U.S. EPA Soil Screening Guidance

U.S. EPA Region 3/ORD Presentations

U.S. EPA Region 3
1650 Arch Street
Philadelphia, PA 19103

May 12 & 13, 1999

Instructors

Dr. David Kargbo
Ms. Nancy Rios-Jafolla
Ms. Bernice Pasquini
Ms. Patricia Flores-Brown
Dr. Anita Singh
Dr. A.K. Singh

Soil Screening Guidance Workshop Agenda

Day 1 (May 12, 1999)

Time	Topic	Presentor(s)
9:00-10:30 am	Overview of SSL Process; Technical Issues and Concepts in SSL Development	David Kargbo
10:30-10:45 am	BREAK	
10:45-11:15 am	Conceptual Site Model (CSM)	Nancy Rios-Jafolla
11:15-12:15 pm	Surface Soil Sampling and Statistics	Anita Singh
12:15-1:00 pm	LUNCH	
1:00-1:15 pm	Ingestion and Dermal SSL	Nancy Rios Jafolla
1:15-2:15 pm	Inhalation and Plant Uptake SSL; Calculated SSL vs. site concentrations	Pat Flores-Brown and Nancy Rios Jafolla
2:15-2:30 pm	BREAK	
2:30-3:15 pm	GW Tech Issues & SSL Development	Bernice Pasquini
3:15 –3:30 pm	Introduction of SSL Case Study	David Kargbo
3:30-4:00 pm	Question & Answer	All presenters

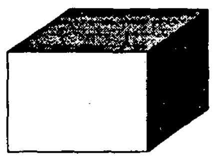
Day 2 (May 13, 1999)

9:00-10:30 am	SSL Case Study: Surface Soil Sampling and Statistics	Anita Singh
10:30-10:45 am	BREAK	
10:45-12:00 pm	SSL Case Study: Effect of SSL Parameters	Dave/Pat/Nancy/Bernice
12:00-1:00 pm	LUNCH	
1:00-2:30 pm	SSL Case Study: SSL Parameters (contd)	Dave/Pat/Nancy/Bernice
2:30-3:00 pm	Panel Discussions	All presenters

U.S. EPA SOIL SCREENING GUIDANCE:
A Technical Overview

by
David M . Kargbo, Ph.D.
Technical Support Section
HSCD, USEPA Region 3, Philadelphia

May, 1999



1. OVERVIEW

- Guidance Documents
- Purpose
- What are SSLs?
- SSL Framework
- When and Where to Use the Guidance
- Decision Process in SSL Determinations
- Contamination Spectrum/Risk Management
- Advantages of the Guidance
- Exposure Pathways
- Site-specific Approach

2. TECHNICAL SSL ISSUES AND CONCEPTS

- Contaminant Fate and Transport Issues
- Background Concentration
- Human Health Issues
- DQO Process
- Collecting Statistically Valid Soil Samples

2. TECHNICAL SSL ISSUES AND CONCEPTS (Continued)

- Soil-to-air Volatilization Factor (VF)
- Particulate Emission Factor (PEF)
- Soil Saturation Limit (Csat)
- Contaminant Dispersion in Air (Q/C term)
- Soil/Water Partition Equation

2. TECHNICAL SSL ISSUES AND CONCEPTS (Continued)

- Contaminant Dilution and Attenuation
- Risk-based SSLs and Mass-balance Violations
- Influence of pH on SSL Calculations
- Sensitivity Analysis

3. DATA REQUIREMENTS

- Source Characteristics
- Soil Characteristics
- Meteorological Data
- Hydrogeological Characteristics

OVERVIEW OF THE U.S. EPA SOIL SCREENING GUIDANCE

1.1 Guidance Documents

- Soil Screening Guidance: Technical Background Document (EPA/540/R-95/128)
- Soil Screening Guidance: User s Guide (EPA/540/R-96/018)

1.2 Purpose

- Standardize and accelerate evaluation and cleanup of contaminated soils
- Provide step-by-step methodology to calculate risk-based, site-specific, soil screening levels (SSLs)
- Provide SSLs in soil that may be used to identify areas needing further investigation at NPL sites.

1.3 What are SSLs?

- 1.3.1 SSLs are risk-based concentrations derived from equations that combine:

  a) Exposure point concentrations
  - measured
  - estimated
  - average concentrations
  - maximum concentrations

1.3 What are SSLs? (continued)

- *b) Chemical Characteristics*

  c) Site Characteristics

  d) EPA toxicity data.

1.3 What are SSLs? (continued)

- **1.3.2. Models and assumptions in SSL calculations are consistent with RME**

1.3 What are SSLs? (continued)

- 1.3.3 Site-specific estimate of RME compared with chemical specific toxicity criterion
  - Ingestion (SFo and RfDs)
  - Inhalation (URFs and RfCs)
  - Mig to GW (MCLGs, MCLs; and HBLs)

1.3 What are SSLs? (continued)

- 1.3.4 Exposure equations and pathways modelled in reverse

- 1.3.5 Potential for additive effects not built in

1.3 What are SSLs? (continued)

- **1.3.6 SSLs generally based on:**
  - **Health-based limits of 10E-06 risk for carcinogens**
  - **Hazard quotient (HQ) of 1 for noncarcinogens**
  - **Non-zero MCLGs, MCLs, or HBLs for migration to ground water**

1.4 SSL Framework and Key Assumptions

- **1.4.1 SSL Framework**
  - ***Tiers***
    - **Tier 1: Generic SSLs**
    - **Tier 2: Site-specific SSLs calculations**
    - **Tier 3: Models for detailed assessment**
  - ***Generic vs. Site-Specific SSLs***
    - **Generic SSLs more conservative than, and can be used in place of, site-specific SSLs**
    - **Caution: Using generic SSLs vs. generating site-specific SSLs**

1.4 SSL Framework and Key Assumptions (contd)

- **1.4.2 Key Assumptions**
  - ***Inhalation and migration to ground water SSL models are designed for use at the early stage of site investigation***
  - ***Source is infinite***

1.4 SSL Framework and Key Assumptions (contd)

- ***Other simplifying assumptions resulting from infinite source assumption***

1.5 When and Where to Use the Guidance

- 1.5.1. When Should the Guidance be Used?

  - ▲ When residential land use assumptions are applicable (but is being updated to be used at non-residential sites)

  - ▲ To determine whether contaminated soil areas warrant further investigation or response

1.5 When and Where to Use the Guidance (continued)

- ▲ State Programs

  - When States screening numbers more stringent than the generic SSLs

  - States may use Guidance in their voluntary cleanup programs

- ▲ Brownfields Program

1.5 When and Where to Use the Guidance (continued)

- SSLs as Action levels in RCRA program
- SSLs in Removal Actions
- SSLs as Preliminary Remediation Goals (PRGs)

1.5 When and Where to Use the Guidance (continued)

- 1.5.2 Why Use the Guidance?
  - Its a tool to facilitate prompt identification of contaminants and exposure areas of concern
  - Its primarily used during the early stages of a remedial investigation at NPL, Removal, and/or RCRA sites
  - It should not replace RI/FS or risk assessment

1.6 SSL Decision Process

- ***Data Interpretation***
  - ▲ Contaminant concentrations < Generic SSLs
    - No further action or study warranted under CERCLA
  - ▲ Contaminant concentrations < Calculated SSLs
    - No further action or study warranted under CERCLA
  - ▲ Contaminant concentrations = or > SSLs
    - further study or investigation, but not necessarily cleanup, is warranted

1.7 Contamination Spectrum and Range of Risk Management

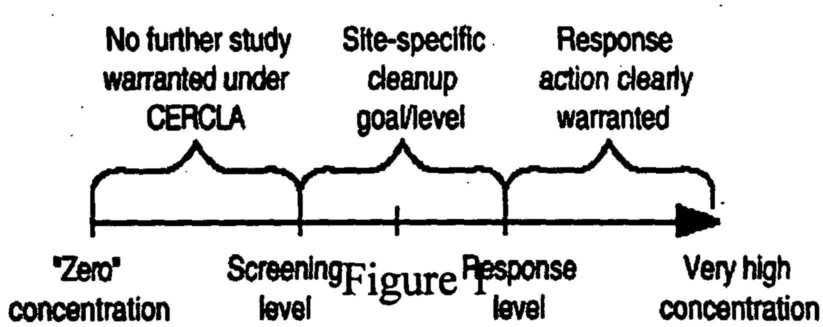


Figure 1

1.8 Advantages of the Guidance

- Standardizes SSL calculation process
- Simple to use
- Can can save resources
- Can save time for site remediation
- Standardizes site remediation process

1.8 Advantages of the Guidance (continued)

- Can be used in later Superfund phases
  - baseline risk assessment
  - feasibility study
  - treatability study
  - remedial design

1.9 Exposure Pathways

- **Quantitative Treatment**
  - **Direct ingestion**
  - **Inhalation of volatiles and fugitive dust**
  - **Ingestion of contaminated groundwater**

1.9 Exposure Pathways (continued)

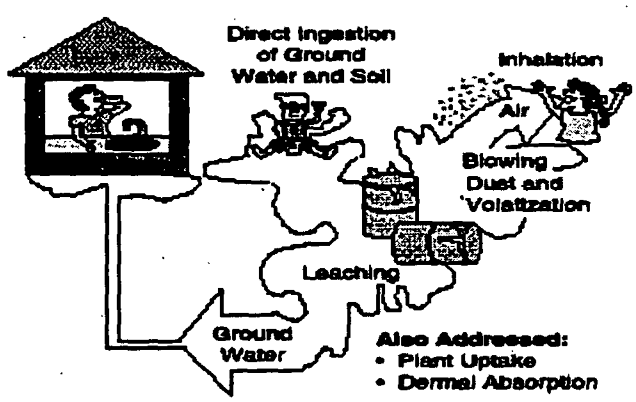


Figure 2

1.9 Exposure Pathways (Contd)

- **Semi-Quantitative Treatment**
  - **Dermal absorption**
  - **Ingestion of contaminated plant material**
  - **Migration of volatiles into basements**

1.9 Exposure Pathways (Contd)

- **Fish consumption**
- **Raising of livestock**
- **Fugutive Dust**

1.9 Exposure Pathways (Contd)

- Not Addressed
  - Ecological Concerns
  - Fish Consumption

1.10 Site-specific Approach

- Step 1: Develop a conceptual site model (CSM)
- Step 2: Compare the CSM to the SSL scenario
- Step 3: Define data collection needs
- Step 4: Sample and analyze soils at site

1.10 Site-specific Approach (contd)

- Step 5: Calculate site-specific SSLs
- Step 6: Compare site soil contaminant concentrations to calculated SSLs
- Step 7: Determine which areas of the site require further study

SIGNIFICANT TECHNICAL ISSUES AND CONCEPTS APPLICABLE TO THE SSL DEVELOPMENT PROCESS CONCEPTS

2.1 Contaminant Fate and Transport Issues

- **Soil Physical Properties**
  - texture
  - structure
  - soil density (particle, bulk)
  - soil porosity (air, water, total)
  - soil moisture

2.1 Contaminant Fate and Transport Issues (continued)

- **Aquifer Properties**
  - hydraulic conductivity
  - aquifer depth
  - disperssivity
  - infiltration/recharge
  - aquifer mixing

2.1 Contaminant Fate and Transport Issues (continued)

- **Chemical Properties and Reactions**
  - volatilization
  - dispersion (in air and water)
  - adsorption/desorption kinetics
  - ionization

2.1 Contaminant Fate and Transport Issues (continued)

- precipitation/dissolution
- cosolvation
- redox
- hydrolysis
- biodegradation

2.2 Background Concentrations

- Approach
- Avoiding clean islands
- Comparing background with generic SSL
- Comparing background with calculated SSL

2.3 Human Health Issues

- Additive Risk
  - For Carcinogens
  - For Non-carcinogens

2.3 Human Health Issues (contd)

- Apportionment

- Fractionization

2.3 Human Health Issues (contd)

- Acute Exposure
  - Major impediments to developing acute SSLs

2.3 Human Health Issues (contd)

- **Route-to-route Extrapolation**
  - **Ingestion SSL vs. Inhalation SSL**
  - **Extrapolated Inhalation SSLs vs. Generic SSLs**

2.4 DQO Process

DATA QUALITY OBJECTIVES

2.5 Collecting Statistically Valid Samples

2.6 Soil-to-air Volatilization Factor (VF)

- **defines the relationship between the concentration of contaminant in soil and the flux of the volatilized contaminant into the air.**
- **Old vs. New**

2.7 Particulate Emission Factor (PEF)

- **Relates the concentration of contaminant in soil to the concentration of dust particles in the air (i.e. windblown dust.)**

2.8 Soil Saturation Limit (Csat)

- **The concentration at which the emission flux from soil to air for a chemical reaches a plateau.**

2.9 Contaminant Dispersion in Air (Q/C term)

- **Q/C simulates dispersion of contaminants in ambient air**

2.10 Soil/Water Partition Equation

- **Definition**
- **Used in Migration to Groundwater Pathway**

2.11 Contaminant Dilution and Attenuation

- **Dilution factor**

- **No attenuation**

2.12 Risk-based SSLs and Mass-balance Violations

-
- **2.12.1 Source depletion time**
  - **chemical volatility**
  - **chemical solubitity**
  - **size of contaminant source**
- **Options for addressing problem**

2.13 Influence of pH on SSL Calculations

- Ionizing organics and metals
- pH-specific Koc values

2.14 Sensitivity Analysis on SSL Parameters

- Sampling Methodologies
- Contaminant Depth

2.14 Sensitivity Analysis on SSL Parameters (continued)

- Contaminated Area
  - Q/C
  - DF
- Location
- Soil pH

DATA REQUIREMENTS IN SSL DETERMINATIONS

3.1 Source Characteristics

- Source Area (A)
- Source Length (L)
- Source Depth

3.2 Soil Characteristics

- Soil texture
- Soil dry bulk density
- Soil moisture
- Soil organic carbon
- Soil pH

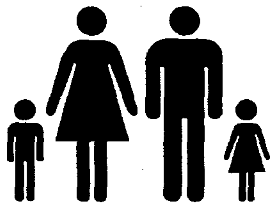
3.3 Meteorological Data

- Air dispersion factor (Q/C term)
- % Vegetative Cover (V)
- Mean Annual Windspeed (Um)
- Equiv. Windspeed at 7 m (Ut)
- Fraction dependent on Um/Ut

3.4 Hydrogeological Characteristics

- Hydrogeologic setting
- Infiltration/recharge rate (I)
- Hydraulic conductivity (K)
- Hydraulic gradient (i)
- Aquifer thickness (d)

Soil Screening Guidance
Step-by-Step Approach
Risk Assessment



by
Nancy Rios Jafolla,Toxicologist
Technical Support Section
HSCD, USEPA Region 3, Philadelphia
May, 1999

Soil Screening Process
Step-by-Step Approach

1. Developing a conceptual site model (CSM)
2. Comparing the CSM to the SSL scenario
3. Defining data collection needs
4. Sampling and analyzing soils at the site
5. Calculating site-specific SSLs
6. Comparing site soil contaminant concentrations to calculated SSLs
7. Determining which areas of the site require further study

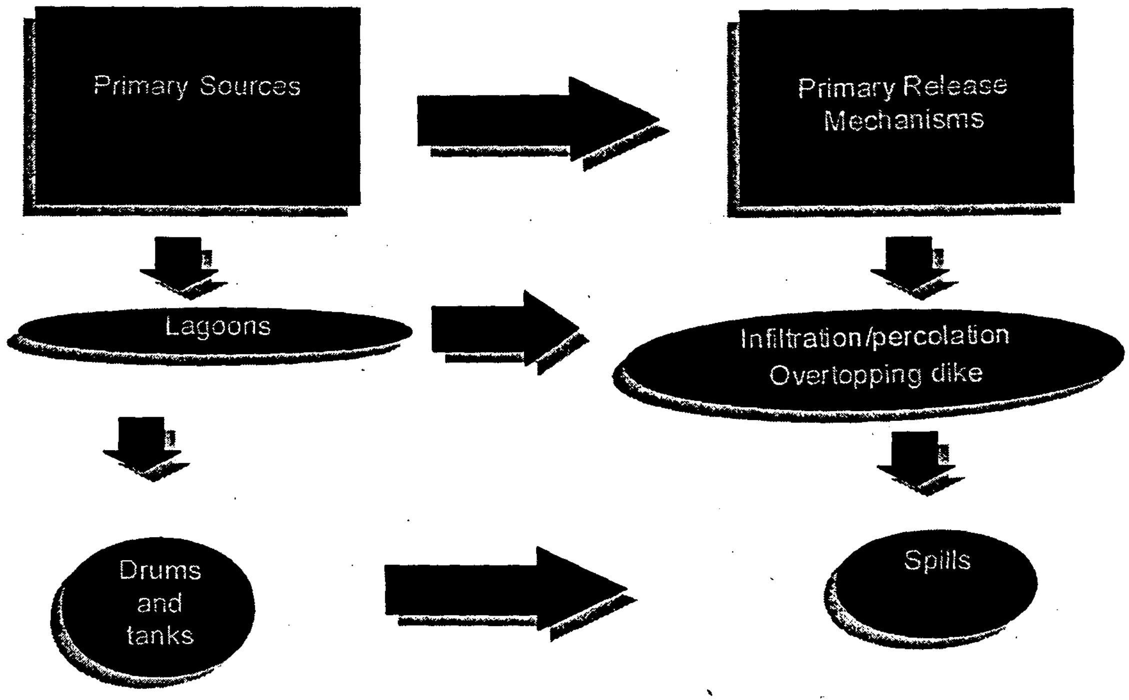
Soil Screening Process Step-by-Step Approach

1. Developing a conceptual site model (CSM)
2. Comparing the CSM to the SSL scenario
3. Defining data collection needs

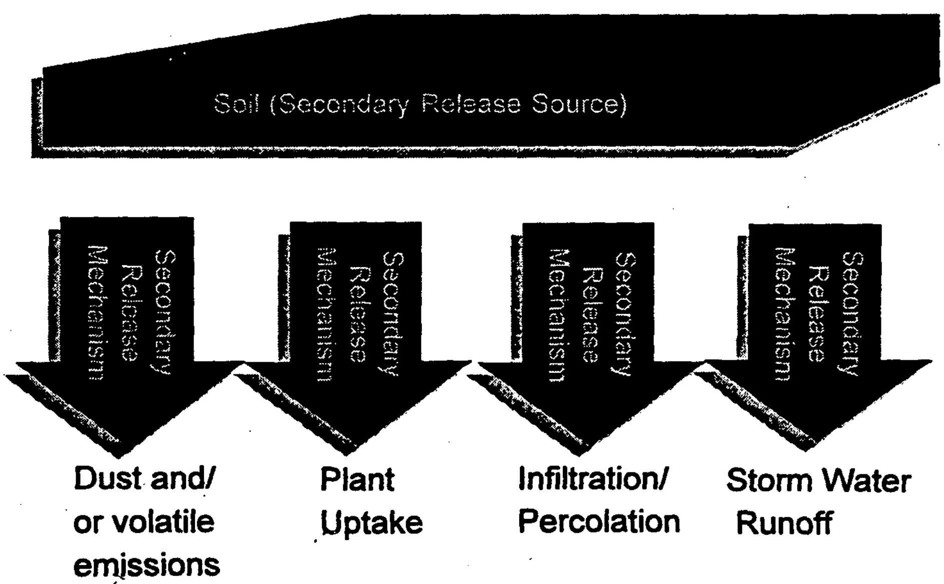
Step 1-Define a Conceptual Site Model (CSM)

- General site information
- Hydrogeologic Characteristics
- Meteorological Characteristics
- Land use-Current and future
- Contaminant sources, distribution and release mechanism,
- Media affected by soil contamination
- Exposure pathways and migration routes, and potential receptors.

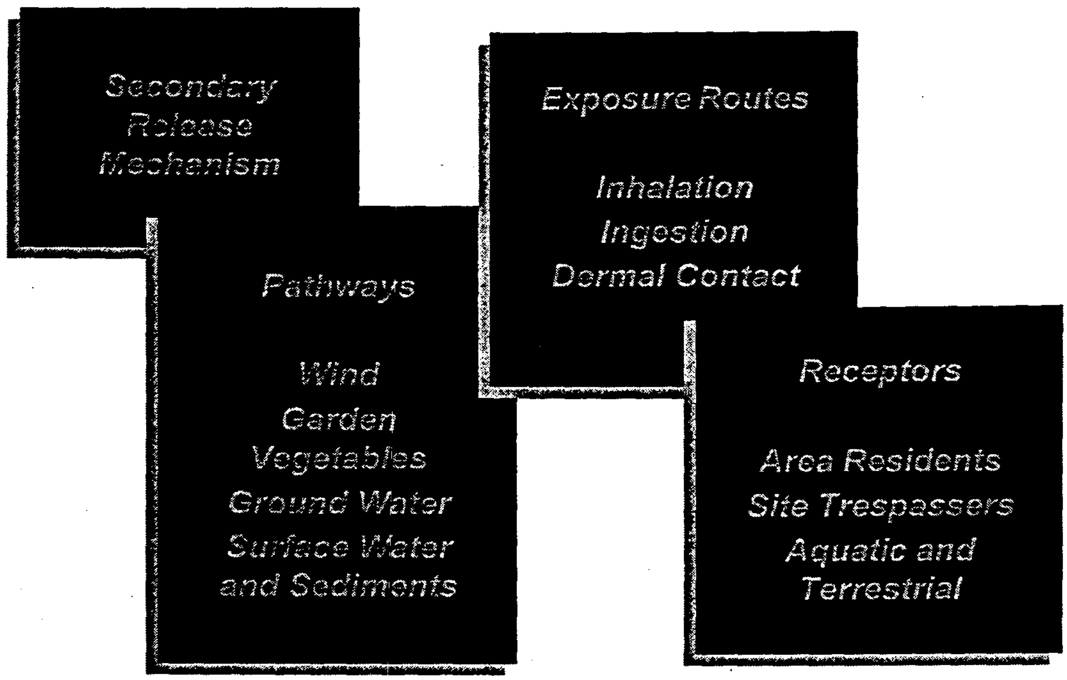
The Conceptual Site Model (CSM)



The Conceptual Site Model (CSM)



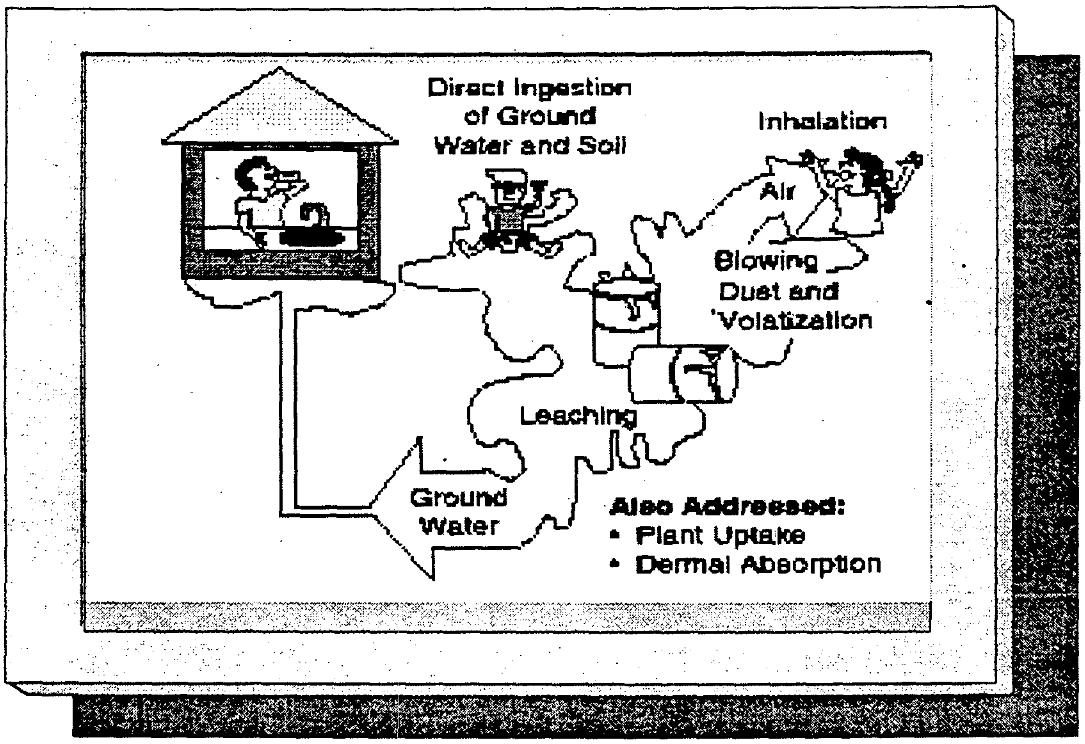
The Conceptual Site Model (CSM)



Step 2: Compare Soil Component of CSM to Soil Screening Scenario.

