Sampling & uncertainty, ”Swedish” perspective

Jenny Norrman, Associate Professor | Engineering Geology | 2020-11-05

Content

- Introduction
- Sampling of soil - uncertainties
- Presentation of selected works (Norrman, Back):
  - Procedure for developing a sampling plan (from 2009)
  - In situ classification of soil (from 2018)
  - Procedure for evaluating data (from 2009)

Use of statistics?

Depends on the purpose of the data sampling

- Searching and identifying – What is there? Where is it?
  - No, mostly not suitable
- Delineation – How spread out is it?
  - No, not suitable
- Characterising – How much is there?
  - Yes, often suitable – but rarely for NAPLs
- Control over time – Is there any change?
  - Yes, often suitable

This Photo by Unknown Author is licensed under CC BY-SA

Different approaches

This Photo by Unknown Author is licensed under CC BY-ND

”Traditional” – go to the field in different rounds

”Dynamic” (US: ”Triad”) – other demands on contracts, type of instruments/field lab, ”real-time” measurements and evaluation

For characterisation of an area/site:

“Traditional” – a collection of observations to characterise an area

“ISM” *(incremental sampling methodology)* – to make few analyses but on samples that have a large sample support (i.e. aim to represent the whole area)

Sampling of different media

Figure from SEPA Report 5894

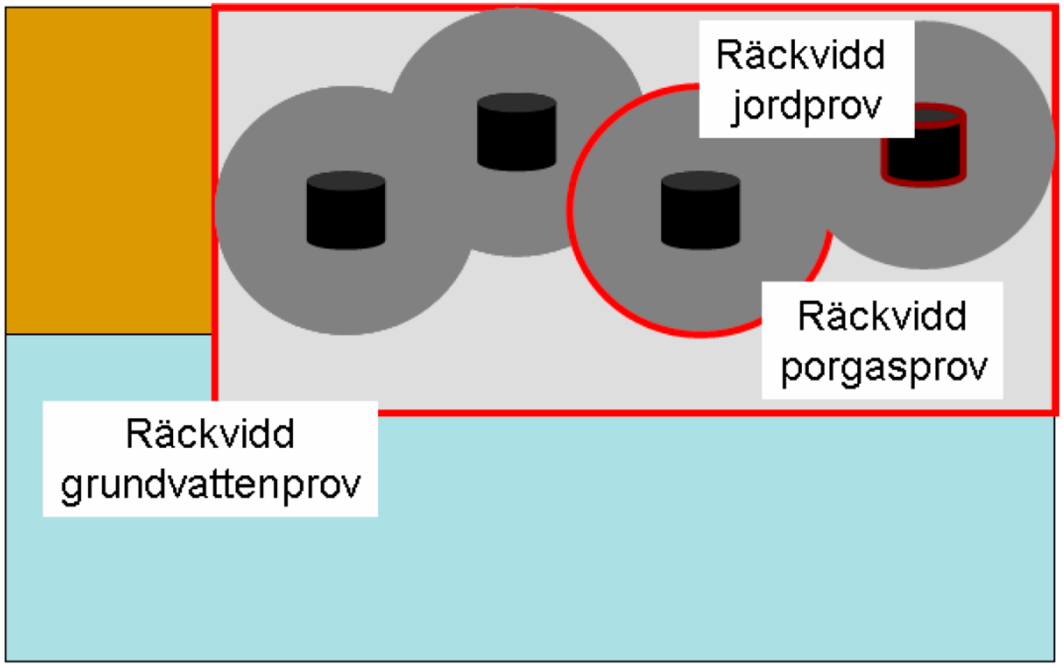
Different representativity of samples in different media

Soil is tricky!

Theory for sampling of particulate matter from 1993 (Pitard)

Research projects in Sweden around 2000 (Back, Gustavsson)

General knowledge in Sweden is increasing, specifically on ISM

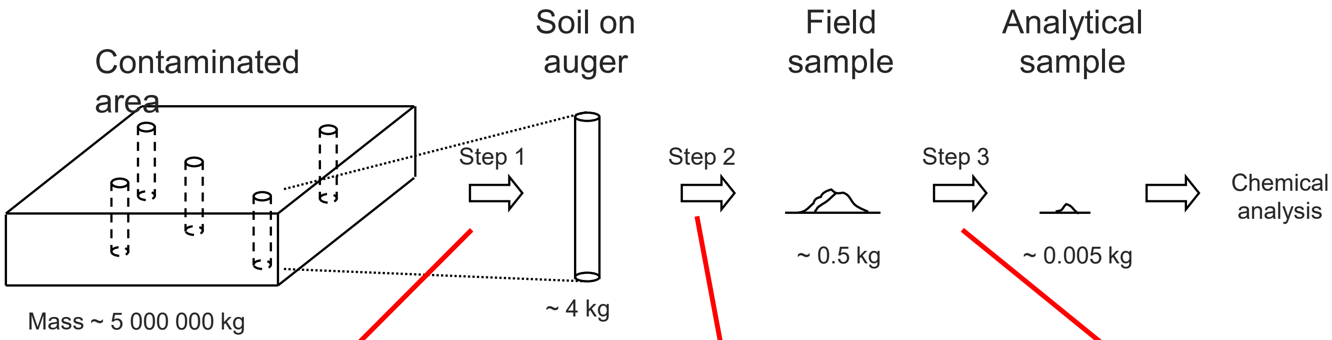


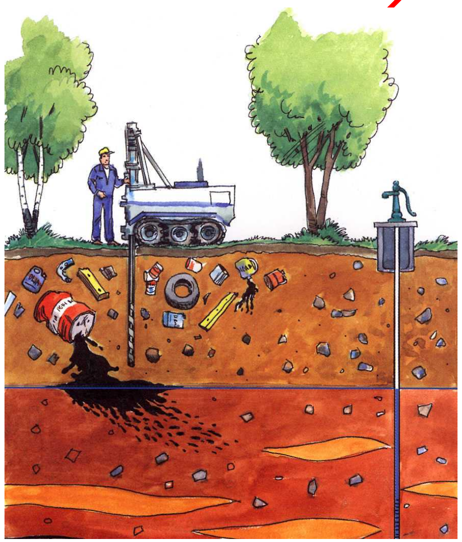
Figur 1. Räckvidd vid provtagning av jord, porgas och grundvatten.

