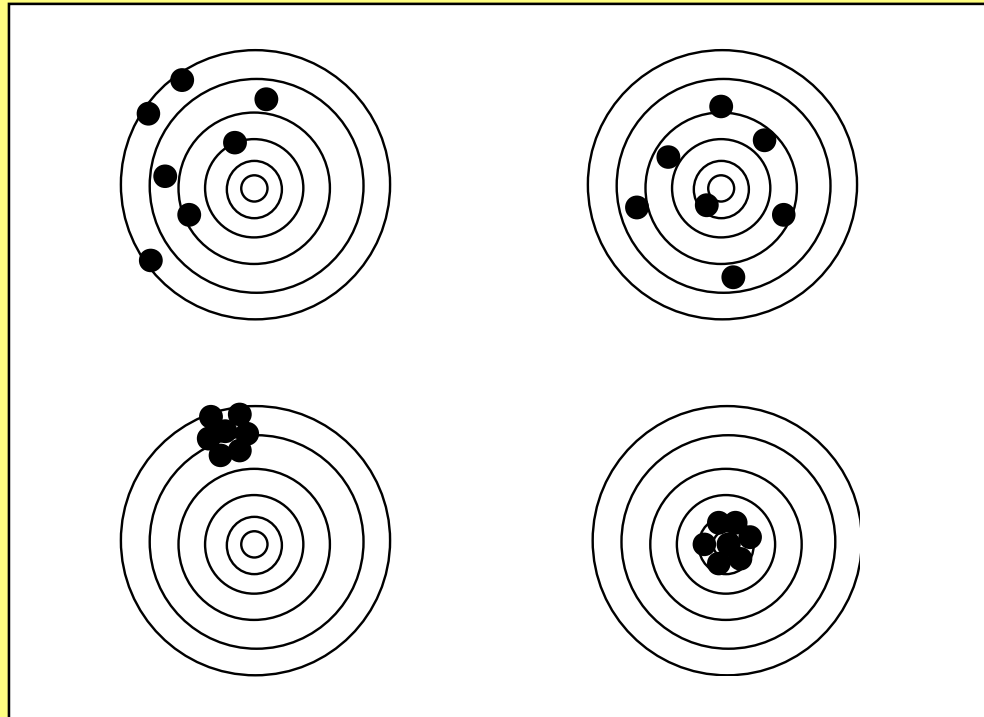




STATENS GEOTEKNISKA INSTITUT
SWEDISH GEOTECHNICAL INSTITUTE



Sampling Strategies and Data Worth Analysis for Contaminated Land

– A literature review

PÄR-ERIK BACK

Varia 500

LINKÖPING 2001

Varia Statens geotekniska institut (SGI)
581 93 Linköping

Beställning SGI
Litteratortjänsten
Tel: 013-20 18 04
Fax: 013-20 19 09
E-post: info@swedgeo.se
Internet: www.swedgeo.se

ISSN 1100-6692
ISRN SGI-VARIA--01/500--SE

Projektnummer SGI 10068
Dnr SGI 1-9912-718



Swedish Geotechnical institute, SGI
Chalmers university of technology, CTH

SGI Varia 500

Sampling strategies and data worth analysis for contaminated land

A literature review

Pär-Erik Back

FOREWORD

Contaminated land and groundwater is a problem of growing concern in our society. An increased environmental awareness, addressed in Sweden's new environmental legislation (*Miljöbalken*, 1999-01-01), has resulted in land and water contamination now being a major factor in land use planning and management, real estate assessment, and property selling. Investigation and remediation of contaminated areas are often associated with high costs. The Swedish EPA currently estimates that there are 22 000 contaminated sites in Sweden, of which approximately 4000 are in need for remediation.

The high costs and large number of contaminated sites are strong incentives for cost-efficient investigation and remediation strategies. *Miljöbalken* states that the environmental value of the remediation process must be higher than the investment costs. Critical issues to be addressed in order to meet the intentions of *Miljöbalken* and to provide cost-efficient handling of contaminated sites to landowners, operators, and the society are:

- What level of certainty is required to make sound decisions at a specific site, i.e. where, how and to what extent should sampling be performed?
- What remediation strategy is the most favorable with respect to investment costs and efficiency to meet specific clean-up goals?

Risk-based decision analysis is a theoretical approach to handle benefits, investment costs, and risk costs in a structured way to identify cost-efficient alternatives. Today, decisions regarding investigation strategies and remediation alternatives are in Sweden most often taken without completely or openly evaluating the cost-efficiency of alternatives.

The present report is one of two literature reviews prepared within the project *Risk-based decision analysis for investigations and remedial actions of contaminated land (Riskbaserad beslutsanalys för undersökningar och åtgärder vid mark- och grundvattenförorenade områden)*. The project is sponsored by the Swedish Geotechnical Institute (SGI) and carried out in co-operation between SGI and the Department of Geology at Chalmers University. The main purpose of the project is to evaluate and describe risk-based methods for cost-effective investigation and remediation strategies with respect to Swedish conditions. The two reports are:

- *Sampling strategies and data worth analysis for contaminated land*, by Pär-Erik Back (Department of Geology, Publ. B 486 and SGI, report Varia 500).
- *Decision analysis under uncertainty*, by Jenny Norrman (Department of Geology, Publ. B 485 and SGI, report Varia 501).

The main purpose of the reports is to provide comprehensive descriptions of the state of the art of decision analysis and sampling strategies for contaminated land. The reports form an important basis for future work, not only in the specific project, but also in the wide topic of risk-based decision analysis.