Identify Pathways Present at the Site Addressed by Guidance.

Pathways Addressed by Guidance.



Direct Ingestion: Non-cancer Risk Equation for the SSL

Ingestion Screening Level (mg/kg)=

$$\frac{THQ \times BW \times AT \times 365\,d/yr}{1/RfDo \times 10^{-6}\,kg/mg \times EF \times ED \times IR}$$

noncancer SSLs use more conservative child receptor

Direct Ingestion: Cancer Risk Equation for the SSL

Ingestion Screening Level (mg/kg)=

$$\frac{TR \times AT \times 365\ d/yr}{SFo \times 10^{-6}\ kg/mg \times EF \times IFsoil/adj}$$

Inhalation Screening Level (mg/kg) Noncancer Risk Equation

Inhalation Screening Level (mg/kg) =

$$\frac{THQ \times AT \times 365\ d/yr}{EF \times ED \times \text{[illegible]} \times (1/PEF\ \text{or}\ 1/VF)}$$

VF=Volatilization Factor

PEF=Particulate Emission Factor

Inhalation Screening Level (mg/kg) Cancer Risk Equation

Inhalation Screening Level (mg/kg) =

$$\frac{TR \times AT \times 365\ d/yr}{URF \times 1000\ \mu g/mg \times EF \times ED \times [(1/PEF)\ or\ (1/VF)]}$$

VF=Volatilization Factor
PEF=Particulate Emission Factor

Pathways not addressed by the Soil Screening Guidance

- **Human/Direct Pathways:**
  - **ingestion and inhalation of fugitive dusts under an acute exposure**
- **Human/Indirect Pathways:**
  - **consumption of nearby meat or dairy products**
  - **fish consumption from nearby surface waters with recreational or subsistence fishing**
- **Ecological Pathways:**
  - **aquatic and terrestrial**

Step 3: Defining Data Collection Needs

Stratify Site Based On Existing Data

"Zero" concentration
Screening level

No Further Study Warranted Under CERCLA

Screening level
Response level

Site-specific Cleanup Level

Response level
Very high concentration

Response action Clearly Warranted

Step 3: Defining Data Collection Needs

- **Media Concentration**
- **Fate and Transport Data**
- **Background Data**

Step 4-Sampling and analyzing soils at the site

Follow DQO Process according to:

Data Quality Objectives for Superfund: Interim Final Guidance

Data Quality Objectives Process

Soil Screening Process Step-by-Step Approach

- **Step 5: Calculating site-specific SSLs**
  - ▲ **SSL risk algorithms for surface and subsurface soil**
    - **direct ingestion**
    - **soil-to-air**
- **Step 6: Comparing site soil contaminant concentrations to calculated SSLs**
  - **Site-specific and Generic SSLs**
  - **Surface and Subsurface Soil**
- **Step 7: Determining which areas of the site require further study**

Step 5 -Calculating Site-Specific SSLs

SSLs are risk-based concentrations that are back-calculated at acceptable risk levels.

Target Cancer Risk is 1E-06
Hazard Index is 1

Step 5 -Calculating Site-Specific SSLs

- **Derived from RME equations and models for a residential exposure that combine:**
  - **air concentrations for particulate and volatile emissions, risk-based ground water concentrations; and**
  - **chemical characteristics (e.g., fate and transport); and**
  - **site characteristics (e.g., size of site, vegetative cover, wind speed); and**
  - **EPA toxicity to compute an acceptable concentration in soil that is compared with the on-site soil concentration.**

Step 5 -Calculating Site-Specific SSLs

- The SSL Guidance calculates SSLs for 110 chemicals found at Superfund Sites.
- SSLs are calculated for surface and subsurface soil exposure pathways.
- SSL Guidance default values are used to generate generic SSLs and can be used to compute additional SSLs for other chemicals.

Step 5 -Calculating Site-Specific SSLs

All SSL site-specific soil parameters reflect average or typical site conditions.

Chronic exposure combines the average concentration with reasonably conservative values for intake and duration.

Step 5 -Calculating Site-Specific SSLs

SSLs based on a Reasonable Maximum Chronic Exposure (RME)

Acute exposure is not considered

Fate and transport properties, volatility and site characteristics are taken into consideration

Step 5 -Calculating Site-Specific SSLs

- **Toxicity Criteria:**
  - **IRIS and HEAST (other sources - NCEA may be used.)**
  - **Nonzero Maximum Contaminant Levels Goals (MCLGs), Maximum Contaminant Levels (MCLs) or Risk-based Concentrations are used for the migration to groundwater pathway.**

Step 5 -Calculating Site-Specific SSLs

- Additive risks are not "built in" to the SSLs calculations.
- Potential for additive effects for multiple chemicals and multiple pathways are not considered.

Step 5 -Calculating Site-Specific SSLs

- Cancer Risk:

  Risks are generally within the acceptable risk range when multiple chemicals are present.

Step 5 -Calculating Site-Specific SSLs

- Noncancer Risk:

  The guidance recommends that the SSL be divided by the number of chemicals affecting the same target organ.

- Region 3 has traditionally used a target hazard quotient of 0.1 for all chemicals.

Step 5 -Calculating Site-Specific SSLs

- Additive risks from multiple pathways are not considered.

- Each SSL exposure pathway is screened separately without consideration to additive exposure from the multiple pathways.

- This may be a concern at some sites.

SSL-Surface Soil

- Direct Ingestion
- Dermal Contact
- Inhalation of Fugitive Dust

Direct Ingestion: Non-cancer Risk Equation for the SSL

Ingestion Screening Level (mg/kg)=

$$\frac{THQ \times BW \times AT \times 365\ d/yr}{1/RfDo \times 10^{-6}\ kg/mg \times EF \times ED \times IR}$$

Noncancer SSLs use more conservative child receptor

Direct Ingestion: Cancer Risk Equation for the SSL

Ingestion Screening Level (mg/kg)=

$$\frac{TR \times AT \times 365\ d/yr}{SFo \times 10^{-6}\ kg/mg \times EF \times IF_{soil/adj}}$$

Cancer SSLs use a time-weighted average soil ingestion rate for child/adult to account for higher exposure during childhood.

Direct Ingestion: Cancer Risk Equation for the SSL Age-Adjusted Ingestion Factor (IF)

IF soil/adj (mg-year/kg-day) =

$$\frac{IR_{soil/age1-6} \times ED_{age1-6}}{BW_{age1-6}} \text{ [illegible] } \frac{IR_{soil/age7-31} \times ED_{age7-31}}{BW_{age7-31}}$$

Dermal Contact

- Absorption must be greater than 10% to equal or exceed the ingestion exposure (assuming 100% absorption of chemicals via ingestion).
- Pentachlorophenol is greater than 10% absorption and is the only SSL meeting criteria of those chemicals for which SSLs were calculated.

Dermal Contact

- SSL is divided by 2 to account for dermal route exposure being equivalent to the ingestion route.
- Region 3 approach for site-specific SSLs follows the Dermal Guidance (1992).

Inhalation Screening Level (mg/kg)-Noncancer Risk Equation Fugitive Dust

Inhalation Screening Level (mg/kg) =

$$\frac{THQ \times AT \times 365\ d/yr}{EF \times ED \times (1/RfC \times 1/PEF)}$$

PEF=Particulate Emission Factor

Inhalation Screening Level (mg/kg) Cancer Risk Equation Fugitive Dust

Inhalation Screening Level (mg/kg) =

$$\frac{TR \times AT \times 365\ d/yr}{URF \times 1000\ \mu g/mg \times EF \times ED \times (1/PEF)}$$

PEF=Particulate Emission Factor

Subsurface Soil

- Inhalation of VOCs
- Ingestion of groundwater contaminants by migration of contaminants through soil to underlying potable aquifer.

Inhalation Screening Level (mg/kg)-Noncancer Risk Equation Volatile Emissions

Inhalation Screening Level (mg/kg) =

$$\frac{THQ \times AT \times 365\ d/yr}{EF \times ED \times (1/RfC \times (1/VF))}$$

VF=Volatilization Factor

SSL is compared with Csat and the Mass Limit SSL

Adjustment for additive risk should not be considered for Csat based SSLs.

Inhalation Screening Level (mg/kg)
Cancer Risk Equation
Volatile Emissions

Inhalation Screening Level (mg/kg) =

$$\frac{TR \times AT \times 365\ d/yr}{URF \times 1000\ \mu g/mg \times EF \times ED \times (1/VF)}$$

VF=Volatilization Factor

SSL is compared with Csat and the Mass Limit SSL

Inhalation SSLs:

- **SSLs based on fugitive dust are higher than the ingestion SSLs.**
- **SSLs based on volatiles are lower than ingestion SSLs.**
- **Generic SSLs for ground water ingestion (DAF of 20) are lower than inhalation SSLs.**

Inhalation SSLs:

- For some contaminants, the lack of inhalation benchmarks may underestimate risks due to inhalation exposure.
- SSLs for ground water can be used for screening when there is ground water contamination and the inhalation pathway may be a concern.
- Route-to-route extrapolation may be performed when there is no ground water contamination.

Inhalation SSLs:

- Route-to-Route Extrapolation: Oral toxicity criteria converted to an inhalation criteria.
- Must account for respiratory tract deposition efficiency and distribution; and
- Physical, biological, and chemical factors; and
- Other aspects of exposure (e.g., discontinuous exposure) that affect uptake and clearance.

Inhalation SSLs:

- Guidance:

- Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry (U.S. EPA, 1994).

Surface/Subsurface Soil: Plant Uptake

- Consumption of garden fruits and vegetables grown in contaminated residential soils.
- Only inorganics considered, empirical data for organics is lacking.

Surface/Subsurface soil: Plant Uptake-Risk Equation

Screening Level (mg/kg) = $\frac{Cplant}{Br}$

Cplant = (mg/kg DW) = $\frac{I \times BW}{F \times CR}$

$\frac{[illegible] \times BW [illegible]}{F \times [illegible] (CR \times Br)}$

Surface/Subsurface soil: Plant Uptake-Risk Equation

Cplant = $\frac{I \times BW}{F \times CR}$

I = $\frac{TR \times AT \times 365\ d/yr [illegible]}{ED \times [illegible] SForal}$ Carcinogens

I = $\frac{HQ \times [illegible] AT \times 365\ d/yr}{[illegible]}$ Non-Carcinogens

Surface/Subsurface Soil: Plant Uptake

- Site specific factors that influence plant uptake and plant contamination concentration
  - pH (influence mobility)
  - Chemical form strongly influence the uptake of metals into plants (influence bioavailability)
  - Plant type (phytotoxicity can influence bioconcentration in plant tissue)

Step 6-Comparing Site Soil Contaminant Concentrations to Calculated SSLs

- Samples from an exposure area is compared to 2SSLs
- When all of the samples are less than 2SSLs, an exposure area is screened out.

Step 6-Comparing Site Soil Contaminant Concentrations to Calculated SSLs

- Several exposure point concentrations can be used to compare the SSLs depending on the site-specific data collected.
- The maximum composite sample concentration for composite samples is used for surface soil SSLs. The Max test is used.

Step 6-Comparing Site Soil Contaminant Concentrations to Calculated SSLs

- The maximum concentration is used with discrete samples at sites with a limited surface soil data set.
- Sites with a limited data set are compared to 1SSLs, not 2SSLs.

Step 6-Comparing Site Soil Contaminant Concentrations to Calculated SSLs

- Subsurface soil data are not composited. The average concentration in a source (as represented by discrete contaminant concentrations averaged within soil borings) is used for the inhalation of volatiles and for the soil-to-ground water SSLs.

- Subsurface soil data are compared to 1SSLs, not 2SSLs.

Step 6-Comparing Site Soil Contaminant Concentrations to Calculated SSLs

- Review the CSM with the actual site data–Is it still reasonable and applicable?

- The gray region has been set between one-half and two times the SSL. Were the desired error rates at the SSL met?

- Were sufficient data collected? Did it pass the DQA process?

Step 7-Addressing Areas Identified for Further Study

- Subject of RI/FS and a baseline risk assessment.
- Data collected for soil screening can be used in RI and risk assessment.

Step 7-Addressing Areas Identified for Further Study

- The 95%UCL or the Max composite sample is used in the RI/FS risk assessment for contaminants of concern (COCs.)
- Additional data may be needed for future investigations.
- SSLs can be used as PRGs after decision is made to remediate if conditions still apply.

The Effects of Shapes on Sample Size

- The following facts become apparent when various shapes and probabilities are assessed:
  1. The number of samples needed increases as the size of the spot which is acceptable to miss decreases.
  2. The number of samples needed increases as the acceptable probability of missing a hot spot decreases.
  3. If the hot spot is circular, fewer numbers of samples are needed than when it is elliptical, and the longer the horizontal axis is in the ellipse, the larger is the number of samples that will be needed for a given probability and grid shape.
  4. A triangular grid is the most efficient and a rectangular grid is the least efficient for finding a hot spot using the same assumptions.

Example 3. Effect of the Shape of a Spot on the Numbers of Samples Needed?

- For a Square Grid with a Sampling Area of 500 square meters, and a Probability of Missing a Hot Spot, if one existed, equal to 0.1, how many Samples are needed to:

- detect a circular hot spot of minimum radius 1, (=152)

- detect an elliptical hot spot, (= 232)

- detect a hot spot which is a long ellipse, (=353).

Example 4. Effect of the Shape of the Grid on the Numbers of Samples Needed?

- For a Sampling Area of 1000 square meters, and a Probability of 0.05 of missing a Circular Hot Spot of Minimum Radius 1 meter, if one existed, how many Samples are needed using:

- a Square Grid, (=360)
- a Triangular Grid, (= 289)
- a Rectangular Grid, (=500)

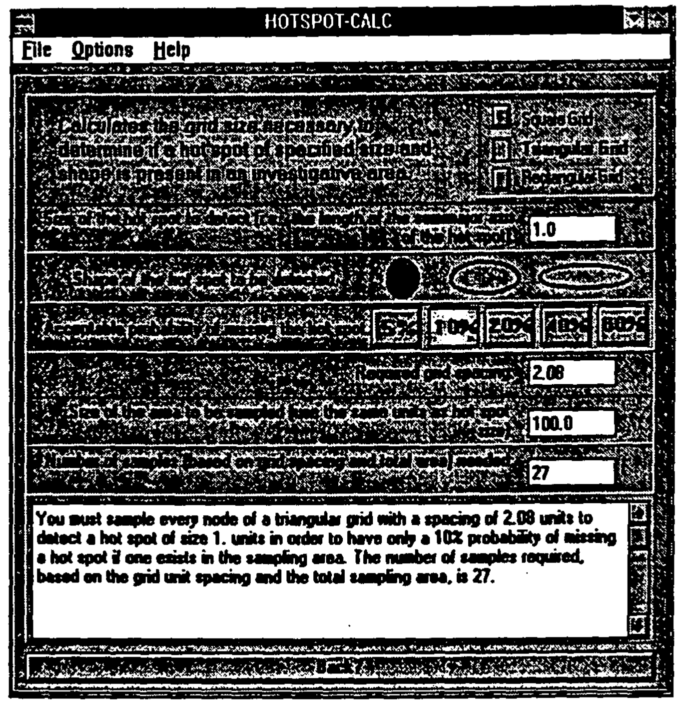
Example 1. Effect of Decreasing the Size of a Spot on the Numbers of Samples Needed?

- For a Square Grid with a Sampling Area of 500 meters, and probability of 0.6 of Missing a Hot Spot, if one existed -
- How many Samples are required for:
- detecting a circular hot spot of minimum radius of 5.0 meters, (=3)
- detecting a circular hot spot of minimum radius of 4.0 meters, (=4)
- detecting a circular hot spot of minimum radius of 3.0 meters, (=7)
- detecting a circular spot of minimum radius of 2.0 meters, (=16)
- detecting a circular spot of minimum radius of 1.0 meters, (=62)
- detecting a circular spot of minimum radius of 0.5 meters, (=245)

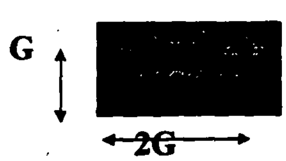
Example 2. Effect of Decreasing the Probability of Missing a Spot on the Numbers of Samples Needed?

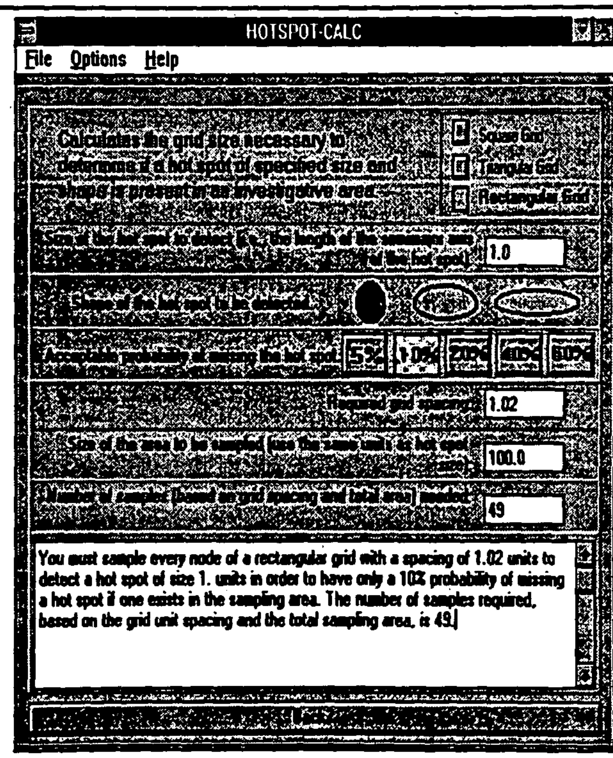
- For a Square Grid with a Sampling Area of 4000 square meters, how many Samples are needed to Detect a Hot Spot of Minimum Radius 2.5:
- for probability of 0.60 of missing a hot spot if one existed, (= 79)
- for probability of 0.40 of missing a hot spot if one existed, (=113)
- for probability of 0.20 of missing a hot spot if one existed, (=160)
- for probability of 0.10 of missing a hot spot if one existed, (=194)
- for probability of 0.05 of missing a hot spot if one existed, (=231)

What if the Grid is Changed to a Triangle?



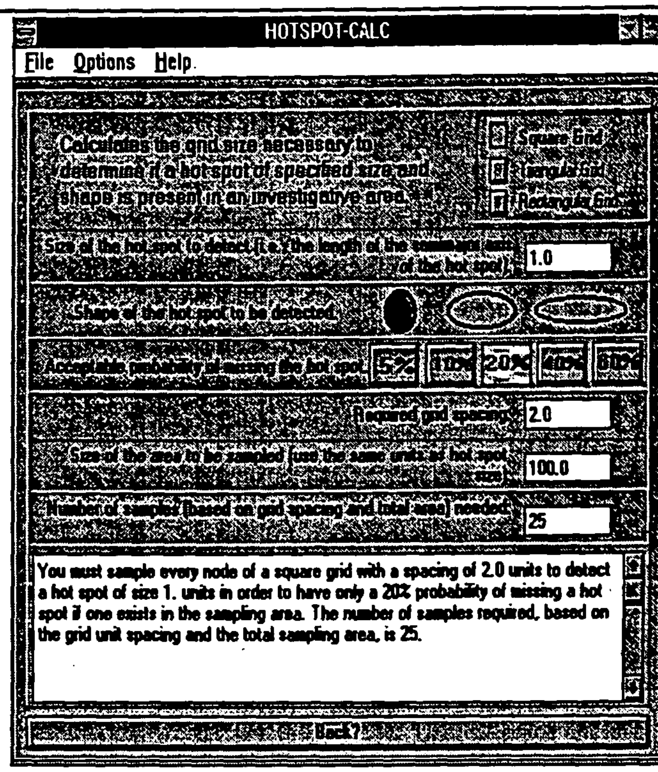
Assume a Rectangular Grid, a Round Spot, and a 10% Probability of Missing the Hot Spot





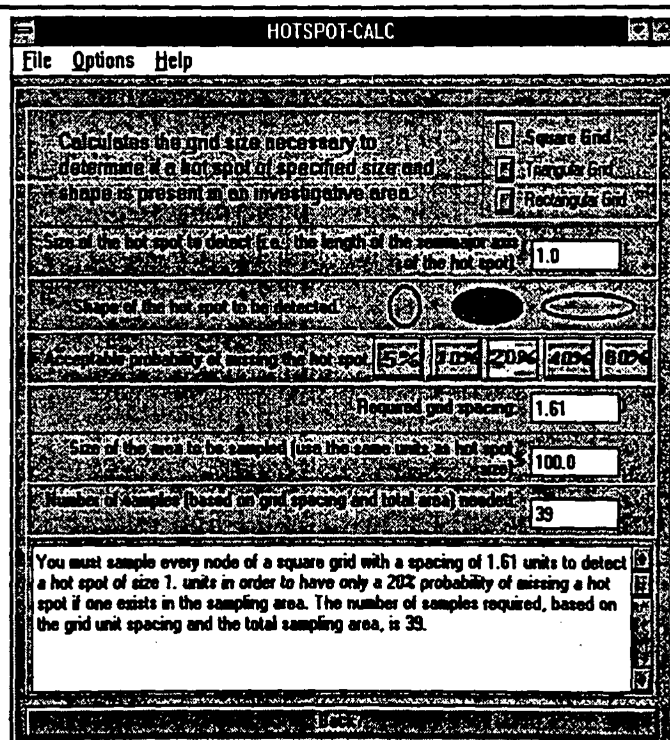
Using a Square Grid, What if the Acceptable Probability of Missing a Hot Spot is Increased?