Some projects/reports etc.

- SEPA Sustainable remediation programme
  - Data evaluation (Report 5932, 2009)
  - Sampling strategies (Report 5888, 2009)
  - Course + excel tool (updated excel tool 2017)
- SGI
  - Sampling strategies, in situ classification (Publ. 40, 2018)
  - Website
- SGF
  - Field handbook (Rapport 2:2013)
  - Dynamic environmental investigations (Rapport 3:2017)
  - Data quality control (Rapport 1:2019)
  - Under development: undersökningsportalen
- Short courses, 1 or 2 days
- Several freeware: ProUCL, SADA, VSP, R

Sampling of soil

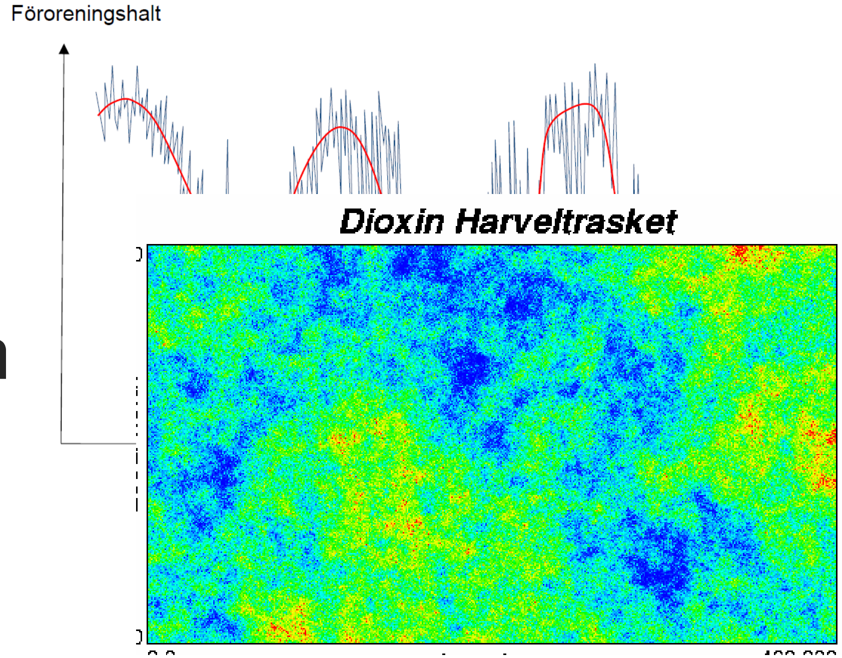




Heterogeneity

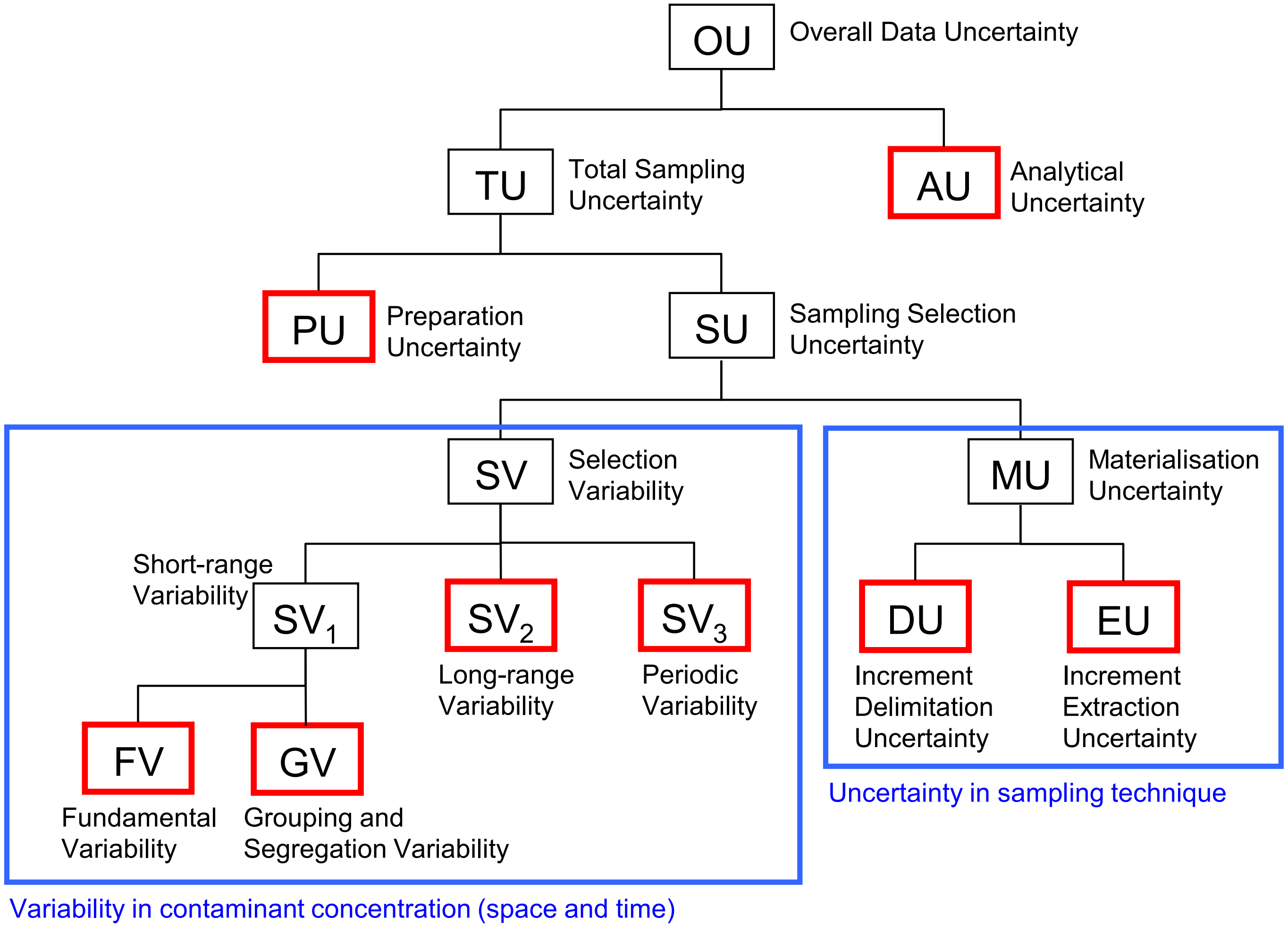
Heterogeneities in the soil will cause uncertainties in our sampling!