Göteborg 2000-12-20

Lars Rosén, Ph.D
Supervisor

ABSTRACT

A literature review of sampling strategies for contaminated soil and groundwater is presented. The different types of uncertainties associated with a site-investigation are reviewed. The uncertainties are classified as pre-sampling, sampling, and post-sampling uncertainties. Theories to quantify different types of uncertainties are reviewed, especially the different sampling uncertainties. The particulate sampling theory was found to be of interest because it identifies all the relevant sampling uncertainties and presents them in a structured way. The question on how many samples to collect under different conditions is addressed. A conclusion is that the sampling objective must be clearly stated prior to the development of any sampling plan. Two of the most common sampling objectives are to detect hot spots and to estimate mean concentrations.

Methods to determine the economic worth of sampling are reviewed. Traditionally, sampling has been performed with the objective to either (1) minimize uncertainty for a fixed sampling budget, or to (2) reach a prespecified accuracy at lowest possible cost. However, in recent years data worth analysis has been used increasingly. The valuable feature of data worth analysis is the ability to assess the worth of a proposed site-investigation program prior to performing any sampling or measurements. Data worth analysis is performed in a risk-based decision framework by combining decision theory with e.g. geological, geochemical, hydrogeological, and economic information. A conclusion is that the use of data worth analysis for contaminated land problems has a potential but the complexity of existing methods has limited its use. To be used on a more regular basis, data worth methods need to be further developed.

A review of more than 30 related software packages is presented in an appendix. It was found that public domain or low cost commercial software packages exist for a number applications related to site-investigations, e.g. sampling design, data evaluation, geostatistical techniques, stochastic simulation, data worth analysis, and decision analysis.

Keywords: sampling, uncertainty, risk, data worth, contamination, soil, groundwater

CONTENTS

FOREWORD

ABSTRACT

1	INTRODUCTION	6
1.1	BACKGROUND.....	6
1.2	THE PROBLEM: DECISION UNDER UNCERTAINTY.....	7
1.3	PURPOSE.....	8
1.4	LIMITATIONS.....	8
2	UNCERTAINTY IN GENERAL	9
2.1	QUANTITIES, PARAMETERS, CONSTANTS AND VARIABLES.....	9
2.2	ERROR VS. UNCERTAINTY, PROBABILITY AND RISK.....	10
2.3	TYPES AND SOURCES OF UNCERTAINTY.....	10
2.4	QUANTIFICATION OF UNCERTAINTY.....	14
2.4.1	<i>Classical statistics</i>	14
2.4.2	<i>Bayesian statistics</i>	15
2.4.3	<i>Spatial statistics</i>	17
2.5	MODELLING UNCERTAINTY.....	20
2.5.1	<i>The deterministic approach</i>	20
2.5.2	<i>The probabilistic approach</i>	20
2.5.3	<i>The possibility approach</i>	21
3	PRE-SAMPLING UNCERTAINTY	22
3.1	INTRODUCTION.....	22
3.2	UNCERTAINTY IN Á-PRIORI KNOWLEDGE.....	22
3.3	UNCERTAINTY IN CONCEPTUAL MODELS.....	22
3.4	COMBINING HARD AND SOFT INFORMATION.....	24
3.5	UNCERTAINTY IN QUANTITATIVE MODELS.....	26
4	SAMPLING UNCERTAINTY	28
4.1	INTRODUCTION.....	28
4.2	SAMPLING KEY TERMS.....	28
4.3	SAMPLING OBJECTIVES.....	29
4.4	PLANNING PHASE (DQOs AND QA/QC).....	30
4.5	SAMPLING PATTERN DESIGN.....	31
4.6	SAMPLING THEORIES.....	35
4.7	PARTICULATE SAMPLING THEORY.....	35
4.7.1	<i>Introduction</i>	35
4.7.2	<i>Sampling errors</i>	36
4.7.3	<i>Heterogeneity</i>	40
4.7.4	<i>Representativeness</i>	41
4.7.5	<i>Sample Support</i>	41
4.7.6	<i>Dimension of lots</i>	43
4.7.7	<i>Application to environmental problems</i>	43
4.8	COMPOSITE SAMPLING.....	43
4.9	MEAN CONCENTRATION – CLASSICAL APPROACHES.....	44
4.9.1	<i>Random sampling uncertainty</i>	44
4.9.2	<i>Systematic sampling uncertainty</i>	45
4.9.3	<i>Geochemical variability (population variability)</i>	47
4.9.4	<i>Models of total random uncertainty</i>	47
4.9.5	<i>Models of total systematic uncertainty</i>	50
4.9.6	<i>Required number of samples</i>	50
4.10	DETECTION OF HOT SPOTS.....	52
4.11	DELINEATION OF CONTAMINATED AREAS.....	54

5	POST-SAMPLING UNCERTAINTY	55
5.1	ERRONEOUS SAMPLES	55
5.2	UNCERTAINTY IN LABORATORY ANALYSES PLANS	55
5.3	FIELD AND LABORATORY ANALYSES	56
5.4	DATA EVALUATION	57
6	DATA WORTH ANALYSIS AND COST EFFECTIVENESS	59
6.1	INTRODUCTION.....	59
6.2	DECISION THEORY	59
6.3	TOOLS FOR DATA WORTH ANALYSIS.....	60
6.3.1	<i>Payoff tables</i>	60
6.3.2	<i>Decision trees and event trees</i>	60
6.3.3	<i>Influence diagrams</i>	61
6.3.4	<i>Expert systems</i>	62
6.4	STRATEGIES FOR DATA WORTH ANALYSIS.....	62
6.4.1	<i>Traditional approach: Fixed cost or fixed uncertainty</i>	62
6.4.2	<i>Misclassification cost and loss functions</i>	63
6.4.3	<i>Uncertainty-based stopping rules</i>	66
6.4.4	<i>Bayesian decision analysis</i>	66
6.4.5	<i>Optimisation approaches</i>	72
6.4.6	<i>Rule-based adaptive sampling</i>	72
6.4.7	<i>Sampling for model discrimination</i>	74
7	DISCUSSION	75
7.1	GENERAL.....	75
7.2	PRIOR INFORMATION AND EXPERT KNOWLEDGE.....	75
7.3	MODELS	75
7.4	PARTICULATE SAMPLING THEORY	75
7.5	SAMPLING ERRORS	76
7.6	SAMPLING STRATEGIES	76
7.7	DATA WORTH.....	76
	REFERENCES.....	78
	APPENDIX: SOFTWARE PACKAGES	