Doubling the probability of missing the spot only decreased the number of samples needed by 6.



What if the Hot Spot is an Ellipse Instead of Circular in Shape?

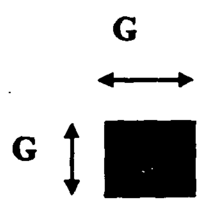
Then the number of samples increases from 25 to 39.

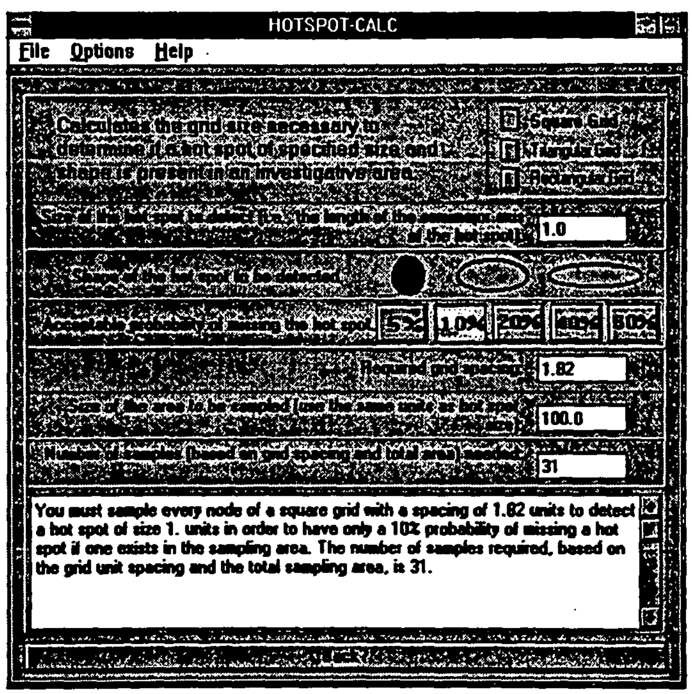


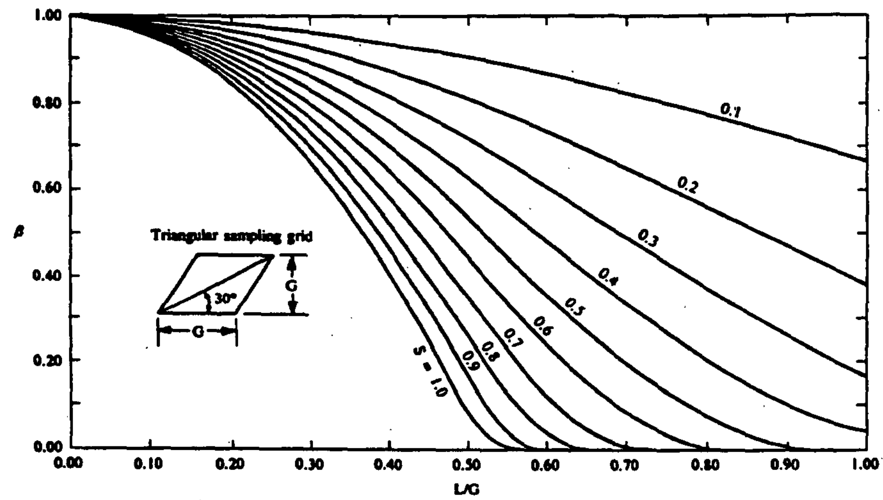
Inputs to *HotSpot-Calc*

- The shape of the grid that will be used:
  - such as triangle, square, or rectangle.
- The size and shape of the spot:
  - such as circle, ellipse, or long ellipse.
- The acceptable probability of **missing** the hot - spot:
  - such as 10%, 20%, etc.
- The size of the area to be sampled:
  - such as 100 square meters, 2 square miles, etc.

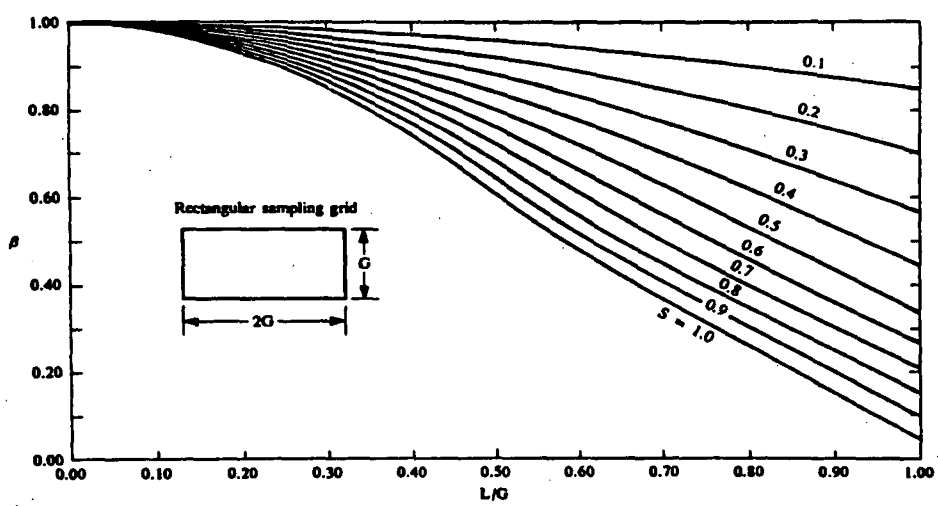
What if the Grid is Changed to a Square?



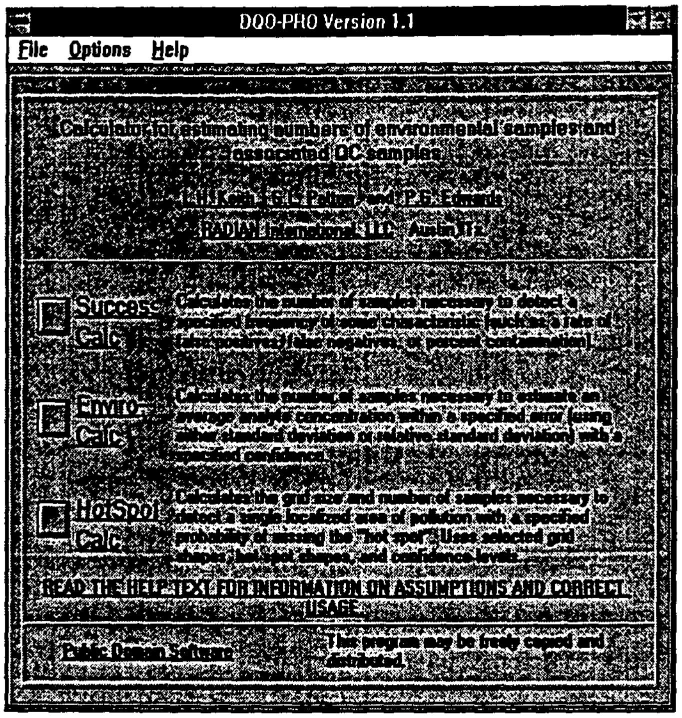


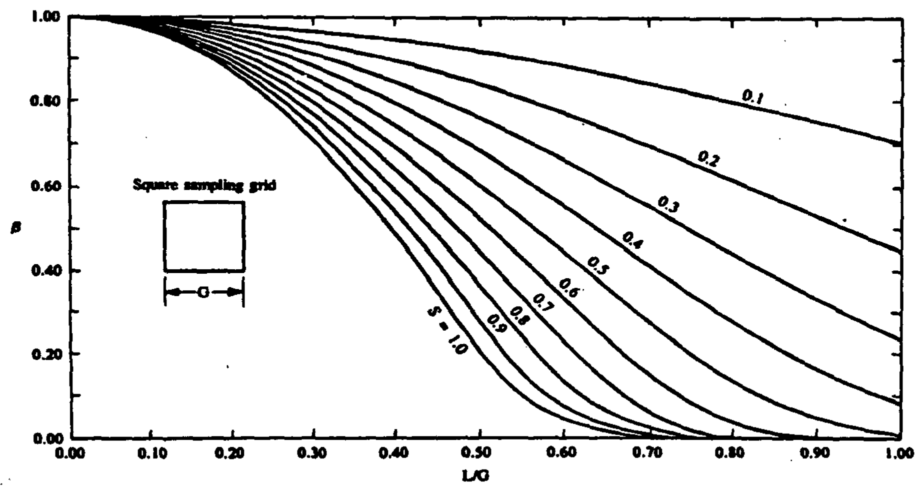


Curves relating L/G to consumer's risk, β, for different target shapes when sampling is on a triangular grid pattern (after Zirschky and Gilbert, 1984, Fig. 4).



Curves relating L/G to consumer's risk, β, for different target shapes when sampling is on a rectangular grid pattern (after Zirschky and Gilbert, 1984, Fig. 5).





Curves relating L/G to consumer's risk, β, for different target shapes when sampling is on a square grid pattern (after Zirschky and Gilbert, 1984, Fig. 3).

HotSpot - Calc Probability

- The probability of finding a hot spot is determined as a function of the specified size and shape of the hot spot, the pattern of the grid (rectangular, square, or triangular), and the relationship between the size of the hot spot and the grid spacing.

- *HotSpot-Calc* is a program developed by Dr. L.H. Keith based on the procedure described in Gilbert (1987). It computes the sample size using the probability of **missing** a hot spot if one exists rather than on the probability of finding one.
  - The program computes the grid spacing for detecting:
    - a circular hot spot (S=1),
    - elliptical hot spot (S~0.7) - fat ellipse, and
    - elliptical hot spot (S ~0.5)- slim long ellipse.
  - For other elliptical shapes consult the nomographs.

Program *HotSpot-Calc*

- Program *HotSpot-Calc* determines the grid size needed to detect the presence of a single localized spot of pollutants ("hot spot") of a specified size and shape with a specified probability of missing its detection if it is present.

- Once the grid spacing,G, is calculated, the number of samples needed to meet the prespecified performance standards is obtained using the equations:

- $n=A/G^2$ for square grid,
- $n=A/(2G^2)$ for rectangular grid, and
- $n=A/(0.866\ G^2)$ for triangular grid.

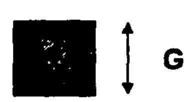
Assumptions for Hot-Spot Detection

- The program *HotSpot-Calc* determines the grid spacing needed to detect the presence of a single hot spot of a specified size and shape with a specified probability of **missing** the hot spot. It is based on the following key assumptions:
  1. That the hot spot is circular (S=1), short elliptical, (S=0.7) or long elliptical (S=0.5) in shape;
  2. That sample measurements are collected on square, rectangular, or triangular grids;
  3. That the definition of a "hot spot" is clear and agreed to by all decision makers; and,
  4. That there are no classification errors (i.e., that there are no false-positive or false-negative measurement errors).

Calculating Numbers of Samples For Hot - Spot Detection

The number of samples required for hot spot sampling is the number of samples required to sample all grid areas at the site for the selected grid spacing. The number of samples required for a **square grid** is approximated by the equation:

$n = A/G^2$



where,

n = number of samples,

A = area to be sampled, in the square of the units for G

and, G = grid spacing.

Hot - Spot Sampling Objectives

- The objective of hot - spot sampling is to determine if localized areas of contamination exist.
  - These localized areas of contamination may be due to spills, leaks, buried waste, or any number of other events where contamination might be confined to a relatively small area.
    - A single site might have multiple hot spots of different origins.
    - Will consider the problem of detecting a single hot spot given that it exists.
  - Dr. L. Keith developed a software, *HotSpot-Cal* to compute the grid size and the sample size needed to detect a hot-spot of a specified size (given that one existed) with probability of missing the spot =ß. The program is in public domain can be down loaded from the internet.

Systematically Sampling a Grid

- Hot - spot sampling involves performing a systematic search of a site for "hot spots" of a certain specified shape (e.g., round, elliptical) and area.
  - The search is conducted by sampling grid nodes on a two-dimensional grid of spacing G, or
  - Samples are taken either in the center of every grid cell or randomly within every cell area.
  - **Shape Of Hot Spot:**
  - M = Length of the semiminor axis of the smallest hot spot need to detected.
  - L = length of semimajor axis of the smallest hot spot critical to detect.
  - Shape, S = Length of semiminor axis/Length of semimajor axis.
  - S: $0<S<=1$. If exact shape is not known, use a conservative shape factor, S=0.5, rather than using a circular or a fatter elliptical shape.

Site-Specific Background/Reference Area

- The background /reference area should be free of the contamination from the site.
- The reference area to be compared with cleanup units (i.e, EA) should be similar to those units in physical, chemical, and biological characteristics.
- The distribution of the COPC in the reference area should be similar to that of the cleanup unit if that cleanup unit had never become contaminated due to the industrial site activities.
- Reference areas are sometimes selected as areas closest to but unaffected by the cleanup unit assuming that spatial proximity implies similarity of concentrations in reference area and the cleanup unit.

Background Levels Exceed SSLs?

- Use hypothesis testing (e.g., two sample t-test, or Wilcoxon's rank sum test) to compare the concentrations of COPC in the site background soils with the respective SSL.
- Using the background data, compute the UCL of the mean contaminant of concern.
    - If UCL < SSL, conclude that background concentrations do not exceed the SSL, and simply proceed with the screening of the cleanup unit, EA, or site under study.
    - If UCL >=SSL, compare the mean background concentration of a COPC with the mean contaminant concentration of the cleanup unit (EA) under study.
    - Use parametric t-test (or non-parametric) to compare the mean concentration background with that of the EA .

Which Procedure(s) to Use?

- In hypothesis testing using composite samples, the Chebychev inequality resulted in the same conclusion as the Max test.
- It is anticipated that procedure based on the Chebychev UCL will control false negative error rate better than the Max test.
- Also, for verification of the attainment of cleanup levels, UCL is compared with C_s (and not $2C_s$).
- **In order to make recommendation for the best procedure meeting the DQOs, power comparison of the various UCLs such as the Chebychev UCL, Adj-CLT, and Max test needs to be made**

Background Levels Exceed SSLs?

- Two types of background contaminants:
  - naturally occurring - organic contaminants, and
  - anthropogenic - contaminants introduced by humans.
  - Use of SSLs as screening thresholds is not appropriate when background contaminant concentration levels are of concern.
- When anthropogenic background concentrations exceed the SSLs, investigation requiring site specific background sampling may be conducted to study the area soils.
- The site-specific background data can be collected using one of the sampling plans (Reference-Based Standards for Soil and Solid Media- Volume 3, 1994) such as:
  - Simple random sampling,or
  - Systematic grid sampling.

Which Procedure(s) to Use?

- The Max test is conservative, and controls Type I error at 2SSL fairly well; but results in a high number of false negatives at SSL/2. This false negative rate increases with the sample size and the standard deviation.

- The sample sizes listed in tables 23, and 25-30 are for low to moderately skewed data sets with CV < =5 (and values of sd,σ of log-transformed data smaller than 2.0).

- However, in environmental applications, samples with values of σ exceeding 2.0 are common.

- Sample sizes listed in these tables are not applicable to skewed distributions with σ exceeding 2.0.

Which Procedure(s) to Use?

- From figures 13 and 14 it is observed that the H-statistic based UCL of the mean does not have adequate power, and therefore cannot be recommended for use for composite samples.

  - The 1994 SSL Guidance document also pointed out need for a correction factor to improve power of test based upon H-UCL.
  - **This needs further investigation to draw conclusions and make recommendations.**

- In a separate study, it is observed that the Chebychev Inequality seems to control the Type I and Type II error rates reasonably well, and that the UCL based upon the Chebychev Inequality provides an adequate coverage for the mean concentration of a cleanup unit (see Singh, Singh, and Engelhard, 1997, 1998).

Site ABCD - LN(0.71, sd=1.78, CV=2.5) COPC=Xylene, CC =0.95, SSL=10 ppm

- Inference based upon right - tailed test: H_0: μ<=5, Vs H_1: μ>5
- Reject H_0 if the test statistic exceeds the critical value.
- Critical value t, Johnson=1.812
- Critical value for adjusted CLT = 1.10
- Critical value for Chen's test =1.645
- Student's t and Adj - CLT = 1.379
- Johnson's modified t-statistic = 1.464
- Chen's t-statistic = 1.977
- **Conclusion based upon** t and modified t: Data not provide enough evidence to reject H_0 and proceed with DQA process.
- **Conclusion based upon** Chen's and Adj-CLT: Reject H_0 and conclude that mean COPC is greater than 5 ppm and the EA needs further investigation.

Site ABCD - LN(0.71, sd=1.78, CV=2.5) COPC=Xylene, CC =0.95, SSL=10 ppm

- The null H_0:μ>= 2SSL =20 is rejected if 95% UCL of mean< 20.
- The 95% UCL based on t-Statistic = 17.16
- The 95% UCL based on regular CLT = 16.52
- The 95% UCL based on Johnson's modified t-Statistic = 18.59
- The 95% UCL based on adjusted CLT = 18.59
- The 95% UCL based on H-statistic (Land's) = 34.74
- The 95% UCL based on Chebychev Inequality using sample arithmetic mean and sd = 27.29
- Conclusion based upon data and H-UCL and Chebychev UCL: Data do not have enough evidence to Reject H_0 and conclude that mean concentration of COPC is greater than 20 ppm.
- Using Adj-CLT and t-tests, conclude that mean < 20, and proceed with DQA.

Site - ABC LN(1.62, sd=2.42, CV=1.5) DQA , CC =0.95, SSL=60 ppm

- Data Quality Assessment for Chen's Test :
- Chen's test did not reject the null hypothesis leading to the conclusion that mean of the COPC may be <=30.
  - Max = 492.7 > 60/sqrt(5) = 26.83, therefore determine a new sample size for CV =5.21 of individual measurements.
  - Consulting tables 25-30 of the SSL Guidance Document, the sample size for CV = 5.21 is not available.

Site ABCD - LN(0.71, sd=1.78, CV=2.5) COPC=Xylene, CC =0.95, SSL=10 ppm

- Inference based upon left -tailed test: H_0: μ>=20, Vs H_1: μ< 20.
- Reject H_0 if test statistic < negative of the critical value.
- Critical value for Student's and Johnson's t =1.812
- Critical value for adjusted CLT = 2.19
- Critical value for Chen's test =1.645
- Student's t and Adj- CLT statistics = -2.56
- Johnson's t-statistic = -2.47
- Max test = 36.12
- **Conclusion based upon Max test**: Do not reject H_0 and conclude that EA has mean > 20; but conclusion using other tests :Reject H_0 and conclude that EA has mean < 20, and proceed with DQA.

Site ABC - LN(1.62, sd=2.42, CV=1.5) COPC=B(a)P, CC =0.95, SSL=60 ppm

- Inference based upon right - tailed test: H_0: μ<=30, Vs H_1: μ>30
- Reject H_0 if the test statistic exceeds the critical value.
- Critical value t, Johnson=1.812
- Critical value for adjusted CLT = 0.6
- Critical value for Chen's test =1.645
- Student's t and Adj- CLT = 0.73
- Johnson's modified t-statistic = 0.894
- Chen's t-statistic = 1.229
- **Conclusion based upon data**
- Chen's test: Data not provide enough evidence to reject H_0 , proceed with DQA. Adjusted CLT: Reject H_0 and conclude that mean COPC is greater than 30 ppm - requiring further investigation.

Site ABC - LN(1.62, sd=2.42, CV=1.5) CC =0.95, SSL=60 ppm

- The null H_0 : μ>=120 is rejected if 95% UCL of μ <120.

- The 95% UCL based on t-Statistic = 151.51
- The 95% UCL based on regular CLT = 143.50
- The 95% UCL based on Johnson's modified t-Statistic = 159.32
- The 95% UCL based on adjusted CLT = 193.59
- The 95% UCL based on H-statistic (Land's) = 265.7
- The 95% UCL based on Chebychev Inequality using sample arithmetic mean and sd = 278.60.
- Conclusion based upon data and all UCLs: Data do not have enough evidence to Reject H_0 and conclude that mean concentration of COPC is greater than 120 ppm and EA needs further investigation.

DQA - Site XYZ - LN(2.563, sd=1.75) CC =0.95, SSL=60 ppm

- Data Quality Assessment :
  - Max = 110.4 > 60/sqrt(5) = 26.83, therefore determine a new sample size for CV = 2.36
  - Max Test : using Table 23 the sample size is about 8-9 for composites of 5 specimens each. The sample size of 10 is > 9, no further investigation needed.
  - Chen's Test: Using tables 25 and 26, it appears that about 6-8 composite samples of size 6-8 of 5 specimens each should be enough for DQA. Since we have 10 composite samples, no further investigation is needed.

Site ABC - LN(1.62, sd=2.42, CV=1.5) COPC=B(a)P, CC =0.95, SSL=60 ppm

- Inference based upon left - tailed test: H_0: μ>=120, Vs H_1: μ< 120.
- Reject H_0 if test statistic < negative of the critical value.

- Critical value for Student's and Johnson's t =1.812
- Critical value for adjusted CLT = 2.69
- Critical value for Chen's test =1.645
- Student's t and Adj- CLT statistics = -1.153
- Johnson's t-statistic = -0.990
- Max test = 492.70

- Conclusion based upon data and all tests: Do not reject H_0 and conclude that EA has mean > 120, and needs further investigation.

Site XYZ - LN(2.563, sd=1.75)
CC =0.95, SSL=60 ppm

- Inference based upon right - tailed test: H_0: μ<=30, Vs H_1: μ>30
- Reject H_0 if the test statistic exceeds the critical value.

- Critical value t, Johnson=1.812
- Critical value for adjusted CLT = 0.83
- Critical value for Chen's test =1.645
- Student's t and Adj- CLT = -0.088
- Johnson's t-test = 0.039
- Chen's t-statistic = 0.036
- Conclusion based upon data and all tests: Data do not provide enough evidence to reject H_0 , and conclude that mean COPC is less than 30 ppm - proceed with DQA to check Type II error of no more than 0.05 at 120.

Site XYZ - LN(2.563, sd=1.75)
CC =0.95, SSL=60 ppm

- Inference based upon the 95% UCL of the mean.
- The null H_0 : μ>=120 is rejected if 95% UCL of μ <120.

- The 95% UCL based on t-Statistic = 46.81
- The 95% UCL based on regular CLT = 45.18
- The 95% UCL based on Johnson's modified t-Statistic = 48.05
- The 95% UCL based on adjusted CLT = 53.12
- The 95% UCL based on H-statistic (Land's) = 67.92
- The 95% UCL based on Chebychev Inequality using sample arithmetic mean and sd = 72.74
- Conclusion based upon data and all UCLs: Reject H_0 and conclude that mean COPC is less than 120 ppm and perform DQA.

DQA Process: Cheb-UCL

- In addition to the condition that UCL < 2SSL, if Max of data <SSL/sqrt [c], then no further DQA or investigation is needed for that EA.

- If Max >=SSL/sqrt [c], then for prespecified performance standards (Type I and II errors) with CV^* for an individual observation as: $CV^* = CV$ sqrt [c], determine a new sample size using the program ProSamp. If new sample size exceeds the the number of samples used, then further investigation of the EA is necessary.