In principal, there is no such thing as complete homogeneity ☹

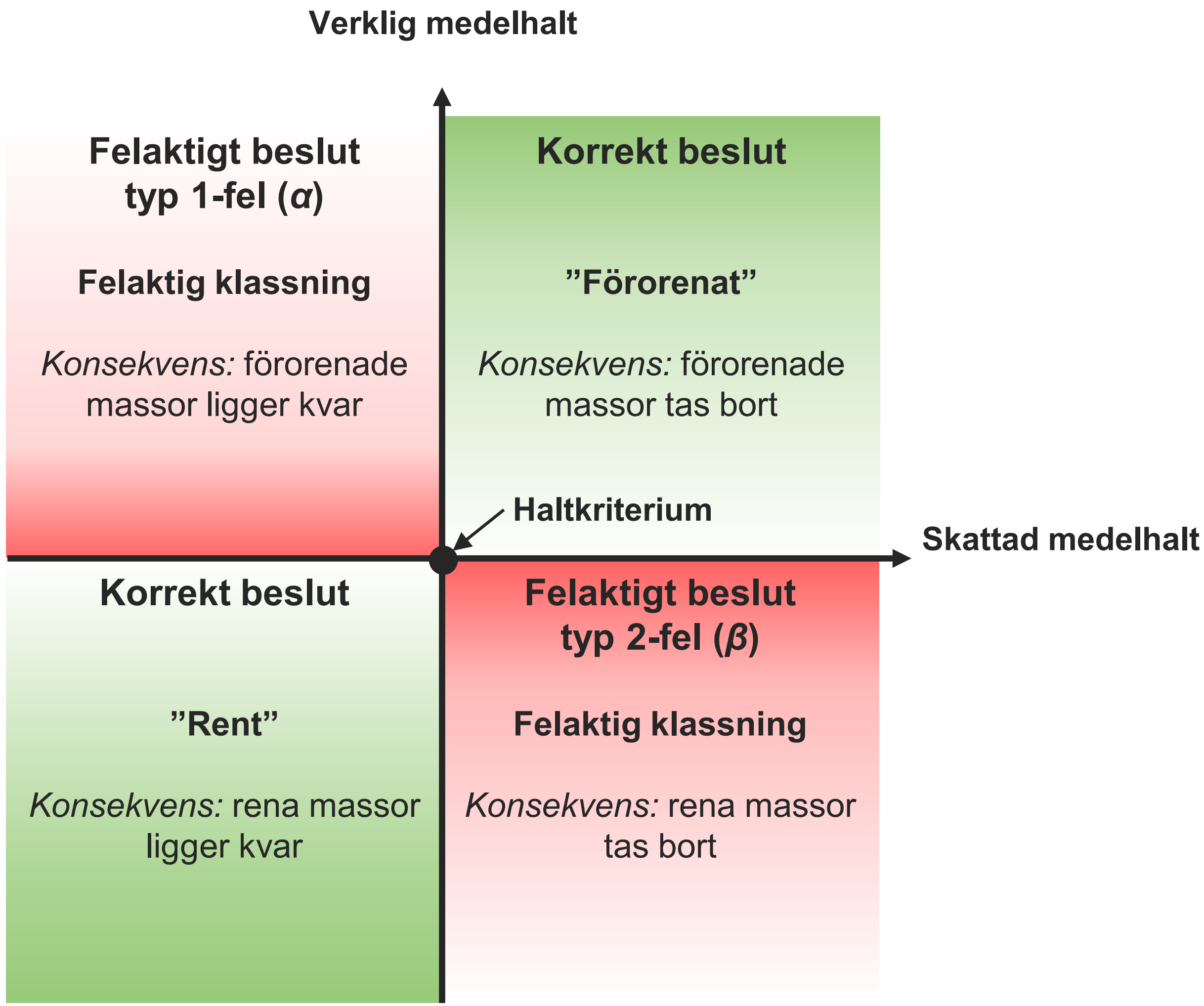


- Heterogeneity on the particle scale (<cm
- Small-scale heterogeneity (cm – m)
- Large-scale heterogeneity (>m)

Sampling uncertainty - the theory of sampling



The ”decision space”



Some selected works

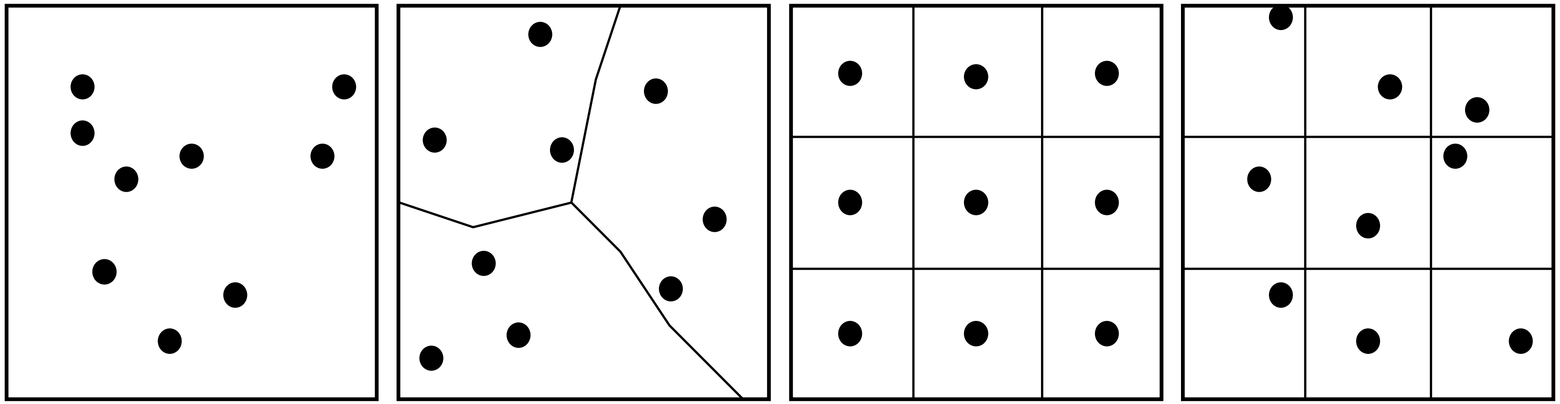
- **A practical approach for developing a sampling plan**
  (SEPA report 5932, Norrman, Back, Engelke, Sego, Wik)
- **Classification of soil in situ**
  (SGI Publ. 40, Back, Norrman, Carling)
- **Suggested approach for statistical data evaluation**
  (SEPA report 5888, Norrman, Purucker, Back, Engelke, Stewart)