1 INTRODUCTION

1.1 Background

Contaminated soil and groundwater is a problem that has received increased attention in the last decade. The risk of exposure to contaminants for humans and the environment makes it necessary to investigate the degree of contamination at a site. Usually the investigation of a contaminated site results in a risk assessment, where the risk for humans and the environment is assessed. Sometimes, a preliminary assessment of the site is performed before any samples or measurements have been taken. This preliminary assessment is based on á-priori knowledge about the site, such as the type of contaminants that has been used at the site, the geologic conditions, spill events etc.

When a site with contaminated soil or groundwater is investigated, a number of decisions have to be made. They are of different nature but typically include which sampling strategy to use (how many samples to take, where to take them, which media to sample, control and duplicate samples etc.) and if remediation of the site is necessary. Usually, many of these, and similar questions, have no easy answer, mainly due to uncertainties in the site-investigation.

All stages of a site-investigation involve uncertainties. These include uncertainties in:

- á-priori knowledge about the site,
- conceptual models of the site,
- sampling, handling, field measurements and laboratory analyses,
- data handling and prediction of contaminant transport, and
- risk assessment

This can be summarised as uncertainties in (1) contaminant characterisation and (2) geological characterisation. The site-investigation results in an estimate of the extent of the contamination at the site. The level of contamination is uncertain but will influence the remediation costs. If there is a large uncertainty in the estimate of contamination, the remediation costs will also be uncertain. Often, the remediation cost will be much higher than anticipated. Therefore, it would be preferable if the uncertainties in a site-investigation could be estimated.

Because of stiff competition on the market the consultant with the cheapest site-investigation (smallest number of samples) often gets the job (Bosman, 1993). A consequence is that it will be difficult to discriminate between “nothing found” because there was nothing there or “nothing found” because of a poor site-investigation (Bosman, 1993). The latter may well be regarded as a success by the involved parties but may in fact lead to long term human health or environmental effects. It is obvious that the uncertainties are large in many site-investigations of contaminated soil and groundwater. Often, the overall uncertainty is underestimated, especially in cases where only a few field samples have been collected, and data analysis cannot recover more information than the samples contain (Flatman and Yfantis, 1996). Collecting only a few samples can therefore result in poor site characterisation, which in turn can result in poor and expensive remediation decisions (James and Gorelick, 1994).

Today, laboratory methodology has reached a point where analytical error contributes only a very small portion of the total variance seen in data (Mason, 1992; Shefsky, 1997). Typically, errors in the taking of field samples are much greater than preparation,

handling, analytical, and data analysis errors (van Ee et al., 1990). Unfortunately, many decisions are made in ignorance or contempt of the uncertainty of the sample data (Taylor, 1996). However, neglecting uncertainties does not mean that they do not exist (Lacasse and Nadim, 1996). Instead, uncertainties need to be considered and if possible be reduced.

There are several ways to reduce the overall uncertainty, but the most effective one is often to increase the number of samples taken, provided that the samples are representative of the sample location. This will usually reduce the uncertainty but increase the costs of sampling and analyses. Also, in many cases the uncertainty can be economically reduced by going from a random to a spatial variable sampling design (Flatman and Yfantis, 1996).

1.2 The problem: Decision under uncertainty

The main problems that will be addressed in this report is:

1. how to quantify the uncertainties that exist in a site-investigation, and
2. how to find an “optimal” level of uncertainty (i.e. optimal to make the proper decision), with respect to sampling.

These problems are a part of the main problem at a contaminated site, namely how to make *decisions under uncertainty*. Therefore, these questions should be handled in a framework based on *decision analysis*. The principle of decision analysis is how to make well-founded, defensible decisions under uncertainty. Typically, a reduction in uncertainty has a cost, which can be weighed against the benefit of making a more well-founded decision. It is practical to use monetary terms to quantify the costs and benefits, as described in chapter 6, but there are several other methods.

The optimal level of uncertainty in a site-investigation is reached when the cost for additional information (from additional sampling etc.) is equal to the expected benefits associated with the new information. Additional information is only cost-efficient up to that point. If the cost for additional information exceeds the benefit, sampling is no longer cost-efficient. *Data worth analysis* is thus the key to a cost-efficient site-investigation. In a decision analysis framework the data worth analysis sets the stopping rule when no additional collection of information should be made. More information will have a cost but the decision will not be sufficiently more well-founded to justify that cost. Decision analysis can also be used to compare alternative data-collection programmes, such as different sampling strategies.

All relevant uncertainties in the pre-sampling, sampling and post-sampling phases of the investigation have to be handled to make the data worth analysis complete. Optimally, it should be possible to quantify all uncertainties involved. Taking all uncertainties into consideration will make it possible to demonstrate the overall uncertainty of a site-investigation, especially the uncertainty associated with the sampling process.

1.3 Purpose

The purpose of the literature review is to present the state-of-the-art regarding uncertainties and data worth analysis in site-investigations. The focus is on sampling and data worth analysis. Both contaminated soil and groundwater are addressed. The literature review covers several related topics, primarily the following:

- Uncertainty in general
- Uncertainty in geologic and conceptual models (e.g. geologic interpretation and description)
- Sampling uncertainty and sampling strategies
- Data worth analysis

A review of relevant software packages is also presented in an appendix.