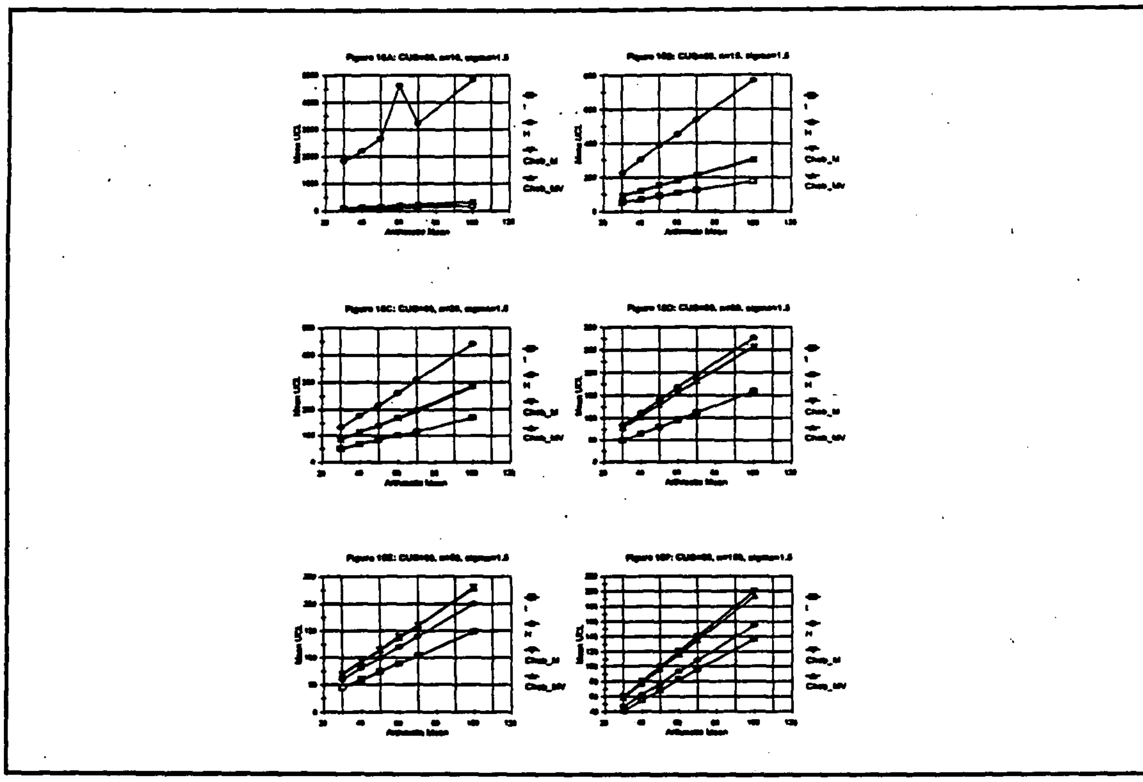
- In this case, additional samples need to be collected and the process repeated to verify if the EA can be screened out using the larger combined sample.

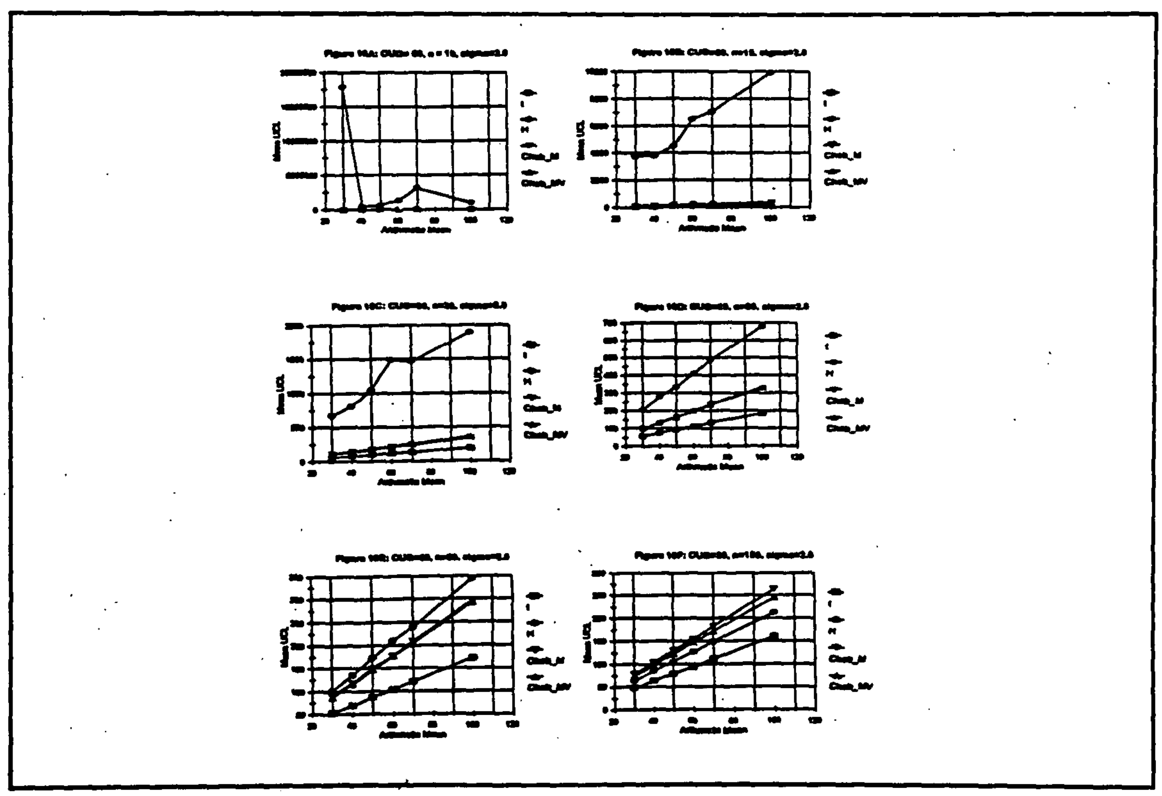
Site XYZ - LN(2.563, sd=1.75) COPC = B(a)P, CC =0.95, SSL=60 ppm

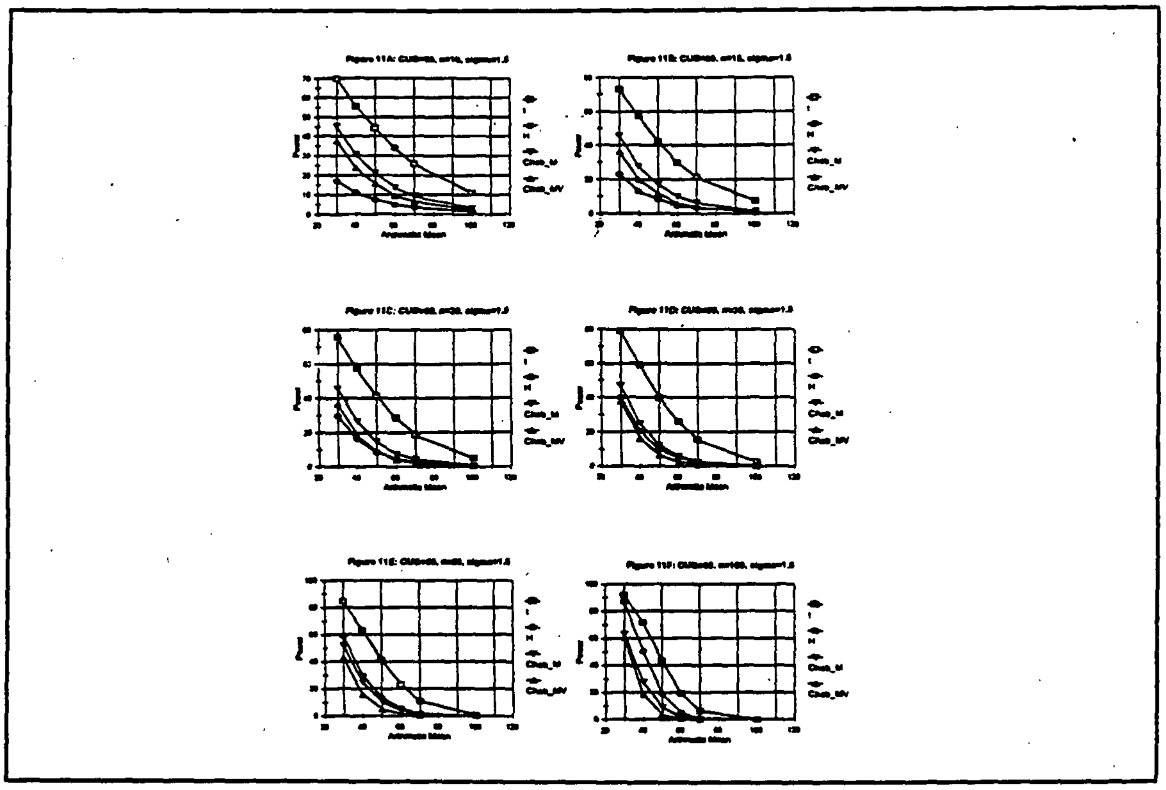
- Inference based upon left - tailed test: H_0: $\mu>=120$, Vs H_1: $\mu< 120$.
- Reject H_0 if test statistic is < negative of the critical value.

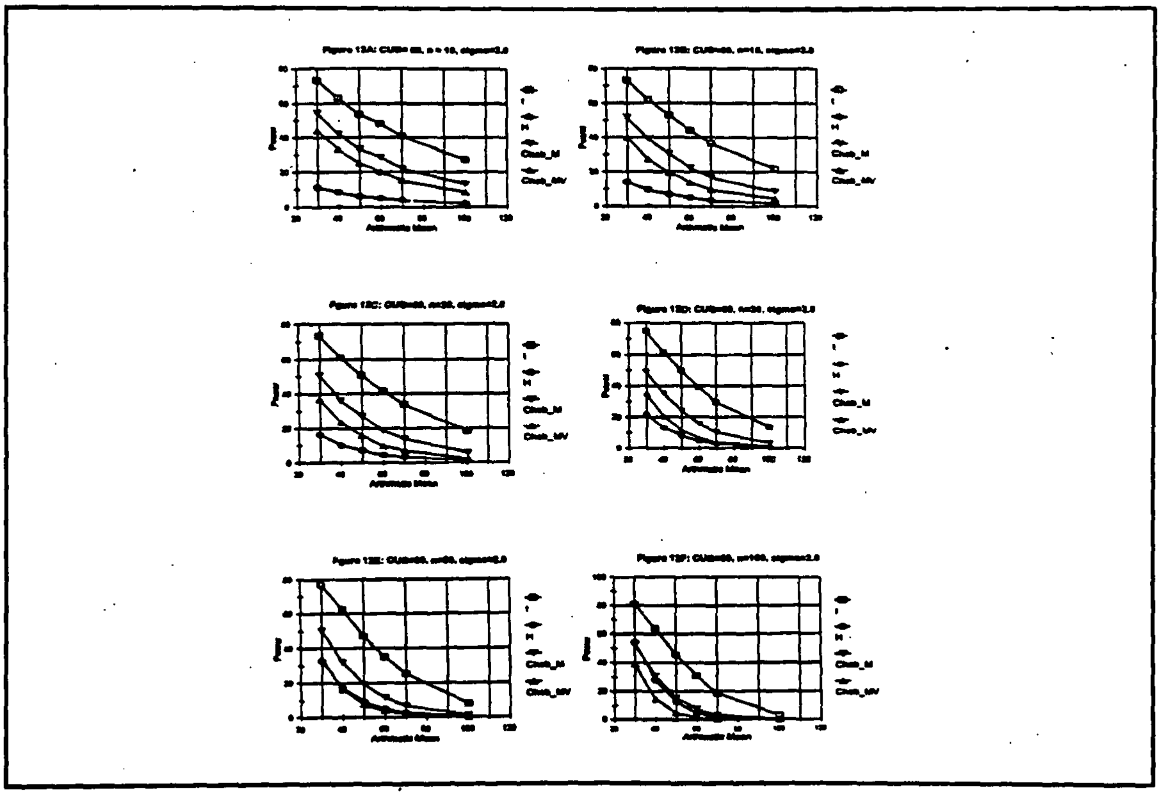
- Critical value for Student's and Johnson's t =1.812
- Critical value for adjusted CLT = 2.46
- Critical value for Chen's test =1.645
- Student's t and Adj- CLT statistics = -9.32
- Johnson's t-statistic = -9.193
- Max test =110.403

- Conclusion based upon data and all tests: Reject H_0 and conclude that mean COPC is less than 120 ppm, and proceed with DQA process to check for Type I error of no more than 0.05 at 120 ppm.









EPC Term- Chebychev UCL of Mean

- The Chebychev inequality results in a conservative estimate of the unknown mean of an EA (Singh, Singh, Engelhard,1997).
- The $(1-1/k^2)100\%$ UCL of the mean is given by $UCL = \bar{x} + k\hat{\sigma}_1/\sqrt{n}$, where σ_1 is the sd of the population of concern. For a 95% UCL of the mean, a conservative value for k~4.472.
- For lognormal populations using discrete samples, Singh, Singh, and Engelhard, 1997, 1998, observed that the Cheb-UCL results in a reasonable conservative estimate of the EPC term with adequate power even for samples of small size. This is especially true when one uses the MVUE of the mean of a lognormal population in place of the sample mean $\bar{x}$.

EPC Term - Chebychev UCL of Mean

- Also, note that compositing is used only when we are dealing with arithmetic mean.
- Therefore, use of the MVUE of population mean based upon lognormal theory may be inadequate when dealing with composite samples. **THIS NEEDS FUTHER INVESTIGATION.**
- For composite samples, the Cheb-UCL should be computed using sample arithmetic mean. If UCL >=2SSL, the EA can not be screened out and will require further investigation.
- For discrete samples, power graphs for lognormal data are given in figures 11a-11f, and 12a-12f, and the graphs for 95% UCL of mean are given in figures 15a-15f, and 16a-16f.

EPC Term - Land's UCL of The Mean

- The UCL of the mean - also called the exposure point concentration (EPC) term can be used to test if an EA can be screened out (RAGS document, 1992).
- Let $x_1, x_2, \ldots, x_n$ represent n discrete or composite samples from an EA with unknown mean μ. Let $y_1, y_2, \ldots y_n$ be the log-transformed data.
- The (1- α)100% H-statistic based UCL of the mean is given by:

$$UCL = \exp[\bar{y} + 0.5s_y^2 + s_y H_{1-\alpha} / \sqrt{(n-1)}]$$

  - If UCL >=2SSL, the EA can not be screened out and will require further investigation.

EPC Term - Land's UCL of The Mean

- However, the H- UCL given above is based upon discrete samples, c=1, and may need some correction factor for c>1. This is still under study and **NEEDS FURTHER INVESTIGATION.**
- In a simulation study on composite samples, it was observed that the procedure based on H-UCL results in a high false negative error rate as it does not have adequate power to reject the null hypothesis when it is false - as can be seen in figures 13-14. This is especially true when sd starts exceeding 1.0 (also see Singh, Singh, and Engelhard, 1997, 1998).
- The Land's procedure cannot be recommended for use to compute the EPC term based upon composite samples without further research in this area.