A practical approach for developing a sampling plan

1. Define the sampling objective.
2. Decide how prior knowledge should be used.
3. Define spatial and temporal boundaries.
4. Decide the sample support.
5. Choose sampling strategy:
   a. Probabilistic
   b. Judgmental
   c. Search-based
6. Choose sampling design and estimate the number of samples required based on the chosen strategy

A practical approach for developing a sampling plan

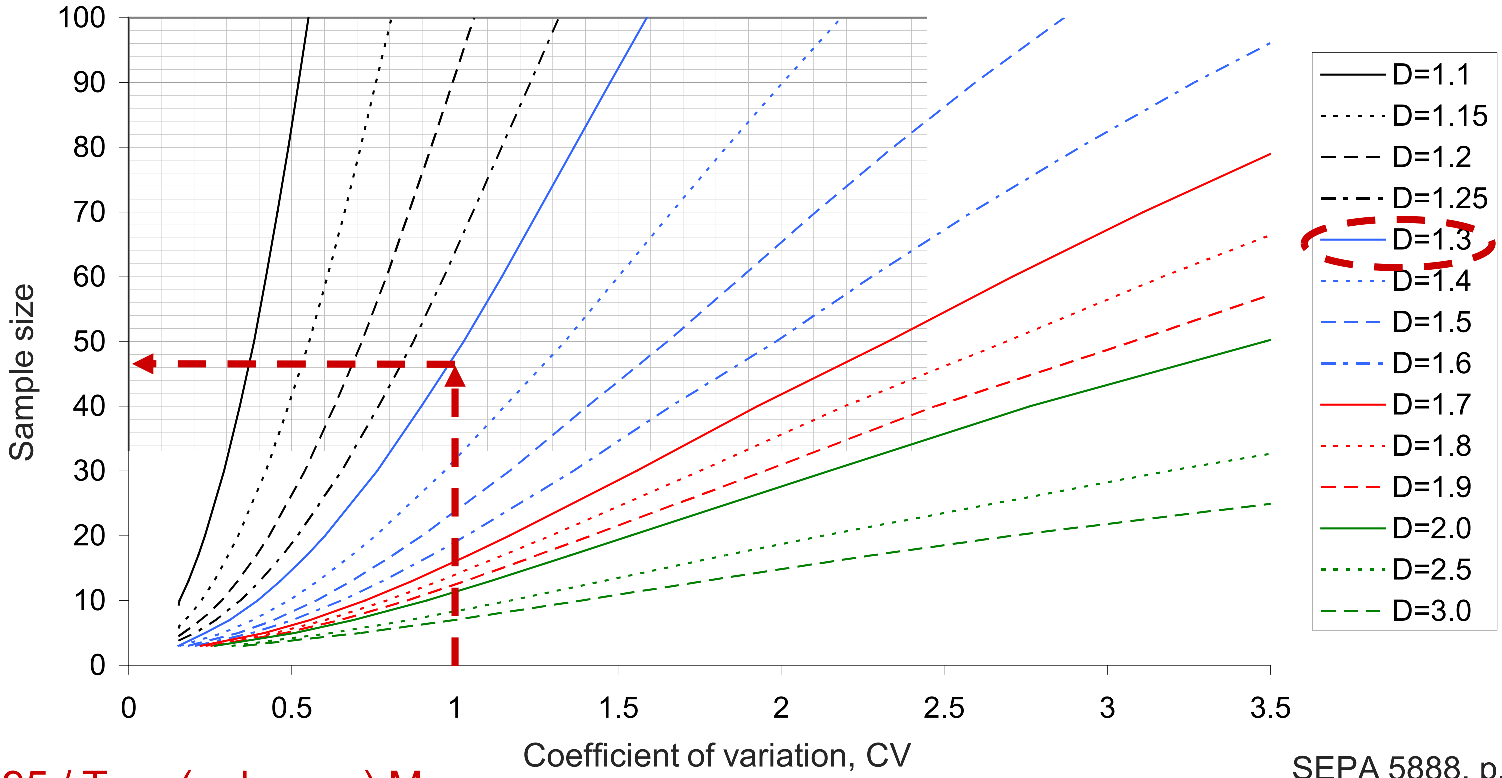
6. Choose sampling design and estimate the number of samples required based on the chosen strategy



Simple random sampling | Stratified random sampling | Systematic sampling | Systematic random sampling

"Samples are like potato chips. You're never satisfied with just one. Every one you take makes you want more. And you're never sure you've had enough until you've had too many!" *J. C. Myers*

Sample size- lognormal population



SEPA 5888, p. 36

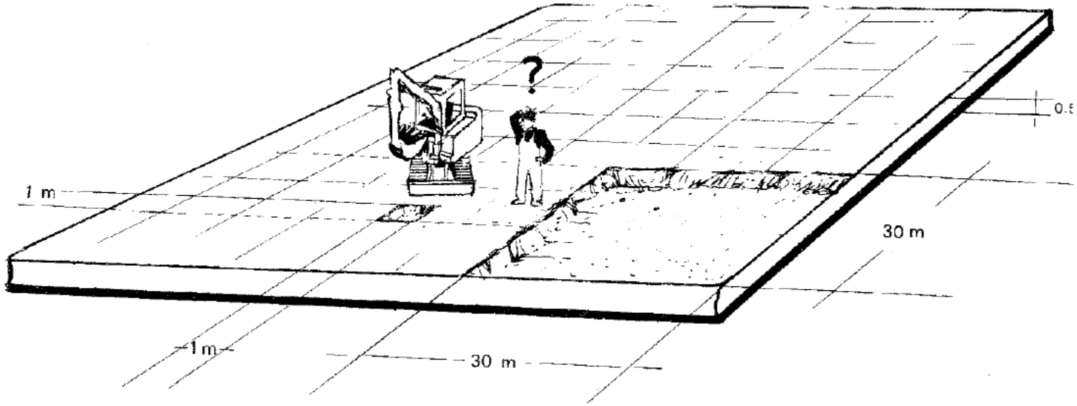
Here: D = UCLM95 / True (unknown) Mean

Example: D =1.3 = up to 30% error accepted: if the true mean concentration is 100 mg/kg, we would accept a calculated UCLM of 130 mg/kg