1.4 Limitations

In the literature review, the focus of uncertainty in site-investigations is on geochemical properties, but geologic and hydraulic properties are also considered. The two latter are associated with the important concepts of *parameter uncertainty* and *model uncertainty*.

Methods to estimate the uncertainty in field and laboratory analyses are only addressed in short. These methods have been described thoroughly in the chemical analysis literature. Also, uncertainty in human health and ecological risk assessments is important to be aware of but is not addressed in the report. This question is more related to remediation decision problems.

In the literature there is an abundance of qualitative routines for sampling but they are of minor concern in this report. The reason is that they rarely support any quantitative estimation of uncertainty.

2 UNCERTAINTY IN GENERAL

2.1 Quantities, parameters, constants and variables

A quantity is something that can be quantified in some way. Parameter is a similar term (note that the term parameter has a more specific definition in statistics). Quantities and parameters can be classified into a number of groups. Morgan and Henrion (1990) distinguish at least 7 different types of quantity:

- ✓ Empirical quantities
- ✓ Defined constants
- ✓ Decision variables
- ✓ Value parameters
- ✓ Index variables
- ✓ Model domain parameters
- ✓ Outcome criteria

These quantities can be explained in the following way, primarily as described by Morgan and Henrion (1990):

Empirical quantities represent measurable properties of the real world, such as hydraulic conductivity. The uncertainty of an empirical quantity can be expressed in the form of a probability density function (PDF). Often, the empirical quantities constitute the majority of quantities in models and they are often uncertain. The commonly used term “parameter uncertainty” refers to uncertainty in empirical quantities.

Defined constants are by definition certain, such as the mathematical constant π or the number of days in December. Many physical constants, such as the gravitational constant, are actually empirical quantities but with only a small degree of uncertainty.

A *decision variable* (or control variable) is a quantity for which it is up to the risk analyst or the decision-maker to select a value. Examples of decision variables at a site-investigation are the number of sampling points and the sampling depth. A decision variable has no inherent uncertainty but the difficulty is to find its “best” value.

Value parameters represent values or preferences of the decision-maker. Examples of value parameters are the discount rate in cost-benefit analysis, parameters of risk tolerance or risk aversion, and “value of life”. It is debatable if value parameters can be treated as probabilistic. Usually it is a serious mistake to treat value parameters in the same way as empirical quantities. However, the difference between a value parameter and an empirical quantity is not always clear. It is also a matter of intent and perspective.

Index variables are used to identify a location in the spatial or temporal domain of a model. Examples include x and y co-ordinates in a 2-dimensional model. Index-variables are certain by definition.

Model domain parameters specify the domain of the modelled system, generally by specifying the range and increments for index variables. Domain parameters define the level of detail of a model, both spatially and temporally. Examples of domain parameters are grid spacing in a model, time increment in transient simulations etc. Usually,

there is uncertainty about the appropriate values for domain parameters but it is inappropriate and impractical to represent the uncertainty with a probability distribution. The choice of value is up to the modeller.

Outcome criteria are variables used to measure the desirability of possible outcomes of a model. Examples include the calculated measure of risk in a risk model. Outcome criteria will be deterministic or probabilistic depending on whether any of the input quantities on which they depend are probabilistic.

Gorelick et al. (1993) make a distinction between parameters and variables in groundwater problems. They use the term parameter for time-independent properties such as hydraulic conductivity, while the term variable is used for time-dependent indicators such as hydraulic head and contaminant concentration.

2.2 Error vs. uncertainty, probability and risk

It is important to notice the difference between terms like error and uncertainty, although they often are used as synonyms. Terms like uncertainty, reliability, confidence and risk are probability-related and refer to *á priori* conditions, i.e. before an event. Probability is related to a statistical confidence before an event (Myers, 1997). Thus, an estimate is only a rational “guess” of the outcome. Error, on the other hand, can only be measured *á posteriori*, i.e. after an event has occurred. It is not possible to know what the errors will be before the event has occurred. Error relates to a known outcome or value and is therefore a more concrete item than the probability-related terms (Myers, 1997). However, uncertainty may be present even after an event has occurred if the error is not completely known.

As the reader will soon be aware of the distinction between error and uncertainty has not been maintained in the rest of this report. The main reason is that the two concepts often are used more or less synonymous by many authors. Therefore, the concept chosen by the author has been kept. The mixed use of uncertainty and error in practical applications can be explained in the following way: Prior to an investigation it is usually known that activities, such as sampling, will result in some error but the magnitude of the error is uncertain. Therefore, trying to estimate the error prior to investigation is a way of handling the uncertainty. This explains why the two terms often are used as synonyms.

Risk is often defined as a combination of the probability of a harmful event to occur and the loss (consequences) of the outcome of this event. However, there are numerous other definitions, which makes it necessary to clearly define the term “risk” in each application it is used to avoid misinterpretation.

2.3 Types and sources of uncertainty

It is not an easy task to define what uncertainty really is. The variety of types and sources of uncertainty, along with the lack of agreed terminology, can generate considerable confusion (Morgan and Henrion, 1990). Rowe (1994) defines uncertainty as absence of information, information that may or may not be obtainable. Taylor (1993) defines uncertainty as a measure of the incompleteness of one’s knowledge or informa-

tion about a quantity whose true value could be established with a perfect measuring device. Though not easily defined, it is important to distinguish between the different types and sources of uncertainty. Morgan and Henrion (1990) argue that probability is an appropriate way to express some of these kinds of uncertainty, but not all of them. In their view, the uncertainties should be handled differently depending on what type of quantity they refer to (see section 2.1).

Rowe (1994) distinguishes four classes of uncertainty:

1. Temporal, i.e. uncertainty in future and past states
2. Structural, i.e. uncertainty due to complexity
3. Metrical, i.e. uncertainty in measurements
4. Translational, i.e. uncertainty in explaining uncertain results.