Figure 39: Ho: mean < 60/2
n-comp=10, sigma=2.0

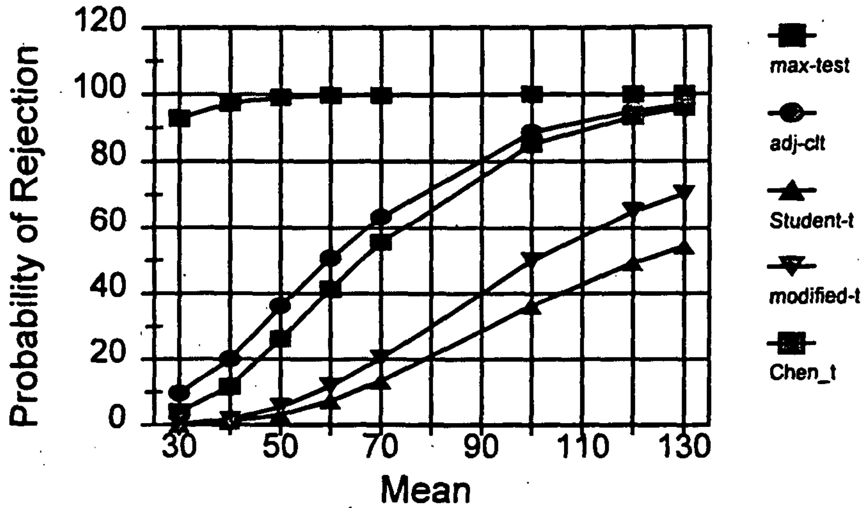


Figure 40: Ho: mean < 60/2
n-comp=10, sigma=2.5

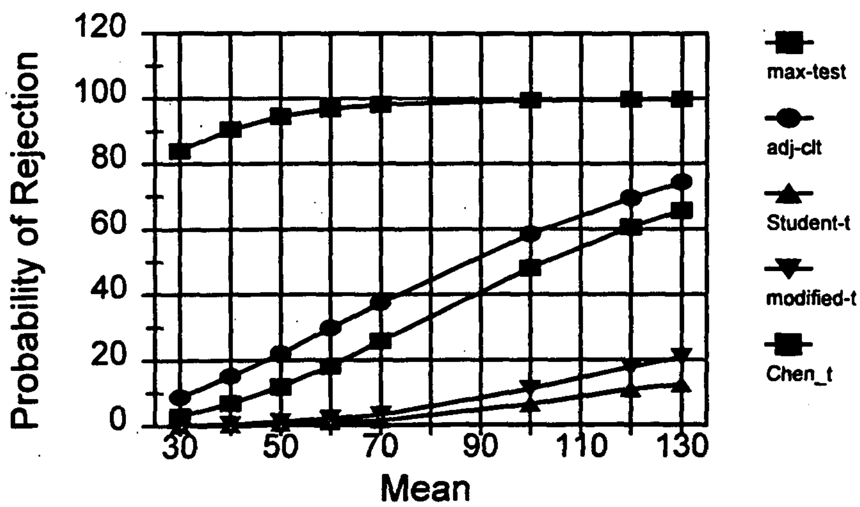


Figure 35: Ho: mean < 60/2
n-comp=8, sigma=2.5

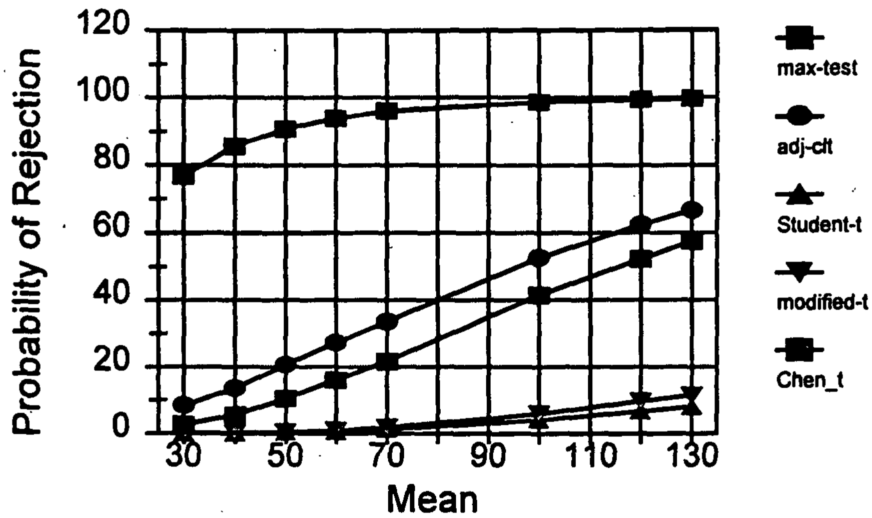


Figure 38: Ho: mean < 60/2
n-comp=10, sigma=1.5

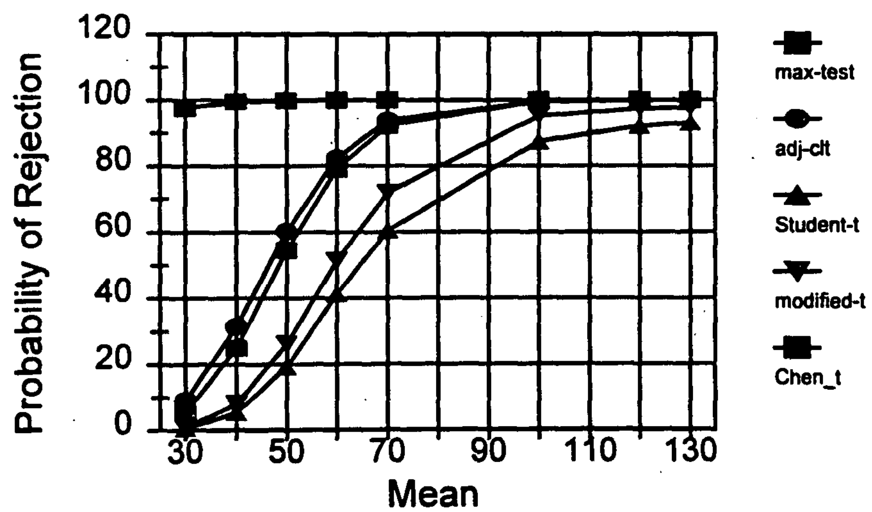


Figure 33: Ho: mean < 60/2
n-comp=8, sigma=1.5

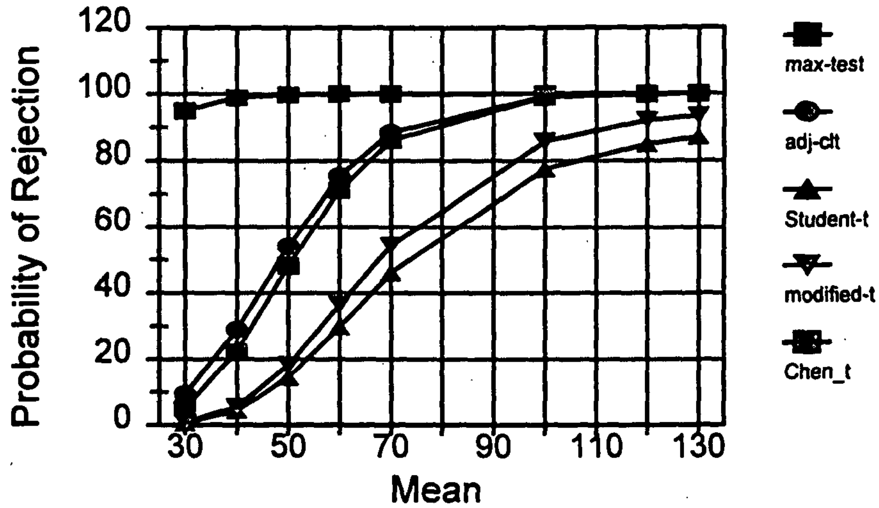


Figure 34: Ho: mean < 60/2
n-comp=8, sigma=2.0

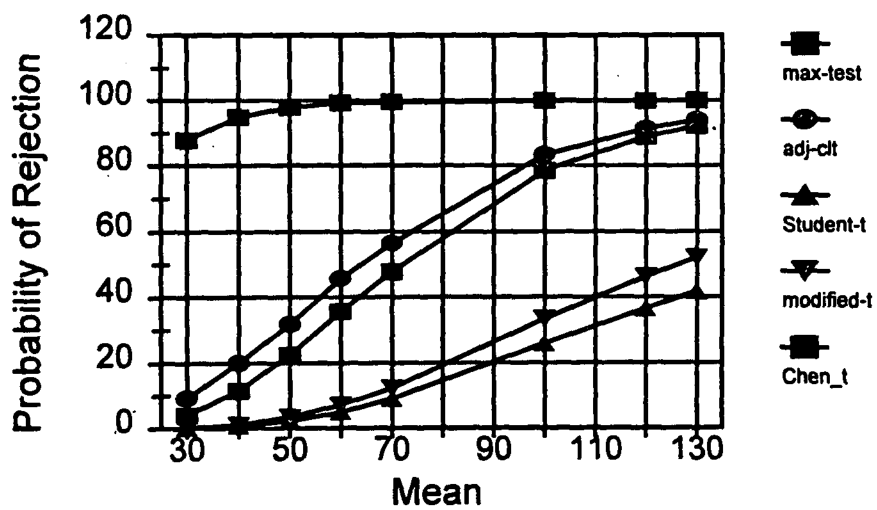


Figure 29: Ho: mean < 60/2
n-comp=5, sigma=2.0

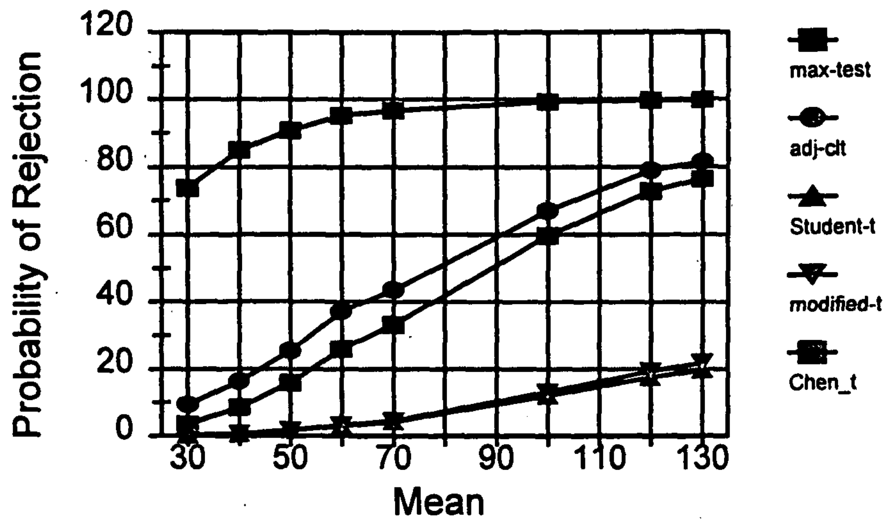
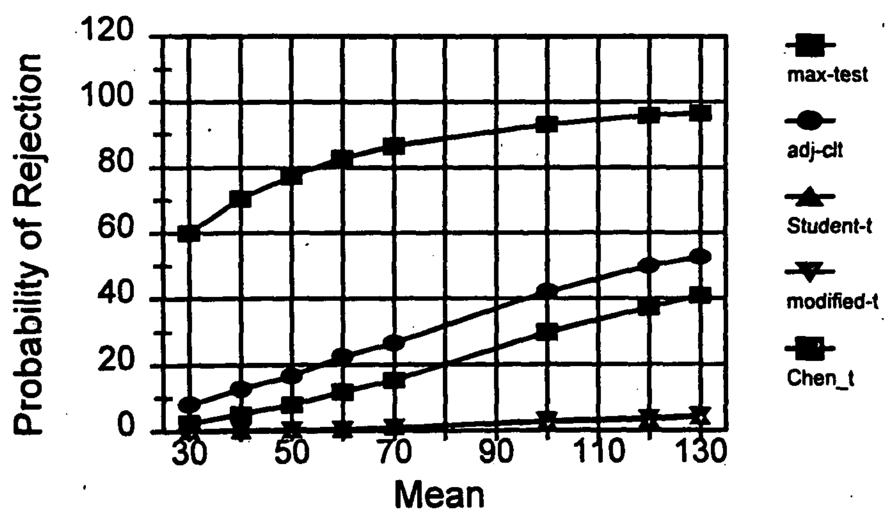


Figure 30: Ho: mean < 60/2
n-comp=5, sigma=2.5

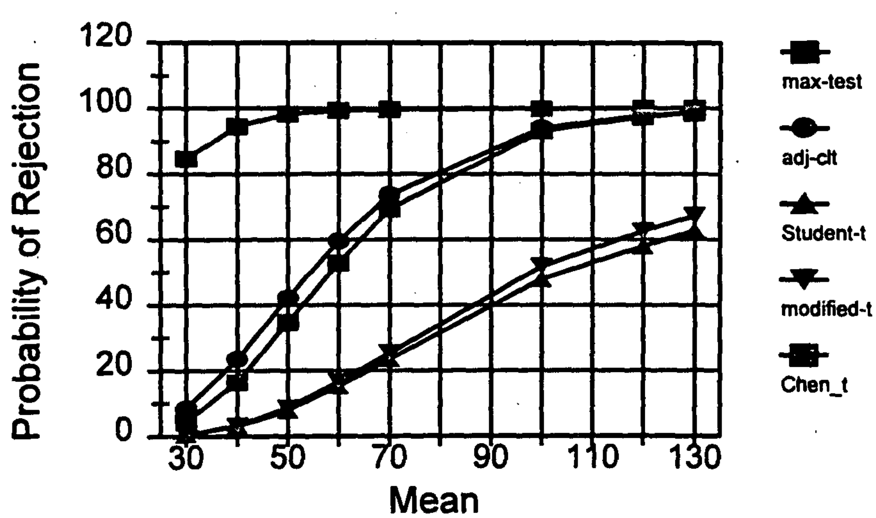


Comparison of Chen's and Right-Tailed Adj-CLT Tests

- For large values of sd exceeding 2.0, number of composite samples needed to achieve a power of 0.95 or more (probability of rejecting H_0 when mean >= 2SSL is less than 0.05) will be greater than 10 for the right-tailed Adj-CLT test and Chen's test. The power increases with the sample size but decreases as sd increases as can be seen in these figures.

- The influence of the number of specimens per composite on the power of the test **NEEDS FURTHER INVESTIGATION.**

Figure 28: Ho: mean < 60/2 n-comp=5, sigma=1.5



DQA for Adj- CLT Left -Tailed Test

- In addition to the condition that the null hypothesis is rejected for an EA to be screened out, if Max <SSL/sqrt [c], then no further DQA or investigation for that EA is needed.
- If Max >=SSL/sqrt [c], then for prespecified performance standards (Type I and II errors) with CV for an individual observation: CV* = CV* sqrt [c], determine a new sample size using the program ProSamp. If new sample size exceeds the sample size used, then further investigation of the EA is necessary.
- In this case, collect additional samples and repeat the testing process to verify if the EA can be screened out using the larger combined sample.

Comparison of Chen's and Right-Tailed Adj-CLT Tests

- From figures 28-30, 33-35, and 38-40, it is obvious that Adj-CLT test possesses higher power than Chen's test.
- **NOTE:** Both Chen's and Adj-CLT tests are consistent, and their the power (probability of rejecting H_0) increases with the sample size, n. The threshold value is SSL, but due to the way hypotheses are defined, the probability of rejecting H_0: $\mu <=$ 0.5SSL (e.g., investigating the site further) when the true mean of the EA is between SSL/2 and SSL increases as the sample size increases. This can be easily seen in figures 29, 34, and 39.
- Therefore, when large samples are available, define the null as H_0: $\mu <=$ SSL rather than H_0: $\mu <=$ 0.5SSL.

Adjusted CLT Left -Tailed Test

- If t >= z_α^* , there is insufficient evidence to reject the null hypothesis H_0 and conclude that EA needs further investigation.

- If t <= z_α^* , H_0 is rejected and the DQA process should be performed to determine if the sample size used is sufficient to achieve 100 α% or less chance of incorrectly rejecting H_0 when the mean COPC = 2SSL.

Adjusted CLT Right -Tailed Test

- For the right - tailed test, null is H_0: mean <= 0.5SSL (not protective of human health), Vs alternative H_1: mean > 0.5SSL, with Type I and Type II error rates as 0.2 and 0.05 at 2SSL.

- The Adj-CLT test statistic, t is given by: $t = \sqrt{n}(\bar{x} - SSL/2)/s$.

- The critical value for test is given by: $z_\alpha^{**} = [z_\alpha - a(1 + 2z_\alpha^2)]$
- Compare t to z_α^{**}

- If t>=z_α^{**}, the null hypothesis H_0 is rejected leading to the conclusion that EA needs further investigation.
- If t < z_α^{**}, the data do not provide enough evidence to reject null hypothesis H_0, and one should proceed with the DQA process.

DQA for Chen's Test

- In addition to the condition that the null hypothesis is not rejected,
  - if Max of data < SSL/sqrt [c], then no DQA is needed and the EA can be screened out without any further investigation.
  - if Max >=SSL/sqrt [c], then for prespecified performance standards (Type I and II error rates), and $CV^* = CV$ sqrt [c] for individual measurements, determine a new sample size using tables 25-30. If the new sample size exceeds the sample size used, further investigation of the EA is necessary
  - In this case,collect additional samples and repeat the hypothesis testing process to verify if the EA can be screened out using the larger combined sample.

Adjusted CLT(Adj-CLT) Left -Tailed Test

- Adj-CLT can be used for both sided tests the Lower as well as the Upper tailed test for unknown mean, μ of skewed distributions. The test can be used for discrete as well as composite samples.
  - For the left-tailed test, the null is H_0: mean >= 2SSL (protective of human health), versus the alternative H_1: mean < 2SSL, with Type I and Type II error rates as 0.05 and 0.2 at SSL/2, respectively.
  - The Adj-CLT test statistic t is given by: $t = \sqrt{n}(\bar{x} - 2SSL) / s$
  - The critical value for the left - tailed test is: $z_a^* = -[z_a + a(1 + 2z_a^2)]$
  - Where the statistic a has been defined earlier.

Chen's Right- Tailed Test

- The test statistic t_2 is then compared with the normal $(1-\alpha)$ 100% critical value z_α
- Where the test statistic t_2 is given by:

$$t_2 = t + a(1 + 2t^2) + 4a^2(t + 2t^3)$$

- and the statistics t and a are given by:

$$a = b / (6.0\sqrt{n}) \qquad t = \sqrt{n}(\bar{x} - 0.5SSL) / s$$

- If the test statistic $t_2 > z_\alpha$, then the null hypothesis is rejected, leading to the conclusion that the EA needs further investigation.

Chen's Test

- If the test statistic $t_2 <= z_\alpha$, the data do not provide enough evidence to reject the null hypothesis, and one should
  - proceed with the DQA process to determine if the sample size used is sufficient to achieve a 100ß % or less chance of incorrectly accepting H_0 when the mean = 2SSL.

DQA for Max Test

- In addition to the condition that Max <2SSL for an EA to be screened out, if Max <SSL/sqrt [c], then no further DQA is needed and the EA needs no further investigation.

- If Max >=SSL/sqrt [c], then for prespecified performance standards (Type I and II errors) and CV^* for an individual observation: $CV^* = CV \text{ sqrt } [c]$, using Table 23, determine a new sample size. If the new sample size exceeds sample size used, further investigation of the EA is required.

- In this case, additional samples need to be collected and the process will be repeated to verify if the EA can be screened out using the larger combined sample.

Chen's Right -Tailed Test

- Chen (JASA,1995) derived an upper tailed test for the unknown mean, μ of skewed distributions. This test can be used for both discrete as well as composite samples.
  - For Chen's test, the null hypothesis is H0: mean <= SSL/2, versus the alternative hypothesis H1: mean > SSL/2 (not protective of human health), with Type I and Type II error rates as 0.2 and 0.05 at 2SSL, respectively.

- Let $x_1, x_2, \ldots, x_n$ represent n discrete or composite samples from an EA with mean μ. The sample mean, variance, and CV are:

$$\bar{x} = \sum x_i / n \qquad s^2 = \sum (x_i - \bar{x})^2 / (n-1) \qquad CV = s/\bar{x}$$

$$\text{and let } b = n\sum (x_i - \bar{x})^3 / [(n-1)(n-2)s^3].$$

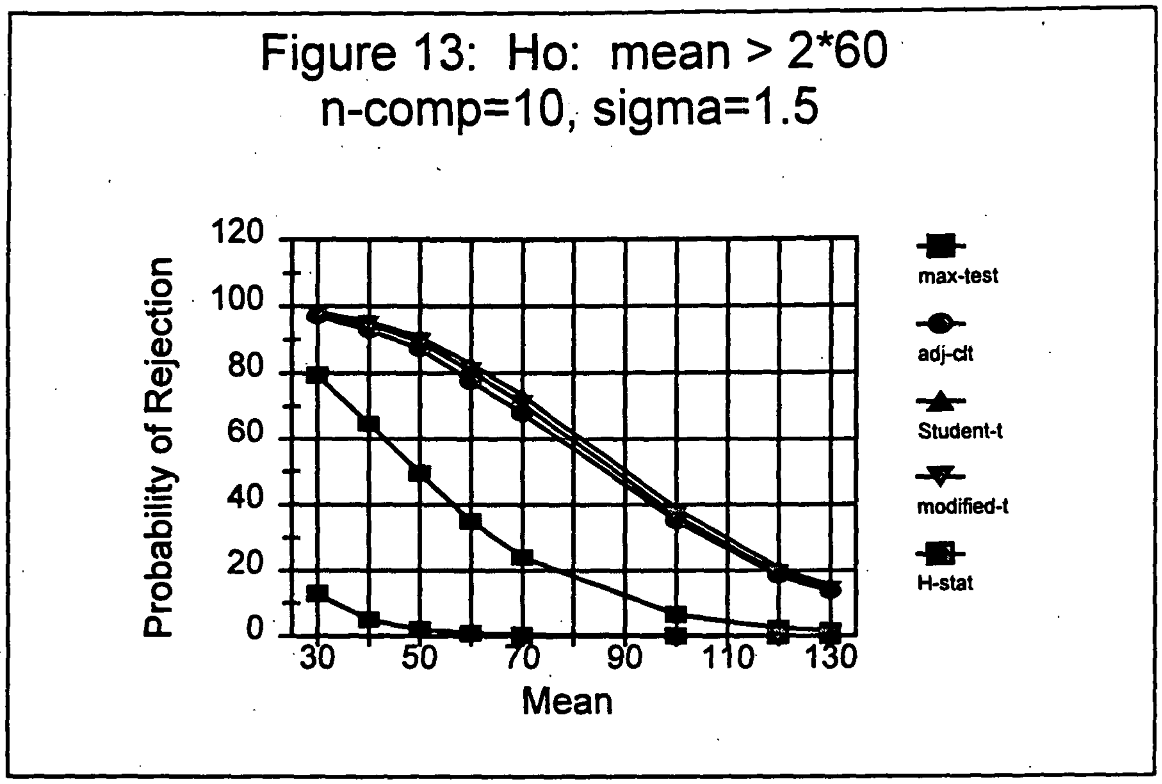


Figure 13: Ho: mean > 2*60
n-comp=10, sigma=1.5
Probability of Rejection
120
100
80
60
40
20
0
30
50
70
90
110
130
Mean
max-test
adj-clt
Student-t
modified-t
H-stat

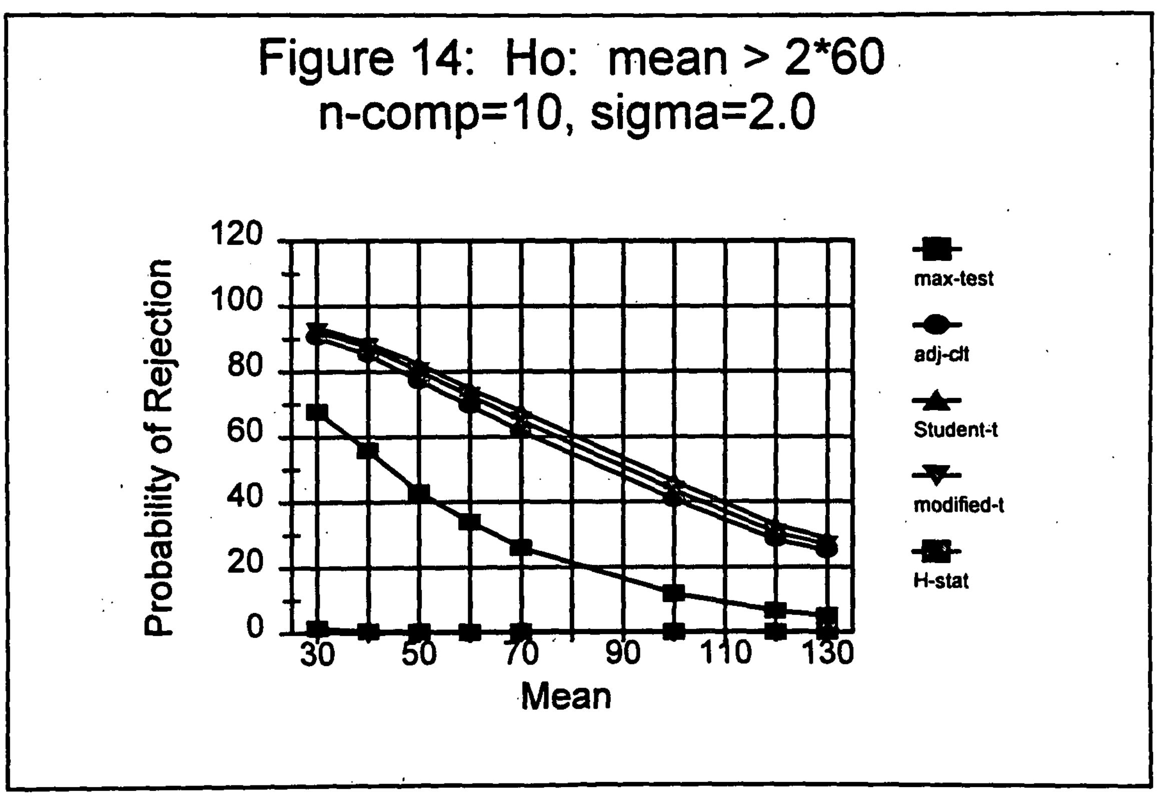


Figure 14: Ho: mean > 2*60
n-comp=10, sigma=2.0
Probability of Rejection
120
100
80
60
40
20
0
30
50
70
90
110
130
Mean
max-test
adj-clt
Student-t
modified-t
H-stat

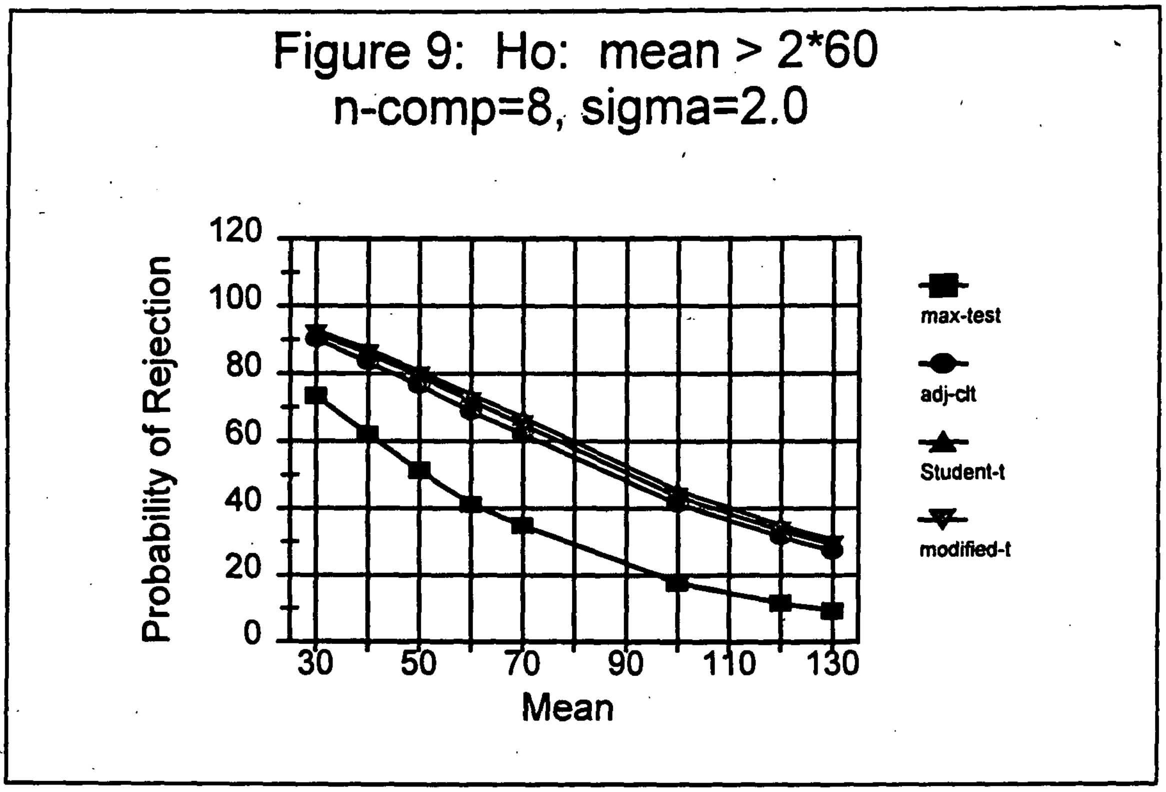


Figure 9: Ho: mean > 2*60
n-comp=8, sigma=2.0
Probability of Rejection
120
100
80
60
40
20
0
30
50
70
90
110
130
Mean
max-test
adj-clt
Student-t
modified-t