Classifcation of soil *in situ*

Classification based on estimation of mean concentration

- To estimate the mean concentration as good as possible



Suggested methodology

- A number of steps with decisions that are needed
- Iterative – most probably need to go back at times
- Diagrams are developed as support

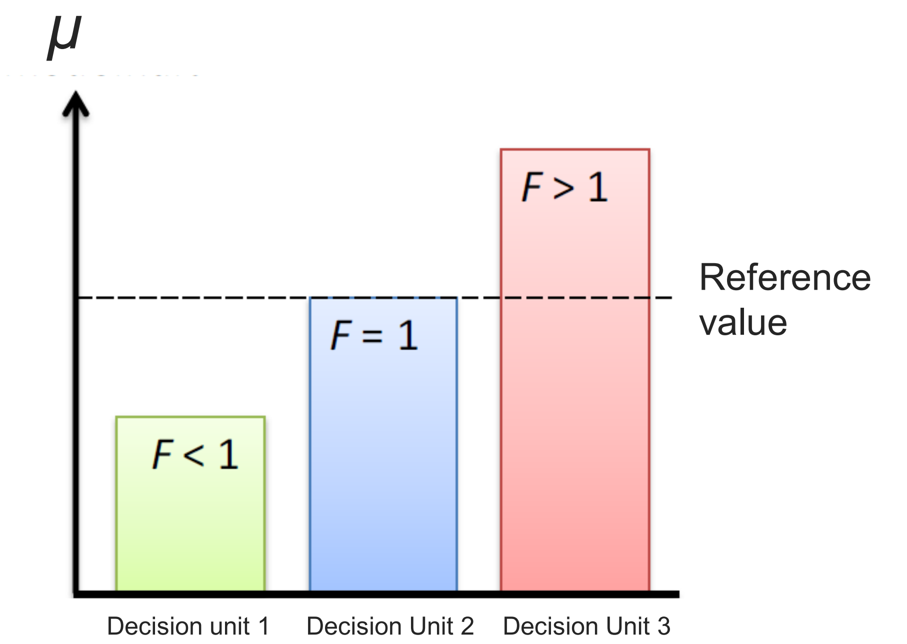
1. Define the purpose of the classification
2. Estimate heterogeneity, variability and contamination level (F)
3. Define the decision units
4. Choose the criteria for classification
5. Choose representative concentration
6. Choose sampling strategy
7. Control the sampling strategy

Level of contamination, F

- F is the ratio between the true (unknown) mean concentration (*μ*) and some reference value (*RV, GV, AL…*)

$$F = \frac{\text{mean concentration}}{\text{reference value}}$$

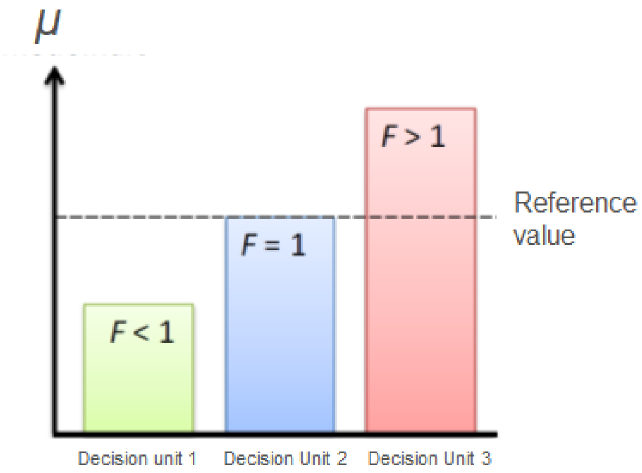
- *Compare with ”risk quotient”*



Why F?

- Useful!
- We never know the exact value of F since μ is always unknown
- But we know that if F is close to 1, it is very difficult to classify a decision unit correctly – but the consequences of type 1-error are low
- And, if F is much lower or higher than 1, classification is quite simple!
- When unknown, choose 1.5