All four classes occur at the same time but for a specific problem one or more dominates.

For human health risk assessments Hoffman and Hammonds (1994) refer to the International Atomic Energy Agency, who distinguishes between Type A and Type B uncertainty. Uncertainty about a deterministic quantity with respect to the assessment end point is called Type B uncertainty (example: specific individual). When the assessment end point is a distribution of actual exposures or risks, the uncertainty is of Type A (example: a group of individuals). To separate these uncertainties in a model, it is necessary to perform stochastic simulation in two dimensions.

Lacasse and Nadim (1996) divide uncertainties associated with geotechnical problems into two categories:

1. aleatory (inherent or natural) uncertainties, i.e. uncertainty that cannot be reduced
2. epistemic (due to lack of knowledge) uncertainties, i.e. uncertainty that can be reduced

This grouping excludes human error, which would fall into a third category.

It is quite common to distinguish between uncertainty in quantity (*parameter uncertainty*) and uncertainty about model structure (*model uncertainty*). Morgan and Henrion (1990) classify uncertainty in empirical quantities (see above) in terms of the sources from which it can arise:

- Random error and statistical variation (precision)
- Systematic error and subjective judgement (bias)
- Linguistic imprecision
- Variability
- Inherent randomness
- Disagreement
- Approximations

Random error and statistical variation is the kind of uncertainty that has been studied the most. No measurement of an empirical quantity can be absolutely exact; there will always be some uncertainty. This is especially true in site-investigations of contaminated land. There are a variety of well-known statistical techniques for quantifying this uncertainty. Random error can be reduced by taking sufficient number of measurements (Morgan and Henrion, 1990). Random error is often expressed as *precision* (Figure 1).

Precision is defined as “[t]he closeness of agreement between independent test results obtained under stipulated conditions” (International Organization for Standardization, 1994). Another name for random uncertainty is *stochastic uncertainty* (U.S. EPA, 1997c).

Systematic error is defined as the difference between the true value of a quantity and the value to which the mean of the measurements converge as more measurements are taken. It cannot be reduced by more measurements. Systematic error often comes to dominate the overall error and it has been found that it is almost always underestimated. This should not be surprising, since systematic errors often are unknown at the time, which calls for subjective estimates of this error. Systematic error is often expressed as *bias* (Figure 1). A positive counterpart to bias has been presented by the International Organization for Standardization (1994) by inventing the term *trueness*. Trueness is defined as “[t]he closeness of agreement between the average value obtained from a large series of test results and an accepted reference value”. Another name for systematic uncertainty is *methodical uncertainty* (U.S. EPA, 1997c).

Together, random errors and systematic errors constitute the *accuracy* of a measurement. Accuracy is a measure of the closeness of a measurement to the true value (Gilbert, 1987), i.e. the absence of error. The International Organization for Standardization (1994) defines accuracy as “[t]he closeness of agreement between a test result and the accepted reference value” and includes trueness and precision. The definition of accuracy is controversial; several definitions exist and they are not in agreement with each other (Pitard, 1993). Pitard (1993) argues that it is incorrect to include the notion of precision in the definition of accuracy since accuracy is independent of precision. U.S. EPA has recommended eliminating the use of the term accuracy because of lack of standard to determine it (Mason, 1992).

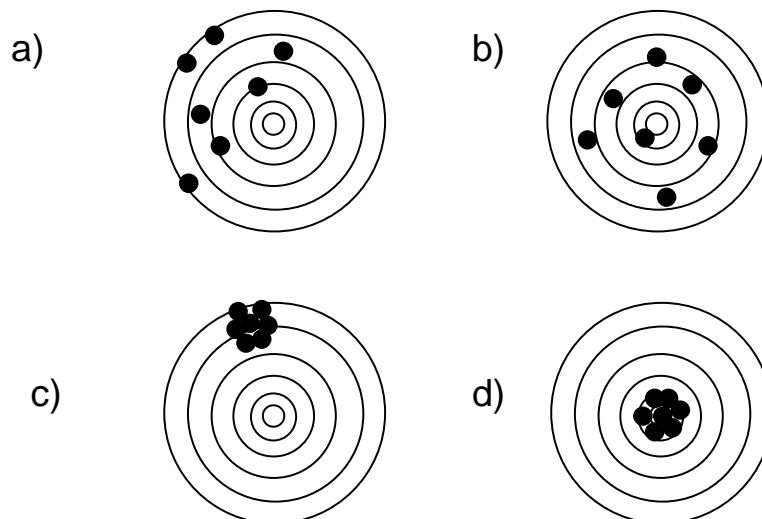


Figure 1. Patterns of shots at a target (after Gilbert, 1987).
 (a) high bias + low precision = low accuracy
 (b) low bias + low precision = low accuracy
 (c) high bias + high precision = low accuracy
 (d) low bias + high precision = high accuracy

Linguistic imprecision is best illustrated by an example. Consider a site where the level of the groundwater table should be determined. The statement “the groundwater level is low” is an example of linguistic imprecision. In a site-investigation linguistic imprecision should be minimised and preferably eliminated.

Variability is the uncertainty due to a quantity that varies over space or time. One example is when the groundwater level in a monitoring well varies over time. Another example is when the concentration of a contaminant varies over space at a contaminated site. Similar terms are *population variability*, *geochemical variability* or *environmental variability*. Taking more samples cannot reduce variability but our knowledge of the variability can be increased (Morgan and Henrion, 1990). Variability relates to random error and statistical variation in spatial statistical problems. Often, the real variation of contaminant concentration is expressed as a variance.

Inherent randomness is sometimes distinguished from other types of uncertainty. It contains randomness that is impossible to reduce by further investigation, in principle or in practice. An example is the inherent randomness in meteorological systems that make it impossible to correctly make long-range weather predictions (Morgan and Henrion, 1990).