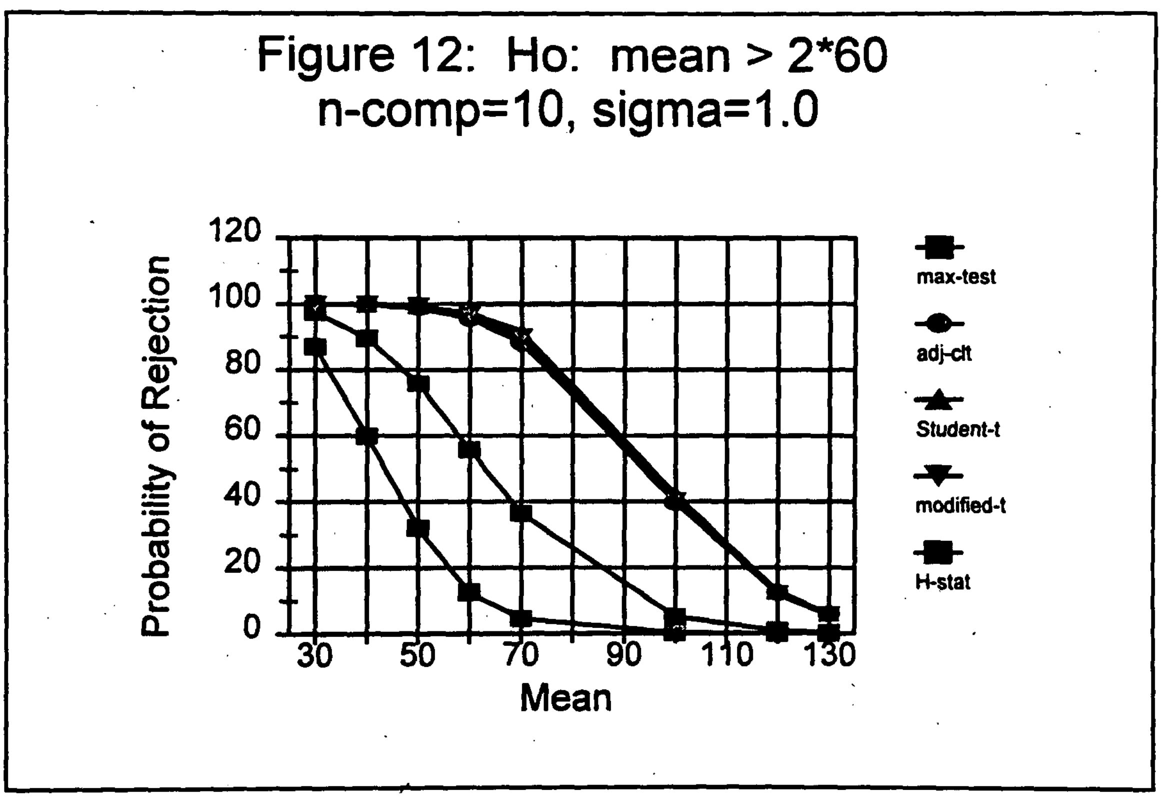


Figure 12: Ho: mean > 2*60
n-comp=10, sigma=1.0
Probability of Rejection
120
100
80
60
40
20
0
30
50
70
90
110
130
Mean
max-test
adj-clt
Student-t
modified-t
H-stat

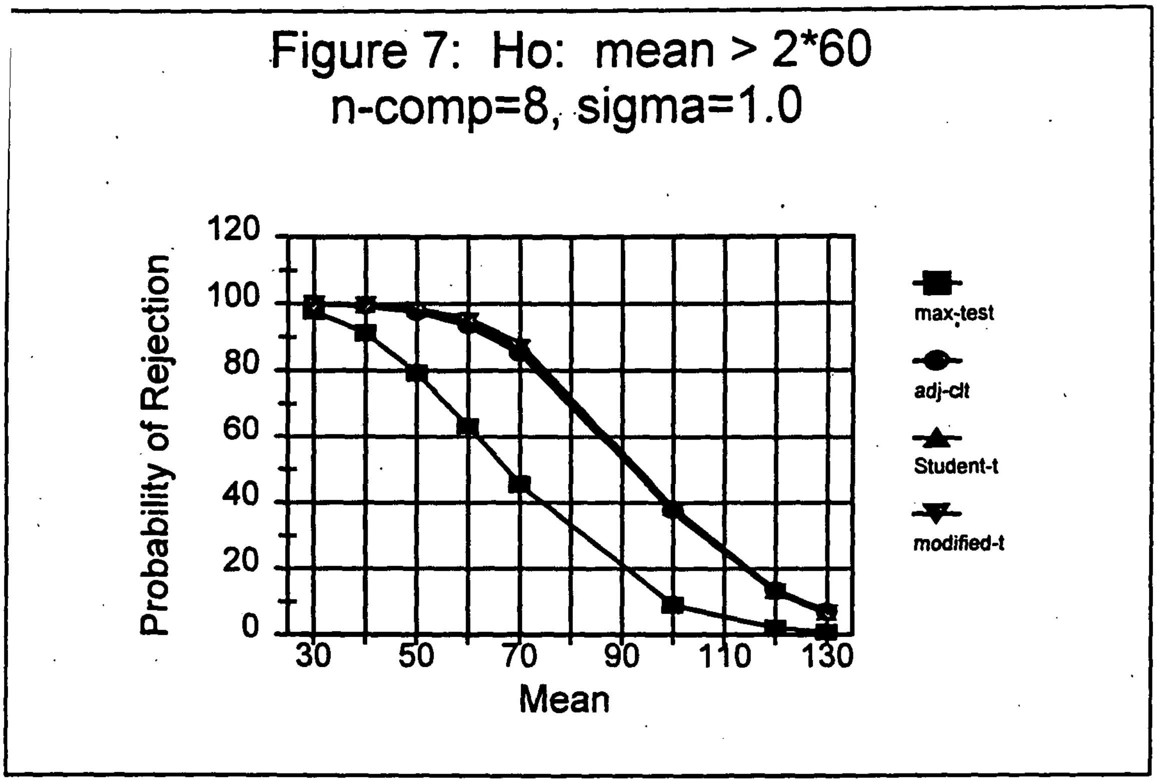


Figure 7: Ho: mean > 2*60
n-comp=8, sigma=1.0
Probability of Rejection
120
100
80
60
40
20
0
30
50
70
90
110
130
Mean
max-test
adj-clt
Student-t
modified-t

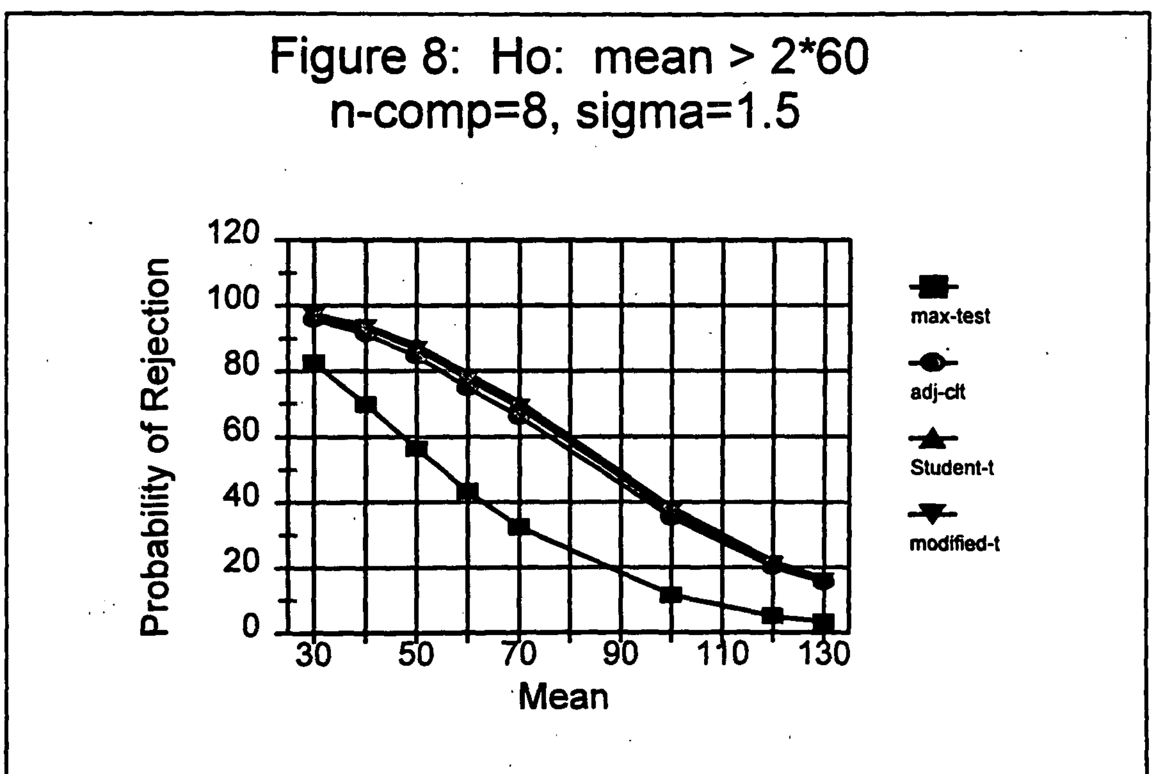


Figure 8: Ho: mean > 2*60
n-comp=8, sigma=1.5
Probability of Rejection
120
100
80
60
40
20
0
30
50
70
90
110
130
Mean
max-test
adj-clt
Student-t
modified-t

Figure 3: Ho: mean > 2*60
n-comp=5, sigma=1.5

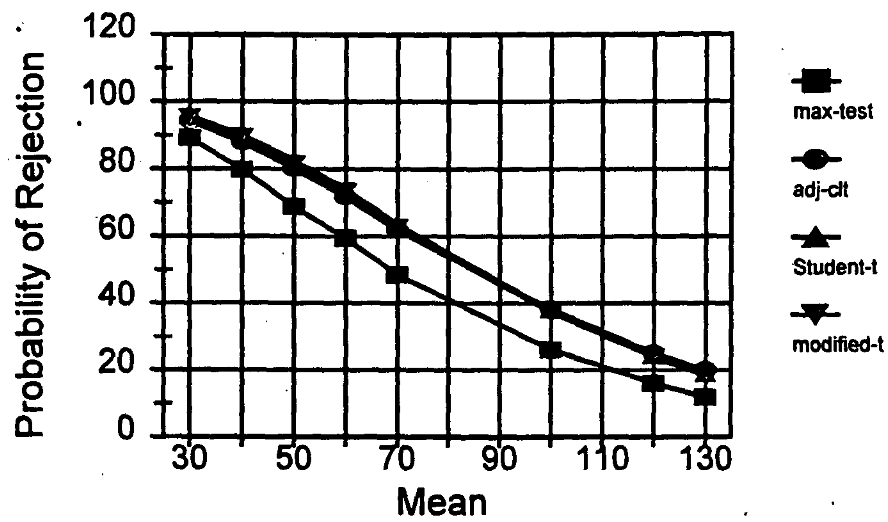
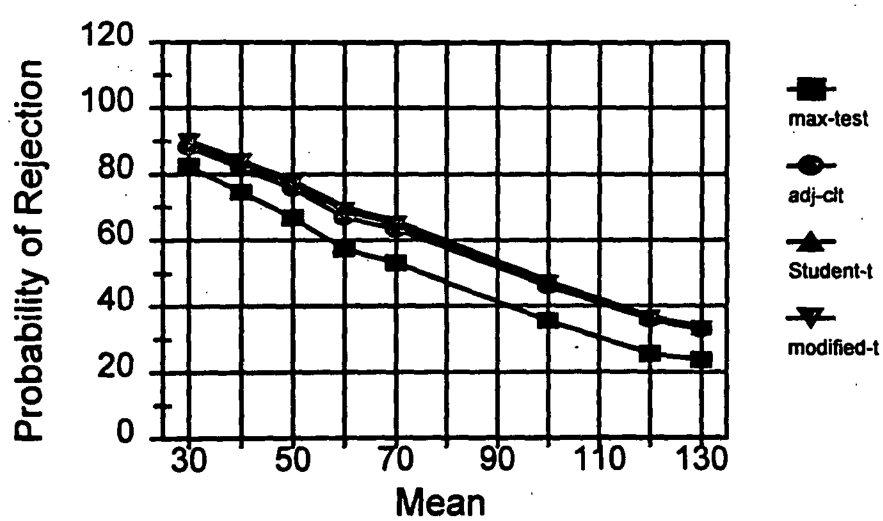


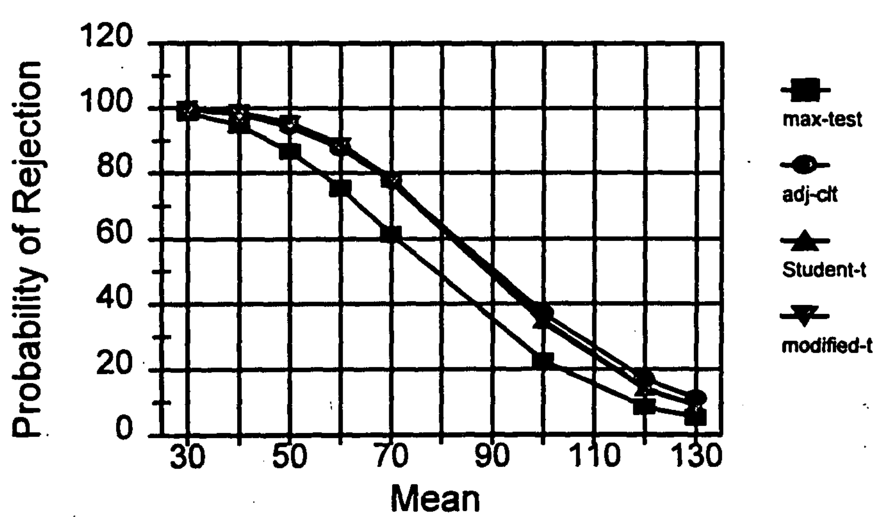
Figure 4: Ho: mean > 2*60
n-comp=5, sigma=2.0



Max Test - for Composite Samples

- Max test is not consistent. For a consistent test, power increases with the sample size.
- For Max test, for fixed value of c (the number of specimens in a composite sample), the Type II error increases (and power decreases) as the number of composite samples n increases as can be seen in figures 2, 3, 7, 8, 12, 13 and 14.
- For values of sigma <=1.0, Max test meets performance standards fairly well; actually all other consistent left - tailed tests (except the H-UCL) perform well for sigma <= 1.0 as can be seen in figures 2, 7, and 12.
- From these figures 2-4, 7-9, and 12-14 note that the Max test does control the Type I error at 2SSL.
- The Type II error rate decreases as specimens, c in a composite sample increases (not in graphs).

Figure 2: Ho: mean > 2*60
n-comp=5, sigma=1.0



Max Test - for Composite samples

- As mentioned earlier, statistical equations may result in a larger number of discrete samples than the resources allow.
  - Compositing is then used to estimate the mean concentration of the COPC in an EA.
  - Using the available information, or an expert opinion get an estimate of CV, so that number, n of composite samples can be determined. A conservative value of CV=2.5 can be used when no information is available.
  - The maximum concentration from composite samples is used as a conservative estimate of the mean of the COPC.
  - The null H_0: mean >= 2SSL, versus H_1: mean < 2SSL, with Type I and Type II error rate as 0.05 and 0.2 at SSL/2.
  - The Max test compares the maximum concentration of the sample with 2SSL.

Max Test - for Composite Samples

- Let $x_1, x_2, \ldots, x_n$ be n composite samples (of c discretes) from an EA with unknown mean μ. Sample mean, variance, and CV are:

$$\bar{x} = \sum x_i / n \qquad s^2 = \sum (x_i - \bar{x})^2 / (n-1) \qquad CV = s/\bar{x}$$

  - Let Max be the maximum of these n composite samples.
  - If Max >= 2SSL, then the EA needs further investigation.
  - If Max < 2SSL, and DQA indicates that the sample size is adequate, then no further investigation is warranted.
  - Max test controls the Type I error rate at 2SSL, but does not provide good control of Type II error at 0.5 SSL .

Screening a Decision Unit- EA Using Statistical Procedures

- Procedures based upon tests of hypotheses.
  - Max Test - composite samples only.
  - Chen's Test - composite or discrete samples.
  - Test based on the adjusted Central Limit Theorem (CLT) - for skewed data distributions - composite or discrete samples.

- Procedures using the UCL of the mean COPC.
  - H-UCL of the mean CPOC for lognormal distribution - for discrete samples.
  - UCL of mean based upon Chebychev Inequality - composite or discrete samples.

Power Comparison of These Procedures

- **Power (=probability of rejecting H_0) Curves.**
- Power curves are used to compare the performance of various procedures. Higher is the power, the better is the procedure.
  - Power curves help to understand the relationship between mean and confidence levels, and
  - determine an adequate sample size needed to meet standards.
  - Note that power of a test increases with the sample size and decreases as the sd increases.

Data Quality Assessment

- The statistical equations can be used to assess the sufficiency of existing data to resolve decisions after sampling and analyses have taken place.
- The purpose of DQA is to evaluate if the DQOs are met, and also to determine if more samples need to be collected so that decisions are acceptable to all relevant parties (e.g., PRP, regulatory agencies).
- The purpose is to help make informed decisions. If you don't like the answers you get and choose to use fewer numbers of samples, that's okay. It's your decision and the purpose of this step is to help make informed decisions whatever they may be.

Screening a Decision Unit- EA Using Statistical Procedures

- Statistical procedures exist to determine if a decision unit can be screened out. These procedures are based upon Upper Confidence Limit (UCL) of mean COPC and tests of hypotheses about the mean concentration of a COPC.
- The SSL Guidance document assumes that data distribution is positively skewed such as lognormal, gamma, and Weibull.
- However the sample sizes given in Table 23, and tables 25-30 of the SSL guidance are based upon less skewed gamma distribution. Depending upon the parameters, a lognormal distribution can be highly skewed and the sample sizes given in tables 25-30 may not be directly applicable.

Systematic Sampling

- Systematic sampling typically involves placing a spatial grid over the site map and selecting a random starting point within one of the grid cells. Sampling points in other cells are placed in a deterministic manner relative to the random starting point.

- These sampling points may be arranged in a pattern of squares, triangles, or rectangles. The result of either approach is a simple pattern of equally spaced points at which sampling will be performed.

- Composites of 4-5 aliquots are sometimes taken within each cell.

Judgmental Sampling

- In authoritative (biased) sampling, an expert familiar with the site dictates where and when to take samples.

- Judgmental sampling data cannot be used to draw statistical conclusions for the site of concern, as the conclusions drawn from judgmental sampling can apply only to those individual samples.

  - For example, if the objective is to identify the location(s) of leaks, one will only be interested in those sampling locations.

- The biased sampling results cannot be used to interpolate and estimate concentrations at other locations throughout the site.

Composite Samples

- To avoid confounding effects, compositing should be avoided when dealing with correlated COPCs.
  - Avoid compositing samples with volatile compounds due to the potential analyte losses which may occur during compositing.
  - Compositing should also be avoided if a parameter other than the mean is of concern (e.g., proportions, sd, geometric mean).
  - Compositing may not be appropriate in cases with heterogeneous soil matrices (e.g.,varying particle sizes, foreign objects, organics).
- Thus, when analytical costs are high, cost-effective plans can sometimes be achieved by compositing physical samples prior to analysis. For the same analytical cost, composite sampling allows a larger number of sampling units and locations to be selected than could have been selected using discrete sampling.

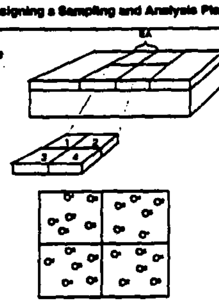
Systematic Sampling

- Systematic sampling using a spatial grid is usually used with contaminated sites to detect hot spots, or for site characterization during RI/FS using geostatistical techniques such as kriging and variogram modeling.
- It may be used to collect soil samples from a landfill, to locate wells for collection of groundwater samples, or to collect aqueous sediments from the bottom of a lake.

Composite Samples

- Compositing represents a physical rather than mathematical mechanism for averaging. In compositing, several individual specimens are physically mixed and homogenized, and one or more subsamples are selected from the mixture for analysis.
  - Note that in surface soil screening the objective is to estimate the mean EA concentration of a COPC, known as exposure point concentration (EPC) term; the physical averaging that occurs during compositing is consistent with the intended use.
  - The individual samples in a composite should be taken across the EA, so that the analytical result of each composite will represent an estimate of the mean concentration of the COPC for the entire EA.

Designing a Sampling and Analysis Plan for Surface Soils



1. **Subdivide Site Into EAs** — For surface soils, the individual unit for decision making is an "EA," or exposure area. It measures 0.5 acre in area or less.
2. **Divide EA Into a Grid** — This step defines the number of specimens (N) that will make up one composite sample.
3. **Organize Surface Sampling Program for EA** — Placement of sample locations on the grid was developed using a default sample size of 6 (which is based on acceptable error rates for a CV of 2.5) and a stratified random sampling pattern.

If the EA CV is suspected to be greater than 2.5, use the table below to select an adequate sample size or refer to the TBD for other sample design options.

Probability of Decision Error at 0.5 SSL and 2 SSL Using Max Test

	CV=2.5[a]		CV=3.0		CV=3.5		CV=4.0	
Sample Size[b]	$\varepsilon_{0.5}$[c]	$\varepsilon_{2.0}$[d]	$\varepsilon_{0.5}$	$\varepsilon_{2.0}$	$\varepsilon_{0.5}$	$\varepsilon_{2.0}$	$\varepsilon_{0.5}$	$\varepsilon_{2.0}$
	C = 4 specimens per composite[e]							
6	0.21	0.06	[illegible]	0.11	0.31	0.11	0.35	0.16
7	[illegible]	0.05	0.31	[illegible]	[illegible]	0.09	0.41	0.15
8	[illegible]	0.04	[illegible]	0.05	0.42	0.07	0.41	0.09
9	[illegible]	0.03	[illegible]	0.04	0.44	0.07	0.46	0.08

[a] The CV is the coefficient of variation for individual, uncomposited measurements across the entire EA, including measurement error.
[b] Sample size (N) = number of composite samples
[c] $\varepsilon_{0.5}$ = Probability of requiring further investigation when the EA mean is 0.5 SSL
[d] $\varepsilon_{2.0}$ = Probability of not requiring further investigation when the EA mean is 2.0 SSL
[e] C = number of specimens per composite sample, when each composite consists of points from a stratified random or systematic grid sample from across the entire EA.

NOTE: All decision error rates are based on 1,000 simulations that assume that each composite is representative of the entire EA, half the EA has concentrations below the limit of detection, and half the EA has concentrations that follow a gamma distribution (a conservative distributional assumption).

Stratify The Population - Surface Soil

- Identify areas which may be contaminated and can not be ruled out from further investigation.
  - Areas that are suspected to be contaminated are the primary subject of surface soil investigation.
  - Sampling scheme discussed in the SS Guidance is most suited for these areas which may be contaminated and cannot be designated as uncontaminated.
  - Geostatistical techniques such as variogram modeling and Kriging can be used to characterize these areas of the site. A systematic grid sampling pattern needs to be used for sample collection. However, spatial statistical techniques are beyond the scope of the SS Guidance.

Simple Random Sampling

- Simple random sampling is the simplest type of probability sampling where every possible sampling unit of the target population has an equal chance of being selected.

- Simple random sampling is often used in the early stages of an investigation in which little is known about any systematic variation within the site - such as those areas which might be contaminated and cannot be ruled out from investigation.

- In order to estimate the average COPC, collect an appropriate number of samples (discrete or composite) needed to meet the performance standards.

- This may result in an extensive sampling effort at high costs which may not be feasible within the available resource constraints.

Stratify the Population - Surface Soils

- Using existing data, maps, expert opinions, and visual inspection, stratify population into homogeneous strata with similar contaminant concentration patterns.
- Various strata may require different levels of investigation.
  - These strata may have different variability (sds), therefore a different sampling design may be needed for each stratum.
  - Since, all EA within a stratum should exhibit similar concentrations for a COPC, one site specific sampling design can be used for all EA within that stratum.
- Thus stratification can characterize the site more effectively and help reduce evaluation and remediation costs .

Stratify The Population - Surface Soil

- Identify areas unlikely to be contaminated by site activities.
  - Undisturbed by site hazardous - waste -generation activities. These areas are typically screened out from further investigations after confirmation. Site managers may take a few confirmatory samples to verify this assumption.
- Identify site areas which are known to be highly contaminated.
  - These are areas directly impacted by site activities, which will be further investigated and characterized during RI/FS.
  - These contaminated areas are targeted for subsurface sampling.

