977. Pacific Northwest National Laboratory. Richland, WA.
- Purucker, S. T., C. J. E. Welsh, R. N. Stewart and P. Starzec (2007). "Use of habitat-contamination spatial correlation to determine when to perform a spatially explicit ecological risk assesemnt." *Ecological Modelling* **204**: 180-192.
- Purucker, T. and R. Stewart (2004). Spatial Analysis and Decision Assistance. Knoxville, TN, The University of Tennessee Research Corporation.
- Rope, S. K., W. J. Arthur, T. H. Craig and E. H. Craig (1988). "Nutrient and trace elements in soil and desert vegetation of southern Idaho." *Environmental Monitoring and Assessment* **10**: 1-24.
- Suter, G. W. (1993). *Ecological Risk Assessment*. Boca Raton FL, Lewis Publishers.
- Swan, A. R. H. and M. Sandilands (1995). *Introduction to Geological Data Analysis*. Oxford, Blackwell Science.
- US EPA (1989). Risk Assessment Guidance for Superfund. Volume I: Human Health Evaluation Manual (Part A). EPA/504/1 -89/002. Office of Emergency and Remedial Response, U. S. E. P. Agency. Washington DC.
- US EPA (1991). Risk Assessment Guidance for Superfund. Volume I: Human Health Evaluation Manual Supplemental Guidance standard Default Exposure Factors. OSWER Directive 9285.6-03. Office of Emergency and Remedial Response.
- US EPA (1996). Soil Screening Guidance: Technical Background Document. EPA/540/R95/128. O. o. E. a. R. Response. Washington DC.
- US EPA (1997a). Guiding Principles for Monte Carlo Analysis. EPA/630/R-97/001. U.S. EPA. Washington DC.
- US EPA (1997b). Guiding Principles for Monte Carlo Analysis. Risk Assessment Forum. EPA/630/R-97/001.

US EPA (2000a). Guidance for Data Quality Assessment: Practical Methods for Data Analysis. EPA QA/G-9. EPA/600/R-96/084. Office of Environmental Information. Washington DC.

US EPA (2000b). Guidance for the Data Quality Objective Process. EPA QA/G-4. Office of Environmental Information. Washington DC.

US EPA (2001). Risk Assessment Guidance for Superfund: Volume III - Part A, Process for Conducting Probabilistic Risk Assessment. EPA-540-R-02-002. O. o. E. a. R. Response. Washington DC.

US EPA (2002). Calculating upper confidence limits for exposure point concentrations at hazardous waste sites. OSWER 9285.5-10. O. o. E. a. R. Response. Washington DC.

US EPA (2006). Data Quality Assessment: Statistical Methods for Practitioners. EPA QA/G-9S. EPA/240/B-06/003. Office of Environmental Information. Washington DC.

Verbruggen, E. M. J., R. Posthumus and A. P. Wezel (2001). Ecotoxicological serious risk concentrations for soil, sediment and (ground)water: updated proposals for first series of compounds. RIVM report 711701 020. National Institute of Public Health and the Environment. Bilthoven, Nederländerna.



# Osäkerheter i riskbedömning och beslutsprocess

RAPPORT 5804

NATURVÅRDSVERKET  
ISBN 978-91-620-5804-3  
ISSN 0282-7298

Rapporten ger en teoretisk bakgrund till osäkerheter i riskbedömning av förorenad mark. Osäkerheten belyses mer i detalj utifrån tre olika utgångspunkter; osäkerheter i jämförelsen mellan föroreningsgrad och uppställt riktvärde, osäkerheter i rumslig utbredning av förorening samt osäkerheter i ekotoxikologisk exponeringsmodell. Med praktiska exempel beskrivs och kvantifieras dessa osäkerheter och deras påverkan på beslut i riskbedömningen. Slutligen belyses inverkan på osäkerheterna av olika val och antaganden i utvärderingsmetodik och riskbedömning.

**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.