Disagreement is a kind of uncertainty, especially in risk and policy analysis. A typical example is the disagreement among experts about a quantity. This type of uncertainty is important in situations where a decision must be made before further investigation about the quantity can be carried out. In risk-based decision analysis there may also be disagreement about how the failure criterion should be defined (see chapter 6) and about acceptable risks.

Approximation uncertainty arises because of the simplifications that are unavoidable when real-world situations are modelled. In many site-investigations the hydraulic conductivity is assumed to be constant in space, which is an approximation. It is often difficult to know how much uncertainty is introduced by a given approximation (Morgan and Henrion, 1990).

Usually, *uncertainty about model structure* is more important than uncertainty about the value of a parameter. In fact, the distinction between model uncertainty and parameter uncertainty can be rather slippery (Morgan and Henrion, 1990). Model uncertainty is due to idealisations made in the physical formulation of a problem (Lacasse and Nadim, 1996). Model uncertainty is further discussed in section 3.5. Sturk (1998) distinguishes between three types of parameter uncertainty; (1) statistical uncertainty, (2) measurement errors, and (3) gross errors.

In site-investigations of contaminated land Ramsey and Argyraki (1997) uses the term *measurement uncertainty* as a way of characterising uncertainty in site-investigations. In this concept field sampling and chemical analysis are just two parts of the same measurement process. Measurement uncertainty is the total uncertainty of the whole measurement process and it has four potential components:

1. Sampling precision (random error)
2. Sampling bias (systematic error)
3. Analytical precision (random error)
4. Analytical bias (systematic error)

Ramsey et al. (1995) applied this view when estimating the uncertainty of the mean concentration of lead and copper at a site, as described in section 4.9.4. Note that other authors may refer to measurement uncertainty as the analytical uncertainty, for example Taylor (1996).

Christian et al. (1994) have categorised uncertainty in soil properties according to Figure 2. These uncertainties can be compared to random error, systematic error, and variability as described above.

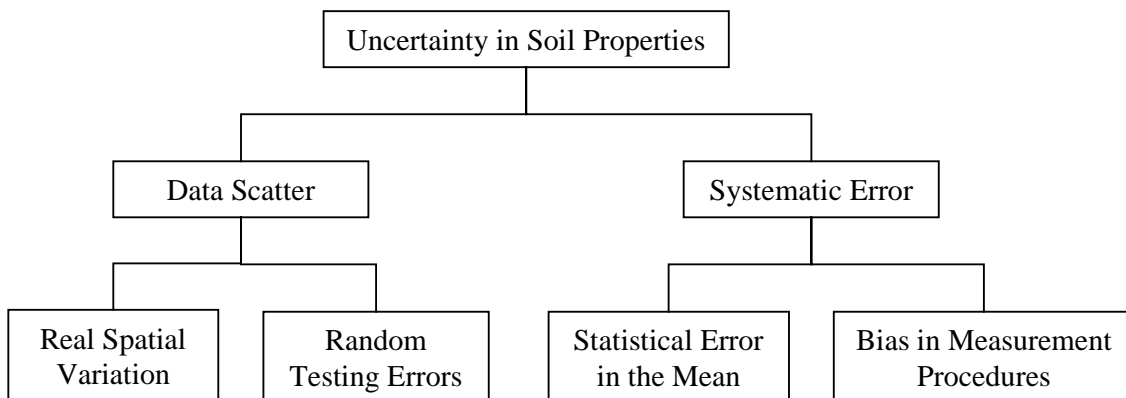


Figure 2. Categories of uncertainty in soil properties (after Christian et al., 1994).

Sturk (1998) classifies uncertainty in geological engineering problems into three classes; (1) inherent variability, (2) modelling uncertainty, and (3) parameter uncertainty. Inherent variability arises because the subsurface is composed of different materials. Modelling and parameter uncertainty is described in chapter 3.

2.4 Quantification of uncertainty

2.4.1 Classical statistics

In classical statistics a frequentistic view is applied. This implies that a probability distribution can only be created by the collection of data. If no data exists there will be no way of quantifying the uncertainty (the uncertainty is infinite), and it will make no sense defining a probability distribution. As more and more data are collected, the confidence in calculated distribution parameters increases. Two of the most important probability distributions are the *Normal* and the *Lognormal* distribution.

Many of the tools used in classical statistics assume normally distributed data, which makes calculations relatively simple compared to if this assumption is not employed. This can be a problem, since contaminant concentrations in soil are inherently lognormally distributed (Ball and Hahn, 1998). The methods of classical statistics will not be

discussed further since they can be found in any textbook on the subject. An excellent presentation of classical statistical methods for environmental sampling problems can be found in Gilbert (1987). U.S. EPA (2000b) describes a large number of statistical measures, plots and statistical tests for environmental sampling.

Thompson and Ramsey (1995) note that uncertainty can be expressed as a standard deviation (the standard uncertainty) or as a half-range (the expanded uncertainty). The latter is obtained by multiplying the standard uncertainty by a coverage factor, normally between 2 and 3.

If few data exist it may be problematic to use classical statistics to estimate the uncertainty.

Lacasse and Nadim (1996) mention that *short-cut estimates* can be used in these cases. This is a method to estimate the standard deviation for limited data sets of symmetrical data. Ball and Hahn (1998) also address the issue of small data sets, but for environmental problems, and a small literature review on the subject is presented. They conclude that the issue of small data sets is not addressed directly in current textbooks on statistics for environmental problems.

Myers (1997) concludes that care should be used when applying classical statistical models to spatial data since classical statistics assumes uncorrelated data while spatial data often is correlated.