Site ABCD - LN(0.71, sd=1.78, CV=2.5)

- Composite surface soil samples are generated from a lognormal population LN(0.71,σ=1.78), SSL=10 ppm with CV= 2.64 of raw individual observations (50 discrete samples) in original units.

- Xylene concentrations of 10 composite surface soil samples of 5 specimens each from site ABCD are:13.12, 3.81, 2.73,1.86, 27.70, 6.55, 36.12, 3.86,5.36, and 1.45, with mean and sd as 10.26, and 12.05 and CV of composites = 1.2.

- The null for Land's UCL test and Max left tailed test: H_0: Mean >=20 ppm, versus H_1: Mean < 20 ppm.
- The null hypothesis for Chen's and Adj-CLT right-tailed test: H_0: Mean <=5 ppm, versus H_1: Mean > 5 ppm.
- Error ate at 2SSL is 0.05 and error rate at 0.5 SSL is 0.2.

Basic Sampling Types

- Surface soil sampling strategy is designed to collect the soil samples needed to evaluate exposure via direct ingestion, dermal contact, and inhalation of fugitive dust pathways.

- There are several types of sampling schemes but they are all combinations or variations of three basic types of sampling:
  - 1. Simple Random Sampling
  - 2. Systematic Sampling, and
  - 3. Judgmental (authoritative) Sampling

- Before using a sampling scheme:
  - Stratify the population of interest into homogeneous regions.
  - Determine the type of samples to be collected - discrete or composites.

Site XYZ - LN(2.56, sd=1.75)

- Composite surface soil samples are generated from a lognormal population LN(2.563, σ=1.75), SSL=60 ppm, and CV for raw individual observations (50 discrete samples) as 2.59.

- B(a)P equivalents of 10 composite surface soil samples of 5 specimens each from site XYZ are:15.672, 16.162,4.984,18.458, 45.210, 7.553, 26.285, 30.503, 110.403, and 16.230 with.sample mean and sd as 29.15, and 30.83, and CV of composites = 1.058.

- The null for Land's UCL test and Max left-tailed test: H_0: Mean B (a)P>=120 ppm, versus H_1: Mean B(a)P < 120 ppm.
- The null hypothesis for right-tailed Chen's test, and Adj-CLT : H_0: Mean B(a)P<=30 ppm, versus H_1: Mean B(a)P >30.
- Error rate at 2SSL is 0.05 and error rate at 0.5 SSL is 0.2.

Site ABC - LN(1.62, 2.42, CV=1.5)

- Composite surface soil samples are generated from a lognormal population LN(1.62, σ=2.42), SSL=60 ppm with CV = 5.31of raw individual observations (50 discrete samples) in original units.

- B(a)P equivalents of 10 composite surface soil samples of 5 specimens each from site ABC are: 492.699, 58.605, 3.733, 15.185, 12.780, 8.555, 24.838, 11.430,10.781, and 10.312 with mean and sd as 64.89, and 151.12 and CV of composites = 2.33.

- The null for Land's UCL test and Max left-tailed test: H_0: Mean B (a)P>=120 ppm, versus H_1: Mean B(a)P < 120 ppm.
- The null hypothesis for right-tailed Chen's test, Adj-CLT, H_0: mean B(a)P<=30 ppm, versus H_1: Mean B(a)P > 30.
- Error rate at 2SSL is 0.05 and error rate at 0.5 SSL is 0.2.

Sample Size Determination

- Statistical equations can be used to:
  - Determine the number of samples (simple random sampling) required to meet DQOs with prescribed Type I and Type II error rates within a tolerable error margin, D = 2SSL-SSL/2.
  - Determine the systematic sampling grid necessary to detect "hot spots".
- The discrete sample size needed for estimating the average concentration of an EA (assuming normal distribution) can be determined using the following equation. This may yield a larger sample size than allowed within the available resources, therefore compositing is sometimes used to reduce analytical costs.

$$n = s^2 (z_{1-\alpha} + z_{1-\beta})^2 \; / \; (2SSL - SSL/2)^2 \; + 0.5 z_{1-\alpha}^{\;2}$$

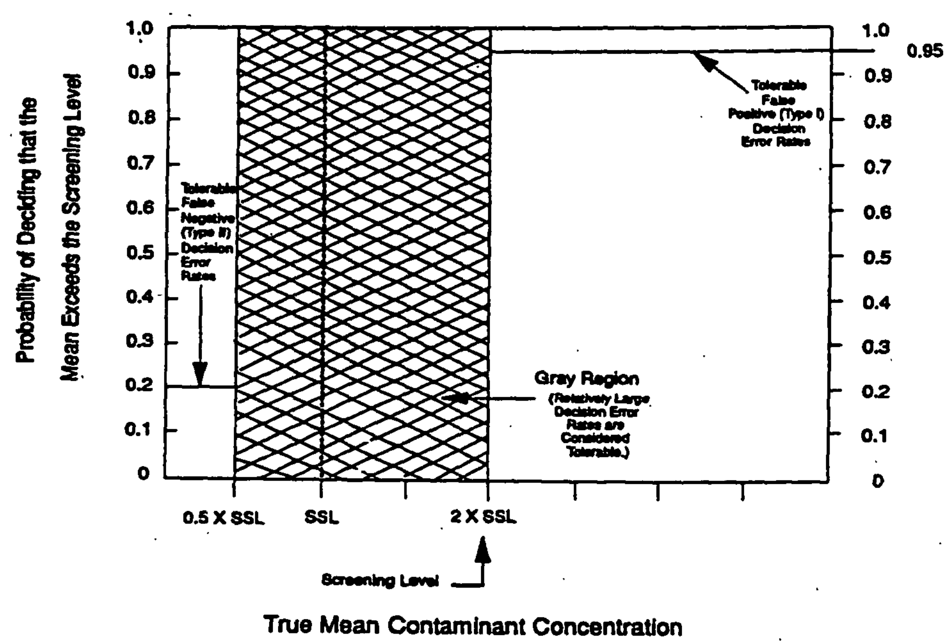
- *s* is obtained using the available information or an expert opinion.

Sample Size Determination

- The sample sizes given in tables 23, 25-30 of the SS Guidance are based upon 1000 simulations of data from a Gamma Distribution.
  - Those samples are driven by the coefficient of variation (CV) of data in original units.
  - A lognormal distribution is characterized by the mean, μ and sd, σ of the log-transformed variable.
  - For a lognormal distribution, (highly skewed - common in environmental applications), CV of data in original unit is a function of the standard deviation (sd), σ and is given by:

$$CV = \sqrt{[\exp(\sigma^2) - 1]}$$

  - For lognormal distribution as σ is increases, CV and skewness both increase. The sample sizes listed in these tables may not be appropriate for highly skewed lognormal distribution.



Decision performance goal diagram.

Optimize the Design to obtain Data

- The design step determines how many samples are needed for decision making and to meet the performance standards, and which type of sampling plan (e.g., simple random, stratified random, judgmental) is required.
- For residential land use, an individual is assumed to move randomly across an EA over time, spending about equivalent amounts of time at each location. Thus for surface soil sampling, the COPC concentration contacted over time is best represented by spatially averaged concentration over the EA.
- Using statistical equations, an optimal sample size can be determined to estimate this average and meeting the performance standards.

Specify Limits on Decision Errors

- Type I decision error for left - tailed test is considered more serious as its consequences include risk to human health and environment, and therefore a stringent limit of 0.05 is used.

– Type II Error (ß) is the probability of not rejecting H_0 when in fact it is false. This type of error is also known as false negative decision error rate.

– Consequences of a false negative decision include unnecessary cleanup expenditure (for Max, Land's tests).

  - Therefore, a less stringent limit of 0.2 is used for Type II error rate, ß.

– Power (1- ß): Power of a test is the probability of rejecting the null hypothesis, H_0. It is desirable for a test to have high power with a value of about 0.20 at SSL/2 and a value of 0.95 or more at 2SSL.

Gray Area - Performance Standards

- Typically, SSL represents a conservative threshold (mean) value for a COPC. Therefore, to be protective of human health and also to guard against unnecessary cleanup expenditure, the SSL Guidance defines the gray area as the interval: SSL/2 to 2SSL.

- When the true mean COPC is in gray area, the consequences of the two decision errors are considered minor which begin to be significant near the boundary points SSL/2 and 2SSL. In gray region, decisions are too close to call as data may not provide conclusive evidence of rejecting or accepting the null, H_0.

- Type I (α) and Type II (ß) error rates are set at 0.05(0.2) and 0.2 (0.05) for left -tailed (right -tailed) test respectively.
  - For left-tailed test: Type I error rate at mean, 2SSL <=0.05.
  - For left-tailed test:Type II error rate at mean, SSL/2 <=0.2.

Hypotheses are Logical Statements About the Mean COPC

- Equivalently, H_0: mean COPC of an EA>= 2 SSL, versus
- The alternative statement, H_1: the EA meets the cleanup goal, or equivalently, H_1: mean COPC of an EA <2SSL.
  - The null hypothesis defined above has critical region in the left tail and is more appropriate for NPL sites.
- However, for Chen's test, the null and alternative hypotheses are defined in a flipped manner, with critical region in the upper tail (therefore called upper - tailed test):
  - H_0: mean COPC of an EA <= SSL/2, versus
  - H_1: mean COPC of an EA > SSL/2.

Specify Limits on Decision Errors

- Due to uncertainty in data, statistical decisions can be made only with certain types of errors: Type I and Type II while testing for two hypotheses.
  - Develop numerical probability limits that express the decision maker's tolerance for committing these two types of errors as a result of uncertainty in data.
  - Type I Error(α) is the probability of rejecting H_0 when in fact it is true. This error is also known as false positive decision error rate.
    - Type I decision error can result in not remediating a polluted area of the site (using UCL, Max, and Land's tests).

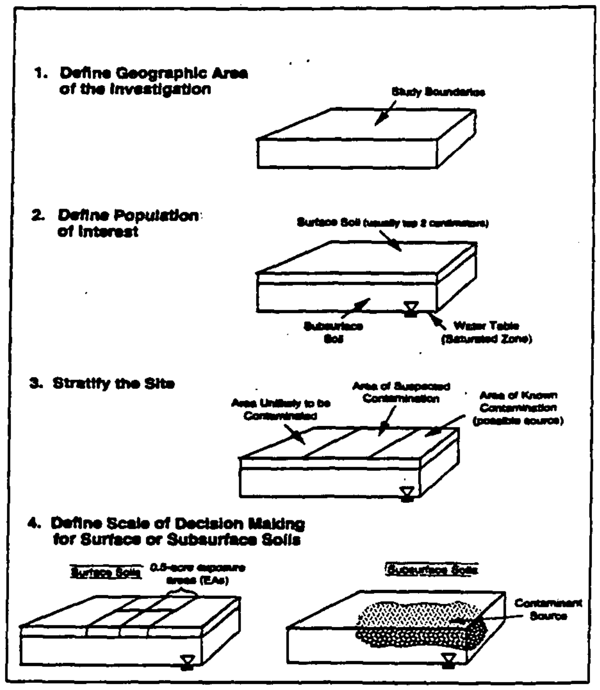
Develop a Decision Rule

- State the objective of data collection - estimation of the mean COPC of an EA for screening purposes.
  - Identify the COPCs - parameters (e.g., mean) of interest and the SSLs with which the parameters will be compared.
- Develop logical statements (hypotheses) about each parameter specifying conditions that would cause the decision maker to choose among alternative actions.
  - Identify all potential actions that could result from data analysis.
    - No action - walk away from the decision unit - EA.
    - Further action needed - investigation, sampling, and possibly remediation.

Hypotheses are Logical Statements About the Mean COPC

- Decision making is done using two statistical hypotheses, the null hypothesis, H_0: the baseline condition, and an alternative hypothesis, H_1: an alternative condition - parameter mean value.
  - Typically the null condition, H_0, is assumed to be true and using the available data, the alternative hypothesis, H_1, bears the burden of proof.
  - To be protective of the environment and human health, at NPL sites the baseline condition, H_0, is stated as :
    - The decision unit (EA) of concern does not meet the cleanup goal and needs further investigation (lower - tailed test).

Defining the Study Boundaries



Translate Objectives into Statistical Hypotheses

- Define logical relations (<, =, and >) specifying how each parameter of interest (e.g., mean COPC) will be compared with the numerical threshold (SSL).
  - Formulate the null hypothesis or the baseline condition : a statement about the population parameter which is presumed to be true unless proved otherwise - an alternative hypothesis (condition) which bears the burden of proof (based upon the collected data).
  - Determine data distribution: normal, lognormal, or other.
  - Identify statistical procedures to be used to draw conclusions.

Identify the Decision

- Does the mean concentration of a COPC in an EA exceed SSL?
- Identify the media, source of contamination, or state records that requires new environmental data to address the contamination problem.
- Identify exposure pathways for surface soil sampling: direct ingestion, dermal absorption, inhalation of fugitive dust.
  - Specify needs for data collection - to estimate the mean COPC.
  - Develop sampling and analysis plan for that decision (surface and subsurface soils, groundwater) to adequately assess contaminant concentrations in that media.

Define the Study Boundaries

- Define spatial and temporal extent of the media under study (e.g., surface or subsurface) that data must represent to make a decision.
- Define the site boundaries.
  - Specify the study area of investigation.
  - Identify the population (e.g., surface soil) of interest.
  - Using all available information and visual inspection, stratify the population into homogeneous sub-areas such as: the clean, contaminated, and regions which may contaminated.
  - Define the smallest scale of decision making unit of each sub-area; for example the 0.5 acre exposure area (EA) for residential land use.

DQO Process in Soil Screening Projects

- The DQO process is a systematic data collection planning process developed by the EPA to ensure that the right type, quality, and quantity of data are collected to support EPA decision making in various environmental applications. There are seven basic elements in the DQO process.

  - State the Problem
  - Identify the Decision
  - Identify the Inputs to the Decision
  - Define the Study Boundaries
  - Develop a Decision Rule
  - Specify Limits on Decision Errors
  - Optimize the Design for Obtaining Data

State the Problem

- Specify the site of concern.
  - Review existing data, identify the population of interest (e.g., segments of the site, surface soils, ground water).

- Summarize the contamination problem requiring investigation and data collection.
  - Identify contaminants of potential concern (COPCs).
  - Identify parameters of interest - e.g., the population mean concentration of a COPC.
  - Compute/Identify numerical value such as the soil screening level (SSL) to which the parameter be compared.
  - Determine if existing data are enough to make this comparison.
  - Identify available resources (e.g., budget, team of experts, time schedule) to address the problem.

Software

- The following software packages can be used to compute the sample size, various test statistics, and the 95% UCLs of the mean.
  - **ProSamp**: Computes the sample size based upon the normal and lognormal distribution assumption for prespecified performance parameters - a common question in Superfund.
  - **ProUCL**:
    - Computes the various $(1-\alpha)100\%$ Upper Confidence Limits (UCLs) of the mean such as:based upon Land's statistic, Chebychev Inequality, t-statistic, Bootstrap and Jackknife procedures, Central Limit Theorem (CLT), Adjusted -CLT, and modified t-statistic for skewed distributions.
    - Computes the test statistics and their critical values for various tests: Max test, Chen's test, t-test, and modified t-test, and Adjusted - CLT for skewed distributions.

Data Collection Needs

- Develop conceptual site model (CSM).
  - Review existing - historical data, state soil surveys, maps, aerial photographs, background data, and confirm information on future residential land use.
  - Consult technical experts - risk assessors, toxicologists, hydro-geologists, and statisticians.
  - Identify sources of contamination, exposure pathways (direct ingestion, dermal contact, inhalation of fugitive dust) and affected media (e.g., surface, sub-surface soils).
  - Identify data gaps.
  - Develop sampling and analysis plan for surface and subsurface soils to adequately assess site contaminant concentrations.

Statistics In Environmental Applications

- Statistical procedures dealing with the estimation and hypotheses testing about the mean of a population of interest (e.g., area of an NPL site) are often used in these applications.

- A 95% Upper Confidence Limit (UCL) of the mean is used:
  - in exposure and risk assessment models to determine the exposure intake to site contaminants,
  - to screen an exposure area (EA) of concern from further investigation by comparing the 95% UCL with the respective soil screening level (SSL) or some action level,
  - to verify the attainment of cleanup levels, and
  - to determine the background level contaminant concentration.

Objectives of Soil Screening Guidance

- The main objective is to provide a tool to help standardize and accelerate the evaluation and cleanup process of contaminated soils at the NPL sites with potential future residential land use.

  - Statistical procedures help identify and verify uncontaminated areas and contaminated areas of the site which may require further investigation and remediation.

  - However, due to data uncertainties, decisions can be made with certain types of decision errors - false positives, and negatives.

  - Statistical issues relevant to SSL guidance will be discussed.

Some Statistical Issues In The USEPA Soil Screening Guidance Document

By
Anita Singh
Lockheed Martin Environmental Services
Las Vegas, Nevada

Statistics In Environmental Applications

- Statistics play an important role in data evaluation and decision making processes at polluted sites.
- Statistical procedures allow extrapolation (estimation) from a set of sampled data to the entire site.
- Statistical procedures can be used to design efficient sampling plans to collect sufficient data: to verify the attainment of cleanup standards, to screen an area of concern from further investigation, and to detect hot spots at polluted sites.
- Spatial mean of an exposure area (EA) best represents the exposure to site contaminants contacted over a period of time.

Patricia Flores-Brown

(Air Modeler)

Region III Air Protection Division
Technical Assessment Branch

Technical SSL Issues and Concepts

The Inhalation Pathway

Particulate **E**mission **F**actor (PEF)

Volatilization **F**actor (VF)

Inhalation of Fugitive Dusts (semivolatile organics and metals in surface soils)

- **Ingestion SSLs are protective for inhalation exposures to fugitive dusts for most organic compounds and metals.**
- **The fugitive dust exposure route need not be routinely considered for organic chemicals and metals in surface soils...** *however chromium is an exception due to the carcinogenicity of* **hexavalent chromium Cr^{+6}.**
- **For most sites, fugitive dust SSLs calculated using the defaults should be adequately protective.**

Derivation of the Particulate Emission Factor - PEF

- **Relates the concentration of contaminant in soil to the concentration of dust particles in the air**
  windblown dust
- **Based on Cowherd's "unlimited reservoir" model .**
- **Represents an annual average emission rate.**

The PEF equation can be broken into two separate models:

- a model to estimate the emissions; and
- a dispersion model (reduced to the term Q/C) that simulates the dispersion of contaminants in ambient air.

PEF (m^3/kg) = Q/C x [illegible] / 0.036 x (1-V) x [illegible]

PEF (m^3/kg) = Q/C x [illegible] / 0.036 x (1-V) x [illegible]

Parameter/Definition (units)	Default
	1.32×10^9 m^3/kg
Q/C = inverse of mean concentration of a 0.5 acre square source (g/m^2-s per kg/m^3) based on 90th percentile (Minneapolis, MN)	90.80 g/m^2-s per kg/m^3
V = fraction of vegetative cover (unitless)	0.5 (50%)
u_m = mean annual windspeed (m/s)	4.69 m/s
u_t = equivalent threshold value of windspeed at 7 m (m/s)	11.32 m/s
F(x) = function depended on u_m/u_t derived using Cowherd et. at. (1985) (unitless)	0.194

PEF Equation Parameters

[illegible]

The generic PEF, using the default values, is 1.32 x 10^9 m^3/kg. It represents an annual average emission rate.

The fraction of vegetative cover, V, ranges from 0 to 1 to represent 0% to 100% land cover.

PEF Equation Parameters

[illegible]

Mean annual windspeed, u_m, ranges in our Region from 2.8 m/s at Elkins, WV to 4.7 m/s at Norfolk

D.C.	3.4 m/s	Scranton	3.8 m/s
Baltimore	4.2 m/s	Lynchburg	3.5 m/s
Harrisburg	3.4 m/s	Norfolk	4.7 m/s
Philadelphia	4.3 m/s	Richmond	3.4 m/s
Pittsburgh	4.2 m/s	Elkins, WV	2.8 m/s

PEF Equation Parameters

$PEF\ (m^3/kg)$ [illegible] $\times\ 3{,}600\ s/h$ [illegible] $\times\ (u_m/u_t)^3 \times F(x)$

Use default values for u_t and $F(x)$. $F(x)$ has a range from 0.19 to about 1.91.

The term $(u_m/u_t)^3$ will range from 0.015 to 0.072 using the windspeeds found in the Region. This is only a difference of a factor of 5.

The Q/C TERM - The Dispersion Model

EPA replaced the Box Model in RAGS Part B with the dispersion model AREA-ST. It has the following characteristics:

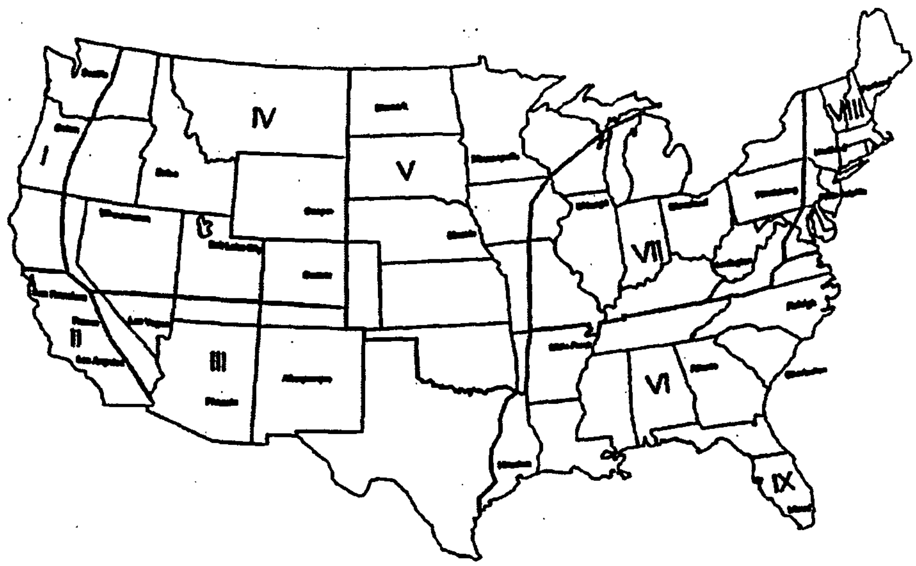
- **dispersion modeling from a ground-level area source**
- **onsite receptors**
- **a long-term/annual average air concentration (necessary for risk assessments)**
- **algorithms for calculating the air concentration for area sources of different shapes and sizes.**

The Q/C TERM - The Dispersion Model

- The dispersion model was run with a full year of meteorological data for 29 U.S. locations selected to be representative of a range of meteorologic conditions across the Nation.
- The results of these modeling runs are presented in Exhibit 11 for square area sources of 0.5 to 30 acres in size.
- When developing a site-specific PEF or VF for the inhalation pathway, place the site into a climatic zone. Then select a Q/C value from Exhibit 11that best represents a site's size and meteorological conditions.