$$F = \frac{\mu}{RV}$$

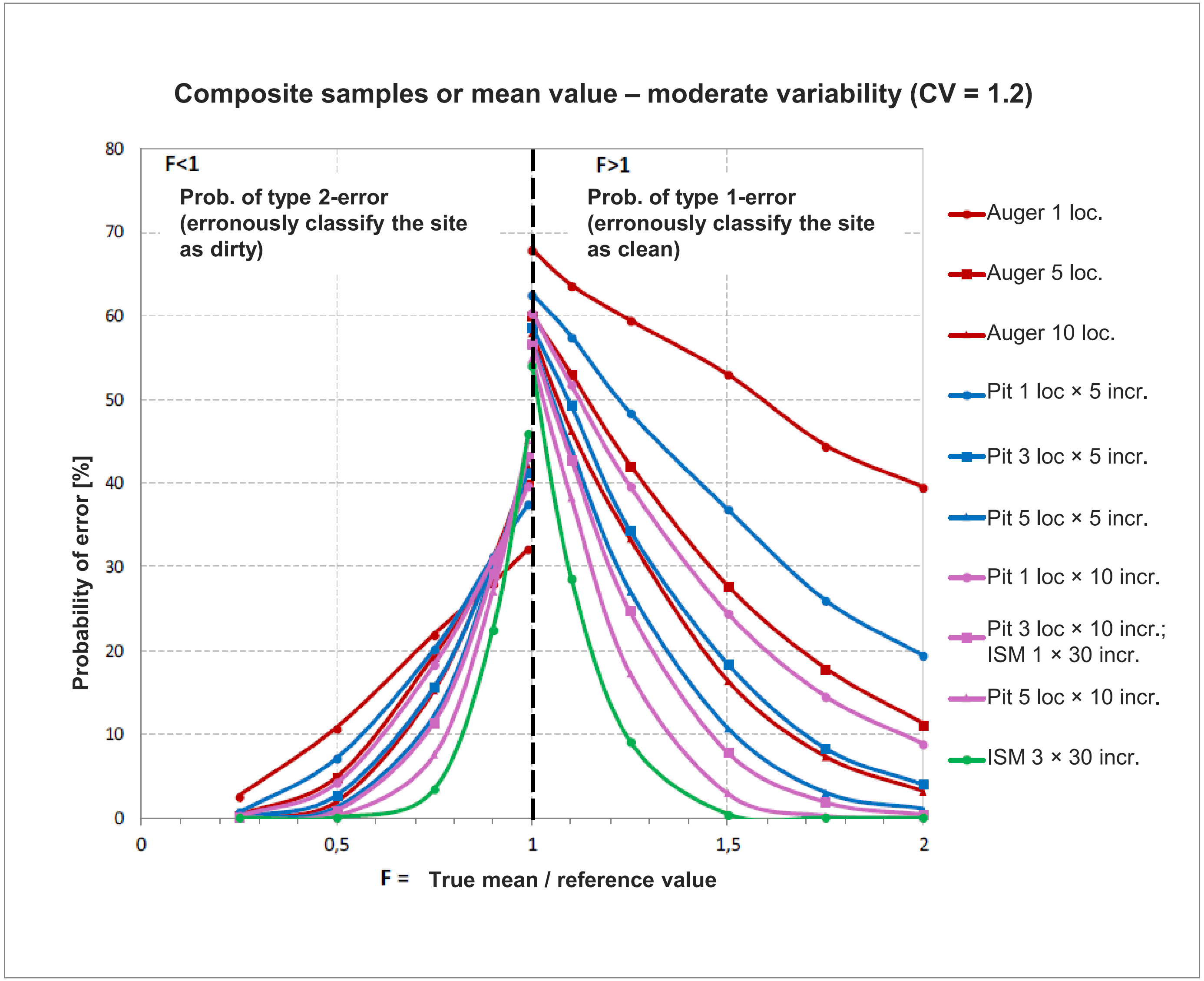


Ex 1. diagram

CV = 1,2 (auger scale)

Mean concentration is estimated from one composite sample or from a calculated mean value of several observations

The better the representativity, the better the symmetry!



Representative concentration

(to be compared to a reference value)

- A lab analysis value from a composite sample
- A calculated mean value based on several separate observations or composite samples

If type 1-errors and type 2-errors are equally bad

- A calculated UCLM value based on several separate observations or composite samples

If type 1-errors are worse

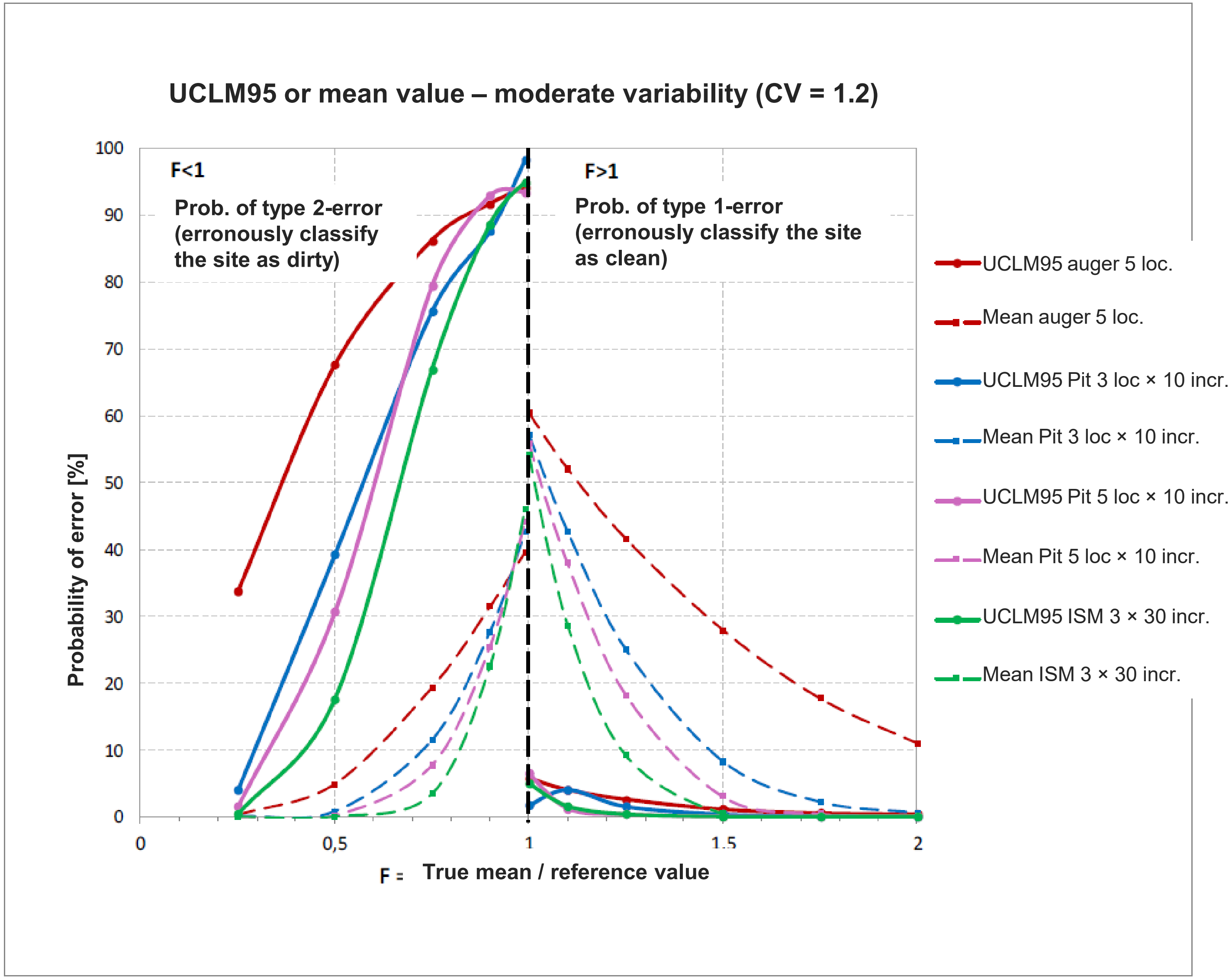
Ex 2. diagram:

CV = 1,2 (auger scale)

The mean concentration is estimated ”safely” by calculating UCLM95 based on several observations (solid lines)

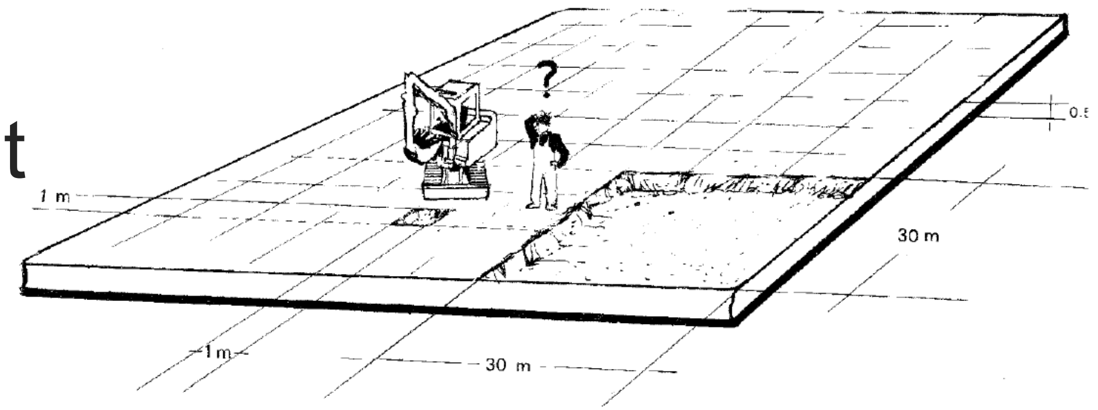
Comparison with just mean value (dotted lines)

UCLM95 protects against type 1-errors



Decision unit

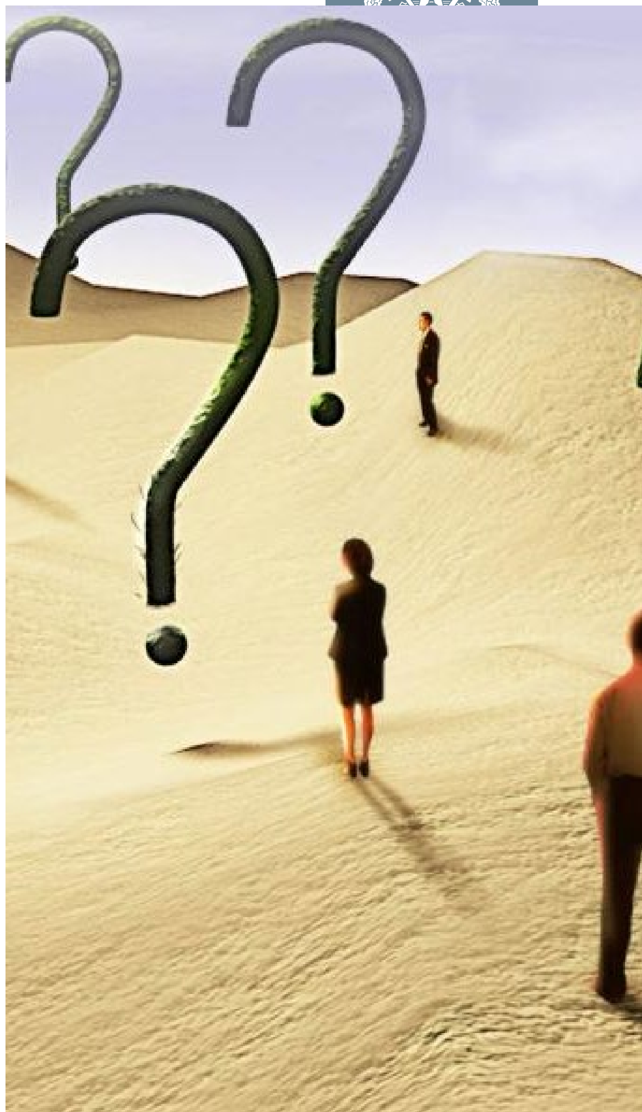
- The volume that should be classified as certain as possible (to avoid errors)
- ”*all soil with mean concentration of substance X above target level Y in Z m^3 should be removed*”
- Volyme Z is the decision unit



Size of the decision unit?

Needs to be defined for each site …

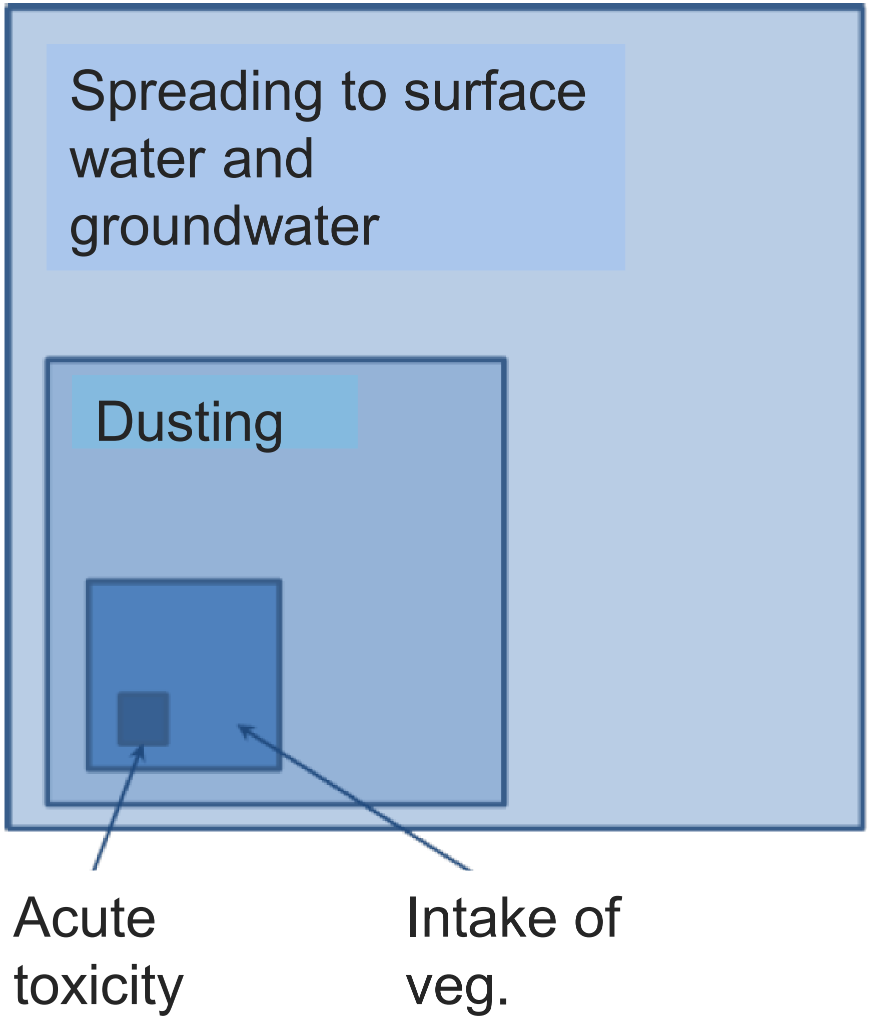
- Purpose (what to leave or classify waste)
- Contaminant: levels and variability
- Most important exposure pathways and risks
- Planned land-use
- Handling of masses, economy …Tricky! ☹