2.4.2 Bayesian statistics

Bayes' theorem

The base for Bayesian statistics is Bayes' theorem. It is a logical extension of the basic rules of probability. Bayes' theorem can be formulated as a conditional probability (Alén, 1998; Vose, 1996):

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)}$$

where $P(A_i|B)$ represents the probability of event A_i given that event B has occurred. An advantage of Bayesian statistics compared to classical statistics is that all sources of information can be considered. This includes direct evidence from statistical sampling as well as indirect evidence of whatever kind available from past experience of the analyst or other experts (Hoffman and Kaplan, 1999).

Bayesian updating

In Bayesian statistics Bayes' theorem is employed to determine *posterior* probability distributions of a variable by combining *priori* information with observed data (Vose, 1996) (a more formally correct term for probability distributions is probability density functions, PDFs). In other words, the prior information is updated with additional data, e.g. from sampling, to reduce the uncertainty. Sometimes an analysis is performed when prior information is available but before additional information has been collected. This stage is called *preposterior* analysis (Freeze et al., 1992). To further reduce the uncertainty another sampling round can be performed. The posterior estimate now becomes

the prior estimate, which can be used together with the new data to produce another posterior estimate with less uncertainty. This step can be repeated several times. McLaughlin et al. (1993) calls this procedure “sequential updating”.

As formulated above, Bayes’ theorem is relatively simple but when applied to continuously distribution functions the mathematics can become laborious. Therefore, risk analysts appear to split into two camps: those who use Bayes’ theorem extensively and those who do not (Vose, 1996).

Prior PDFs

There is an important aspect of Bayesian statistics that often is used even if calculation with Bayes’ theorem is not performed. In contrary to classical statistics it is possible to define probability density functions (PDFs) although no “hard data” exists. Subjective PDFs can be used to characterise an individual’s belief about the value of a parameter (Hammit, 1995). Such *prior distributions* should reflect the prior knowledge before measurements have been made, something that is not possible in classical statistics. Therefore, the difference between classical and Bayesian statistics is more or less philosophical since the mathematics are the same. However, some authors, e.g. Rowe (1994), argue that probability distributions should only be used to address future temporal uncertainty since probability does not exist in the past, and that they are improperly used otherwise.

Prior distributions are defined at early stages in a project when measurements are scarce. If sufficient *hard data* is available it is relatively easy to define a probability distribution by classical statistical methods. *Hard data* are direct measurements or observations (Koltermann and Gorelick, 1996). If measurements or similar hard data is not available, so called *soft data* can be used to define a prior distribution. The soft data are indirect information such as historical information, expert opinions, professional experience and all other information that may be difficult to quantify. Freeze et al. (1992) and Johnson (1996) also refer to indirect measurements as soft data, such as geophysical measurements.

A distribution based on soft data is to some extent subjective or judgmental but should contain the uncertainty associated with the information. Judgmental approaches to generate probability distributions can be formal or informal in nature (Taylor, 1993). Formal methods are to be preferred. Several formal methods to assign subjective probabilities and construct subjective probability distributions exist. A review of these methods is given by Olsson (2000). Vose (1996) also discusses how distributions are defined from expert opinion and the sources of error in subjective estimation. Many procedures for combining information from multiple sources to construct prior distributions have been proposed but none are clearly superior (Hammit and Shlyakhter, 1999). It is beneficial if objective and subjective information can be considered together when distributions are developed (Taylor, 1993).

Examples of probability distributions can be found in many textbooks on probability theory. Typical examples of distributions constructed from soft data are the *non-parametric distributions*, which include the Cumulative, Discrete, Histogram, General, Triangular, and Uniform (Rectangular) distributions (Vose, 1996). Common *parametric distributions* include the Exponential, Normal (Gaussian) and Lognormal distributions.

Vose (1996) argues that parametric distributions based on subjective information should be used with caution.

The shape of the prior distribution is very important since it reflects the current information. However, there is a well-documented tendency of both experts and lay people to underestimate uncertainty in their knowledge of quantitative information (Hammitt, 1995; Hammitt and Shlyakhter, 1999). This will often result in a prior distribution that is too narrow (Taylor, 1993). It has been concluded that empirical distributions of measurement and forecast errors have much longer tails than can be described by a Normal distribution (Hammitt and Shlyakhter, 1999). Hammitt (1995) suggests that the tendency for overconfidence must be taken into account when the prior distribution is estimated. Some researchers recommend that the uncertainty should be increased deliberately to combat this bias (Taylor, 1993).

In Bayesian statistics all uncertain quantities are often assigned probability distributions. This makes analytical calculations troublesome, and often even impossible. Therefore, different types of techniques for propagation and simulation of uncertainty has been developed. Morgan and Henrion (1990) provide an introduction and review of the principal techniques. A common name for many of these techniques is *stochastic simulation* (Koltermann and Gorelick, 1996). Some of these include Monte Carlo simulation (U.S. EPA, 1997b; Vose, 1996), Monte Carlo simulation with Latin Hypercube sampling (Vose, 1996) and first-order second-moment (FOSM) (James and Oldenburg, 1997). The latter is less computationally intensive than Monte Carlo simulation. A simpler simulation technique that requires even less computational capacity is the Point Estimate Method (Alén, 1998; Harr, 1987).

Hoffman and Hammonds (1994) and Taylor (1993) argue that it under certain circumstances may be useful to separate uncertainty (due to incomplete knowledge) and variability (due to natural variation) when Monte Carlo simulation is performed. One such example can be simulations in human health risk assessments.

2.4.3 Spatial statistics

The most well known field of spatial statistics is geostatistics. Originally, geostatistics developed as a tool to estimate ore reserves. The geostatistical methods have been derived from existing classical statistical theory and methods (Myers, 1997). In the past 20 years, the geostatistical literature has grown enormously and many developments have been made (Flatman and Yfantis, 1996). Here, only a few fundamental concepts will be described briefly. An excellent introduction to geostatistics is given by Isaaks and Srivastava (1989). The basic concepts of geostatistics are described and provided with examples by Englund and Sparks (1991) as well as by Deutsch and Journel (1998). A more advanced presentation is given by for example Cressie (1993).