Exhibit 11 - Q/C Values by Source Area, City, and Climatic Zone

U.S. Climatic Zones



The Q/C TERM - The Dispersion Model

To develop a reasonably conservative default Q/C for calculating generic PEF driven SSLs, a default site (Minneapolis, MN) was chosen that best approximated the 90th percentile of the 29 normalized concentrations (kg/m^3 per g/cm^2-s).

The inverse of this concentration results in a default Q/C value of 90.80 (kg/m^3 per g/cm^2-s) for a 0.5 acre site.

Inhalation of Volatiles (volatilization of organic compounds from soils)

- The VF or volatilization factor is used to define the relationship between the concentration of contaminant in soil and the flux of the volatilized contaminant into the air.
- The VF is based on the assumption of an infinite contaminant source and vapor phase diffusion as the transport mechanism.
- The model calculates the maximum flux of a contaminant from contaminated soil and considers soil moisture conditions integral in calculating VF.

Inhalation of Volatiles (volatilization of organic compounds from soils)

- The VF equation can be broken into two separate models:
- a model to estimate the emissions; and
- a dispersion model (reduced to the term Q/C) that simulates the dispersion of contaminants in ambient air.

The Soil Saturation Limit - C_{sat}

- Before using VF, C_{sat} must be calculated to ensure that VF is applicable.
- At C_{sat}, the emission flux from soil to air for a chemical reaches a plateau.
- Volatile emissions will not increase above this level no matter how much chemical is added to the soil.

The Soil Saturation Limit - C_{sat}

- C_{sat} concentrations represent an upper limit to the applicability of the SSLs VF model because a basic principle of the model (Henry's Law) does not apply when contaminants are present in free phase.
- VF-based inhalation SSLs are reliable only if they are at or below C_{sat}.
- Because VF-base SSLs are not accurate for soil concentrations above C_{sat}, these SSLs should be compared to C_{sat} concentrations before they are used for soil screening.

Derivation of the Volatilization Factor

[illegible]

[illegible]

[redacted]

[redacted]

[illegible]

[illegible]

[redacted]

[illegible]

- VF is calculated using chemical-specific properties and either site-measured or default values for soil moisture, dry bulk density, and fraction of organic carbon in soil.
  - ▲ Other than initial soil concentration, air -filled porosity, θ_a, is the most significant soil parameter affecting the final steady-state flux of volatile contminants from soil.
  - ▲ The higher the air-filled porosity, the greater the emission flux of volatile constituents.

VF Equation Parameters

- Among the soil parameters used to calculate VF, annual average water-filled soil porosity (θ_w) has the most significant effect on air-filled soil porosity (θ_a) and hence volatile contaminant emissions. The default value of θ_w (0.15) corresponds to an average annual soil water content of 10 weight percent.
- The soil bulk density (ρ_b) has too limited a range for surface soils (generally between 1.3 and 1.7 g/cm^3) to affect results with nearly the significance of soil moisture content. Therefore, a default bulk density of 1.5 g/cm^3 was chosen to calculate generic SSLs.

VF Equation Parameters

- The default value for f_{oc} (0.006 or 0.6 percent) is the mean value for the top 0.3m of Class B soils.
- To develop a reasonably conservative default Q/C for calculating generic SSLs, a default site (Los Angeles, CA) was chosen that best approximated the 90th percentile of the 29 normalized concentrations (kg/m^3 per g/cm^2-s). The inverse of this concentration results in a default Q/C value of 68.81 g/m^2-s per kg/m^3 for a 0.5 acre site.

Mass-Limit SSLs

- The use of infinite source models to estimate volatilization can violate mass balance considerations, especially for small sources.
- Mass-limit SSLs provide a lower limit to SSLs when the volume of the source is known or can be estimated reliably.
- A mass-limit SSL represents the level of contaminant in the subsurface that is still protective when the entire volume of contaminantion volatilizes over the 30-year exposure duration and the level of contaminant at the receptor does not exceed the health-based limit.

Mass-Limit SSLs

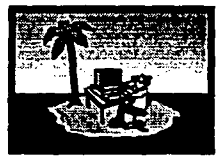
To use mass-limit SSLs:

a. **determine the area and depth of the source,**
b. **calculate both standard and mass-limit SSLs,**
c. **compare them for each chemical of concern, and**
d. **select the higher of the two values.**

Mass-Limit Volatilization Factor

[illegible] / (ρ_b x d x 10^6 g/Mg)

SOIL SCREENING GUIDANCE:



The Soil to Ground Water Migration Pathway
Presented by
Bernice Pasquini
Technical Support Section
HSCD, USEPA Region 3, Philadelphia
May 1999

Subsurface Soil

- Two exposure pathways are evaluated for subsurface soil
  - Inhalation of volatiles.
  - Ingestion of ground water contaminated by leachate produced from contaminated soils.

$Csat = \frac{S}{\rho_b}(Kd\rho_b + \theta_w + H'\theta_a)$

- A soil saturation limit (Csat) is calculated to determine whether the inhalation SSL is applicable for the site.
  - **Definition:** Chemical Concentration at which soil pore air and water are saturated with the chemical and the adsorptive limits of the soil have been reached.
    - soil concentrations > Csat-based SSL, may be indication of DNAPL.
    - SSL defaults to Csat when SSL > Csat

$Csat = \frac{S}{\rho_b}(Kd\rho_b + \theta_w + H'\theta_a)$

Parameter/Definition (units)	Default
Csat/soil saturation concentration (mg/kg)	--
S/solubility in water (mg/L-water)	chemical specific
ρ_b/dry soil bulk density (kg/L)	1.5
Kd/ soil-water partition coefficient (L/kg)	Koc*foc
Koc/soil organic carbon/water partition coefficient (L/kg)	chemical specific
foc/fraction organic carbon in soil (g/g)	0.006 (0.6%)
θ_w/water-filled soil porosity	0.15
H'/dimensionless Henry's Law constant	chemical specific
θ_a/air-filled soil porosity	$n-\theta_w$
n/total soil porosity	$1-(\rho_b/\rho_s)$
ρ_s/soil particle density	2.65

Soil Saturation Limit (Csat)

- **Physical State of Some Organic SSL Chemicals**

Compound	Melting Point (°C)	DNAPL-TYPE COMPOUND?
Benzene	5.5	Yes or No
TCE	-73	Yes or No
benzo(a) pyrene	176.5	Yes or No
anthracene	216.4	Yes or No

Migration to Ground Water SSL: Approach One

- **Soil/Water Partition Equation:**

$$SSL(mg/kg) = C_w[Kd + \frac{(\theta_w + \theta_a \cdot H')}{\rho b}]$$

- ▲ SSL for inorganics (Hg is exception), H'=0
- ▲ If soil gas is lost during sampling, θ_a=0

Migration to Ground Water SSL: Approach Two

- ▲ Leach Tests: Perform leach tests from site contaminated soil.
  - Do not need to collect soil parameters.
  - Still must calculate Dilution factor (need to collect aquifer parameters) and C_w
  - Compare leach test extract concentrations to C_w

Migration to Ground Water SSL-Inherent Assumptions

- Infinite source
- Contamination distributed uniformly
- No attenuation of contamination in soil or ground water
- Instantaneous and linear equilibrium soil/water partitioning
- unconfined, unconsolidated, homogeneous and isotropic aquifer
- receptor well at downgradient edge and screened in plume
- No NAPLs present

$$SSL(mg/kg) = C_w[Kd + \frac{(\theta_w + \theta_a * H')}{\rho_b}]$$

Parameter/Definition (units)	Default
C_w/target leachate concentration (mg/L)	nonzero MCLG, MCL, or HBL*DF
Kd/soil-water partition coefficient (L/kg)	Koc*foc
Koc/soil organic carbon/water partition coefficient (L/kg)	chemical specific
foc/fraction organic carbon in soil (g/g)	0.002 (0.2%)
θ_w/water-filled soil porosity	0.3
θ_a/air-filled soil porosity	n-θw
ρb/dry soil bulk density (kg/L)	1.5
n/soil porosity	$1-(\rho_b/\rho_s)$
ρ_s/soil particle density (kg/L)	2.65
H'/dimensionless Henry's law constant	chemical specific

Kd--Soil-Water Partition Coefficient

- **Non-ionizing Organic Compounds**
  - **Kd=Koc*foc**
  - **Koc is not influenced by pH**
- **Ionizing Organic Compounds**
  - **Kd=Koc*foc**
  - **Koc is influenced by pH**
  - **amines, carboxylic acids, and phenols**
  - **compounds ionize under certain pH conditions**
  - **ionized and neutral species have different sorption coefficients**

Predicted Soil Organic Carbon/Water Partition Coefficients (Koc, L/kg) as a Function of pH: Ionizing Organics

Compound	pH=4.9	pH=6.8	pH=8.0
Benzoic acid	5.5	0.6	0.5
2-chlorophenol	398	388	286
2,4-dichlorophenol	159	147	72
pentachlorophenol	9055	592	410
2,4,6-trichlorophenol	1040	381	131

Kd--Soil-Water Partition Coefficient

- **Inorgranic Compounds (Metals)**
  - **Kd affected by**
    - **pH, sorption to clays, organic matter, ORP, chemical form of metal**
    - **MINTEQ (speciation model) used to estimate Kd for different pHs**

Derivation of the Dilution Factor

- Contaminant dilution when mixing with clean ground water is the only attenuation process adressed in the Dilution Factor equation.
- No default values assigned to input parameters due to uncertainties associated with large variability of hydrogeologic input parameters that affect contaminant migration in ground water.
- DF default for source up to 0.5 acres is 20.
- Because migration to ground water SSLs are most sensitive to the DF, a site specific DF should be calculated on a site-by-site basis.

[redacted] [redacted]

Dilution Factor (DF) $= 1 + \frac{Kid}{IL}$

Parameter/Definition (units)	Default
dilution factor (unitless)	20 (0.5 acres)
K/aquifer hydraulic conductivity (m/yr)	site specific
i/hydraulic gradient (m/m)	site specific
I/infiltration rate (m/yr)	site specific
d/mixing zone depth (m)	Equation12 in Users Guide
L/source length parallel to ground water flow (m)	site specific

[redacted] [redacted]

Estimation of Mixing Zone Depth

- Mixing Zone Depth (d) equation relates this depth to aquifer thickness, infiltration rate, source length, hydraulic conductivity and hydraulic gradient.

$$d = (0.0112\ L^2)^{0.5} + d_a \{1-\exp[(-LI)/(Kid_a)]\}$$

Parameter/Definition (units)
d/mixing zone depth (m)
L/source length parallel to ground water flow
K/aquifer hydraulic conductivity (m/yr)
I/infiltration rate (m/yr)
i/hydraulic gradient (m/m)
d_a/aquifer thickness (m)

- Aquifer thickness should be the upper limit for the [redacted] zone depth.

Mass-Limit SSLs

- Use of infinite source models to estimate volatilization and migration to ground water can violate mass balance, especially for small sources.

- Migration to ground water mass limit SSL is the concentration of a contaminant in the subsurface that is still protective when the entire volume of contamination leaches over the 70-year exposure duration and the level at the receptor does not exceed the health-based limit.

$$\textbf{Mass Limit SSL} = \frac{(C_w * I * ED)}{\rho_b * d_s}$$

Parameter/Definition (units)	Default
C_w/target soil leachate concentration (mg/L)	nonzero MCLG, MCL, or HBL * DF
d_s/depth of source (m)	site-specific
I/infiltration rate (m/yr)	0.18
ED/exposure duration (yr)	70
ρ_b/dry soil bulk density (kg/L)	1.5

Mass Limit SSLs

- **Determine the area and depth of source.**
  - **Actual depth of contamination is unknown, a conservative estimate should be used.**
    - **maximum possible depth in unsaturated zone**
    - **average water table depth -- unless the depth of source is suspected to be within the saturated zone (i.e. below water table).**
- **Both the standard and Mass Limit SSLs should be calculated.**
- **Compare these SSLs for each chemical of concern.**
- **Select the higher of the two values.**

Subsurface Soil Sampling Strategy

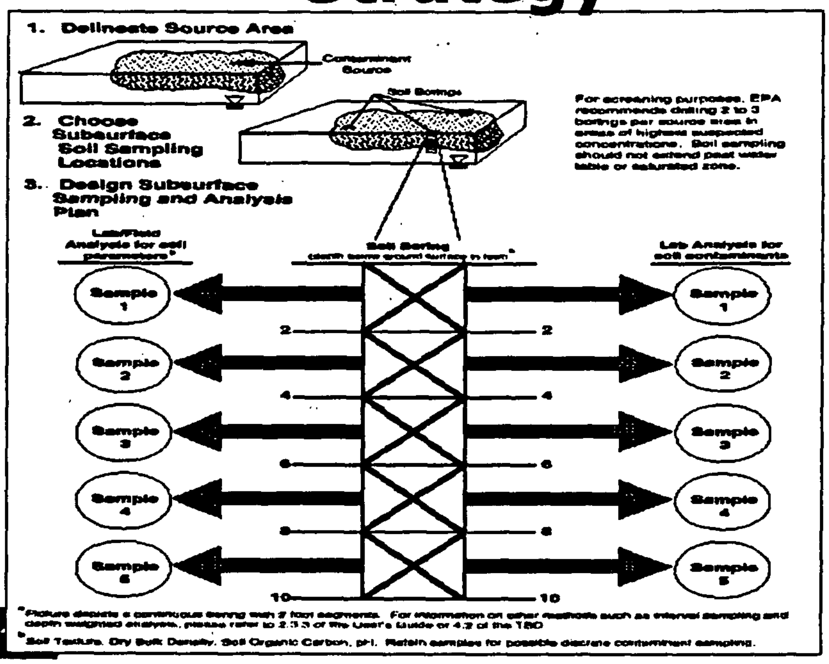
- Develop SSLs for each source
- Collect 2-3 soil borings at suspected source
- Highest mean soil boring contaminant concentration used to screen with SSL
- Maximum depth of contamination encountered < water table depth
- VOC contamination
  - ▲ soil gas surveys and matrix sampling

Development of Contaminant Concentration

- Average contaminant concentration when all sampling intervals are the same.
- When sampling intervals are not equal calculate the depth-weighted average $(\bar{c})$

$$\bar{c} = \frac{\sum_{i=1}^{n} l_i c_i}{\sum_{i=1}^{n} l_i}$$

Subsurface Sampling Strategy



Summary of Migration to GW Pathway SSL

- Important to collect site specific data
  - characterizing soils: foc, pH, dry soil bulk density, soil texture and moisture content
  - characterizing aquifer: hydraulic conductivity, Infiltration rate, aquifer thickness
- Process
  - Compare Csat to SSL for Inhalation and default to the lower of the two as the SSL
  - Calculate mass limit SSL and compare to standard SSL; use the higher of the two as the SSL

US EPA SOIL SCREENING GUIDANCE WORKSHOP

Case Study
(Parameter Simulation Exercises)

US EPA Training Site XYZ is a former wood treater site located 5 miles from a residential neighborhood. There is nothing in the zoning ordinance that will prevent future development of the site for residential use. The owner/ operator treated wood at the site since 1962. Seven years ago, results of water from a well downgradient from Site XYZ were found to contain several chemicals above drinking water standards. These chemicals include:*chromium VI (Cr); arsenic (As); mercury (Hg); benzo(a)pyrene; benzene; 2,4,6-trichlorophenol; trichloroethylene (TCE);* and *xylene (mixed)*

The site was inspected by the State PA/SI program personnel. Some of the above chemicals were found in both the dissolved and NAPL phases in the aquifer. However, the NAPLs were removed under a removal action coordinated between the State and Federal government. All of the above chemicals (with the exception of benzene, TCE, and xylene) have been identified in site surface soils. On the other hand, all of the above chemicals have been identified in site subsurface soils to a depth of 2 m. Depth to groundwater is, on average, 25 m. Contaminant distribution in on-site soils is non uniform.

The site is located in the Coastal Plain Sediments region with geologic formations of a thick regolith of sandy loam over an unconfined sandy aquifer. Other hydrogeologic parameters pertinent to this simulations (K, I, i, d) are as provided in the Attachment 1.

Average particle density (based on literature values) is 2.65 g/cm^3. Values of other predominant soil characteristics are provided in Attachment 2.

A review of available data indicates site contamination of both surface soils (Attachment 3) and subsurface soils (Attachment 4). One exposure area (Source No 1) identified and evaluated for this exercise is about 2025 m^2 (0.5 acre) with a length source parallel to groundwater (L) of 45 m. Exposure and benchmark parameters are as provided in Attachment 5.

Meteorologically, the site is similar to a site placed in Zone V with climatic conditions that are close to those in Minneapolis. The Q/C value is 90.80 (g/m^2-s per kg/m^3) for a 0.5-acre exposure area. Additional meteorological parameters calculated for Site XYZ include:

fraction of vegetative cover (V) of 0.5;
mean annual windspeed (Um) of 4.69 m/s;
equivalent threshold value of windspeed at 7 m (Ut) of 11.32 m/s;
function dependent on Um/Ut of 0.194

Based on the above information and additional information from similar sites close to Site XYZ, the following is known about source area:

1. Land use is currently industrial but with a high likelihood of being residential in the future;
2. Media affected include soil (surface and subsurface) and groundwater;
3. Contaminant release mechanisms include
   - chemical leaching to groundwater supplies,
   - volatilization of chemicals, and
   - fugitive dusts
4. Applicable exposure pathways include
   - soil ingestion,
   - inhalation of fugitive dust, and
   - migration to groundwater
5. No ecological concerns or acute effects are known or determined.

Simulation Exercises

I. Using the minimum and maximum of each of the ranges provided for each parameter in Attachment and given the above information, perform simulations on parameters for:

   a) the groundwater pathway; and
   b) the inhalation pathway.

II. From the output, answer the following, determine the parameter (for each pathway) that is most sensitive towards influencing changes in SSL?

Attachment 1

Ground Water Parameters for Site EPA Training Site XYZ

Heath Region

Hydrogeologic Setting

	Typical	*Minimum*	*Maximum*
Hydraulic Conductivity (m/y)	350		
Hydraulic Gradient (m/m)	0.09		
Aquifer Thickness (m)	15		
Infiltration Rate (m/y)	0.18		

Soil Parameters for Site EPA Training Site XYZ

Source Name	*Source Area*	*Source Depth*	*Source Length*	*Air Porosity*	*pH*	*Organic C*	*Water Content*	*Bulk Density*
Simulation 1 (Default)	2023.5	2	45	0.28	6.80	0.0060	0.15	1.50

Units: Source Area (m2); Source Length = source length parallel to groundwater (m); Source Depth (m); Air Porosity (unitless); Organic C = fraction of organic carbon (g/g); Water Content = average water content (L/L); Bulk Density = dry bulk density (g/cm3)

Attachment 2

Attachment 3

Surface Soil Data Report*
EPA Training Site XYZ

Source No 1

Contaminant	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Background*
2,4,6-Trichlorophenol	147.35	240.49	3456.12	86.80	99.48	1545.45	120.49	77.12	133.60	155.78	6
Arsenic (as carcinogen)	3.79	5.28	3.21	3.22	8.32	3.71	3.41	2.87	2.50	2.66	0.5
Benzo[a]pyrene	353.95	139.91	42.79	43.83	6.43	6.75	1.35	2.49	466.72	7.43	1.9
Chromium VI and compounds	34.98	366.23	30.55	99.35	51.37	107.67	127.70	111.05	249.02	8606.10	12
Mercury (inorganic)	4.00	5.00	3.00	2.40		5.00	2.00	2.20	1.50		

* All concentrations are expressed in Mg/Kg

Subsurface Soil Data Report*

EPA Training Site XYZ

Source Area No. 1

Contaminant	CAS No	Int1	Sample 1	Int2	Sample2	Int3	Sample3	Int4	Sample4	Int5	Sample 5	Int6	Sample 6	Int7	Sample 7	Int8	Sample 8	Int9	Sample 9	Int10	Sample 10	Background
2,4,6-Trichlorophenol	88062	*1*	5.00	*2*	44.00	*1*	3.00	*2*	4.00	*2*	3.00	*2*	3.00			*2*	4.00					
Arsenic (as carcinogen)	7440382	*1*	2.00	*1*	3.00	*1*	2.00	*1*	4.00	*1*	5.00	*1*	6.00	*2*	3.00	*2*	8.00					
Benzene	71432	*2*	6.00	*1*	44.00	*2*	23.00	*1*	5.00	*2*	67.00	*1*	5.00		65.00							
Benzo[a]pyrene	50328	*1*	45.00	*1*	32.00	*2*	33.00	*1*	76.00	*2*	12.00	*1*	34.00	*2*	23.00	*1*	21.00	*1*	56.00			
Chromium VI and compounds	18540299	*2*	45.00	*1*	334.00	*1*	21.00	*2*	4.00	*2*	44.00	*2*	4.00									
Mercury (inorganic)	7439976	*2*	4.00			*2*	3.00	*1*	4.00	*2*	3.00	*1*	3.00	*2*	5.00	*2*	5.00					
Trichloroethylene (TCE)	79016	*3*	5.00	*1*	4.00	*1*	3.00	*1*	3.00	*1*	3.00	*1*	4.00	*2*	98.00	*1*	3.00					
Xylene (mixed)	1330207	*3*	4.00	*2*	54.00	*3*	44.00	*1*	33.00	*1*	33.00	*1*	2322.00	*2*	7.00	*1*	43.00					

Attachment 4

* All concentrations are expressed in Mg/Kg and sampling interval (Int) in meters (m)

Attachment 5

Exposure Parameters for Site EPA Training Site XYZ

Exposure and Benchmark Parameters

Exposure Factors

	Adult	*Child*	*Occupational*	*Residential*
BW/Body Weight (kg)	70.0	15.0		
SA/Surface Area (cm^2/d)	5700	2900		
IRA/Inhalation Rate (m^3/d)	20.0	10.0		
IRS/Soil Ingestion (mg/d)	100.0	200.0	50.0	
ED/Exposure Duration (yr)		6.0	25.0	30.0
ATc/Average Time, carcinogen (yr)			70.0	70.0
EF/Exposure Frequency (d/yr)			250.0	350.0

Other Parameters and Benchmarks

AF/Adherence Factor (mg/cm^2)	0.30
TR/Target Cancer Risk	1.00E-06
THQ/Target Hazard Quotient	1.00

Attachment 6

Soil Screening Guidance Course

Parameter Simulations

	Parameter	*Units*	*Initial Value*	*Range*
		I.	**Groundwater pathway**	
1.	Hydraulic Gradient (i)	-	0.09	0.005 – 0.09
2.	Infiltration Rate (I)	m/yr	0.18	0.09 - 0.25
3.	Hydraulic Conductivity (K)	m/yr	350	350 - 5
4.	Soil pH	-	6.8	5.5 - 8
5.	Depth of Contamination	m	2	0.1 - 8
6.	Organic carbon (foc)	%	0.2	0.01 - 0.10
		II.	**Inhalation pathway**	
1.	Contaminated Area (Q/C)	g/m^2-s /kg/m^3	90.8 (0.5 acre)	53.9 – 90.8
2.	Soil pH	-	6.8	5.5 - 8
3.	Depth of Contamination	m	2	0.1 - 8
4.	Organic carbon (foc)	%	0.6	0.01 - 0.10