Exposure unit/area

- The smallest soil volume that is a problem from a risk perspective
- Contaminant variations within such unit is per defeinition not important!
- The exposure unit is typically much larger than the volume that separate observations represent

- *More research is needed!!*



Suggested approach for data evaluation

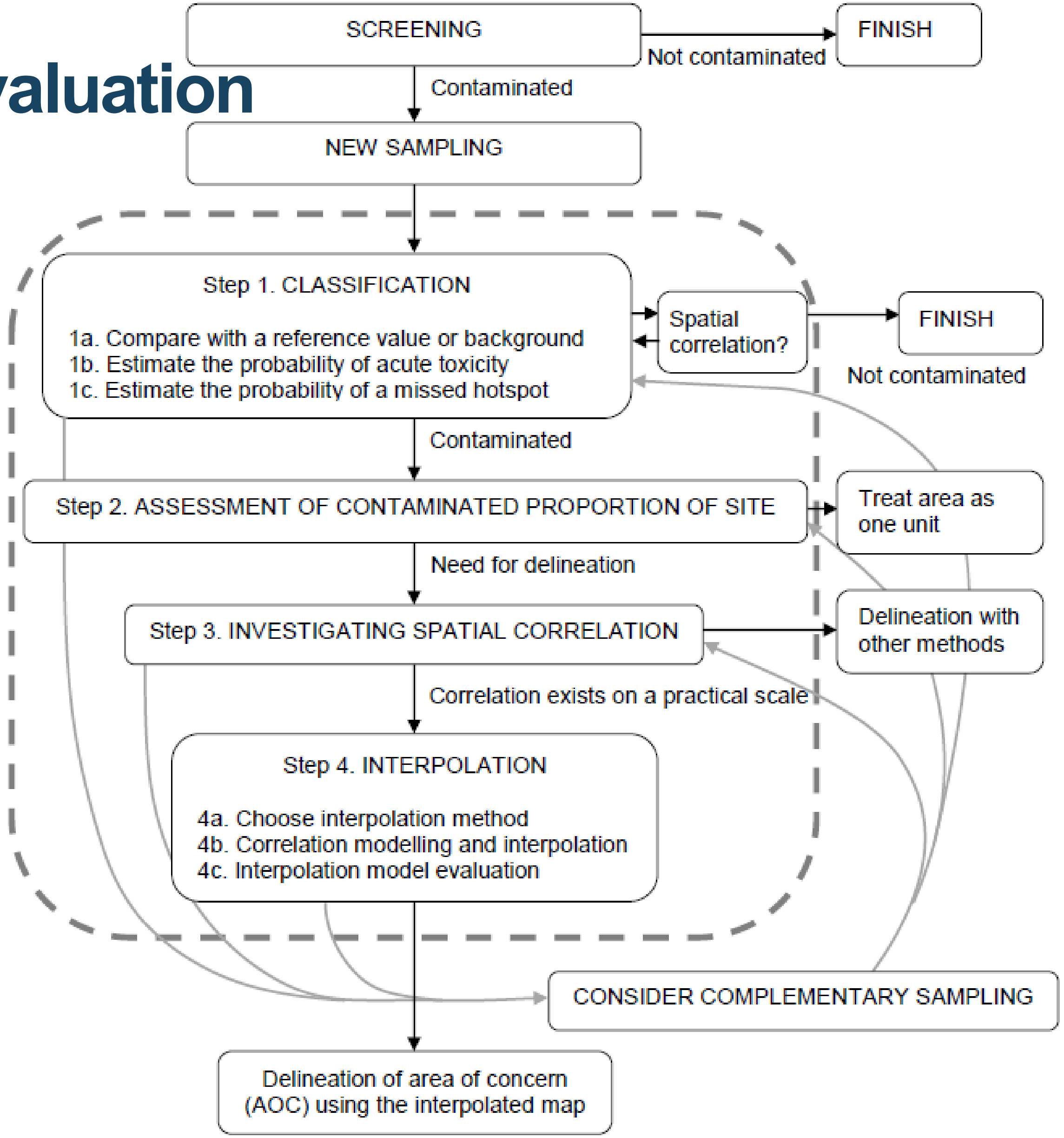
Step-wise procedure

1: Classification

2: Assessment of contaminated proportion of site

3: Spatial correlation?

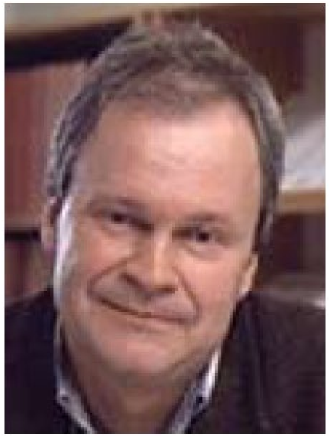
4: Interpolation, AOC



Acknowledgements to co-workers

- Pär-Erik Back, SGI

- Lars Rosén, Chalmers
- Tommy Norberg, previously Chalmers



AVANCEZ
1829
CHALMERS
UNIVERSITY OF TECHNOLOGY