Variogram

The *variogram*, or *semivariogram*, is often used to analyse the spatial or temporal correlations (Figure 3). The computation, interpretation and modelling of variograms is the “heart” of a geostatistical study (Englund and Sparks, 1991). The horizontal axis of the variogram is called the *lag* axis and the vertical axis *gamma* axis (Flatman and Yfantis, 1996). The experimental points are computed by averaging data grouped in distance class intervals (lags on the horizontal axis). The variance is displayed on the vertical

axis. The result is a graph displaying variance as a function of separation distance between samples. The variance in Figure 3 is a measure of the structural variation (increasing variance with increase in separation distance between sample locations). A variogram model can be fitted to the experimental data. A recommendation is that a minimum of 30 data pairs should be used in each lag (Chang et al., 1998).

As illustrated in Figure 3 the rise in variance has an upper bound known as the *sill*. The value on the lag axis corresponding to the sill is called the *range* of correlation or correlation distance. When the separation distance is smaller than the range the variance will increase with the distance. In this case there is a spatial correlation between sampling points. For distances greater than the range the variance will be constant, i.e. there is no longer a correlation between points at this separation distance. The correlation distance at soil-contaminated sites is usually small because of heterogeneities and non-uniform contamination release.

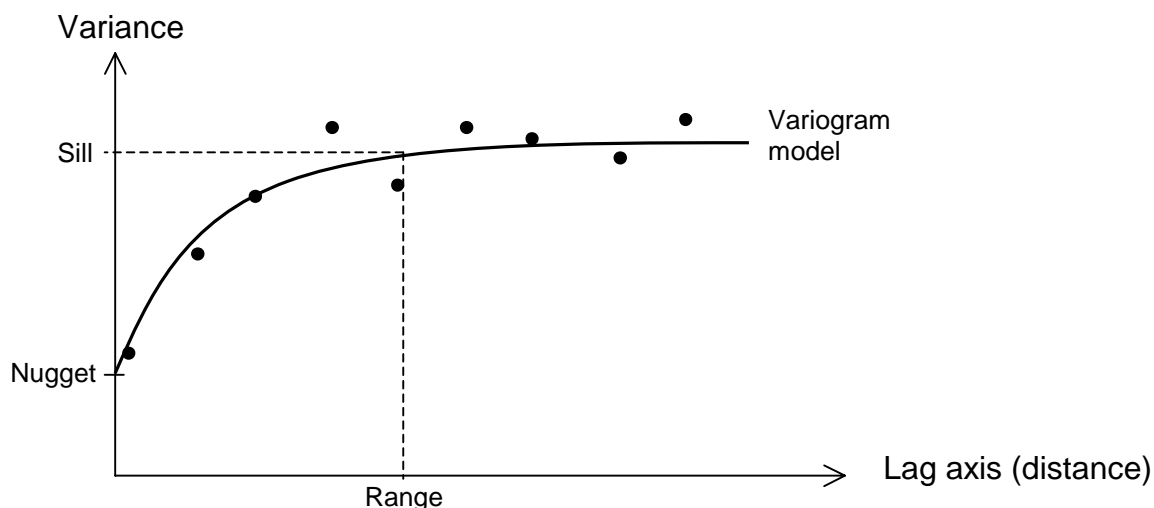


Figure 3. Example of a variogram with sill, range and nugget defined.

At a separation distance of zero one would assume that the variance also would be zero. However, in practice this is often not the case (Figure 3). The intersection of the variogram model on the gamma axis is called *nugget*. The nugget represents the experimental error and field variation within the minimum sampling spacing (Chang et al., 1998). It is constant for all lags (Flatman and Yfantis, 1996). In many cases the nugget effect is the result of sampling errors, i.e. the fundamental error (*FE*) and the grouping and segregation error (*GE*) (Myers, 1997), see section 4.7.2. The nugget effect is reduced if these errors are minimised.

Chang et al. (1998) cites Y.J. Chien et al. who suggest that the ratio of nugget variance to sill variance can be used as a criterion to classify the spatial dependence of soil properties:

- <25 % = strong spatial dependence
- 25-75 % = moderate spatial dependence
- >75 % = weak spatial dependence

Myers (1997) exemplifies how the variogram can be used to quantify the uncertainty in estimated concentration values at unsampled locations.

Kriging

Kriging is a linear-weighted average interpolation technique used in geostatistics to estimate unknown points or blocks from surrounding sample data. Traditionally, kriging has been used for mapping purposes. By using a spatial correlation function derived from the variogram the kriging algorithm computes the set of sample weights that minimise the interpolation error (Flatman and Yfantis, 1996). Several different types of kriging can be performed and the literature about kriging techniques is extensive. Deutsch and Journel (1998) shortly describe several kriging techniques, such as simple kriging, ordinary kriging, kriging with a trend model, kriging the trend, kriging with an external drift, factorial kriging, cokriging, nonlinear kriging, indicator kriging, indicator cokriging, probability kriging, soft kriging by the Markov-Bayes model, block kriging, and cross validation. Freeze et al. (1990) distinguish between three types of kriging; (1) *simple kriging* in which the mean is stationary and known, (2) ordinary kriging in which the mean is stationary but unknown, and (3) universal kriging in which the mean may inhibit a drift or trend.

As a mapping tool, kriging is not significantly better than other interpolation techniques that can account for anisotropy, data clustering etc. (Deutsch and Journel, 1998). However, in contrary to other techniques, the kriging algorithm provides an error variance. Unfortunately this variance cannot generally be used as a measure of estimation accuracy.

Geostatistics can be used for estimation of *local uncertainty*, such as uncertainty regarding delineation of contaminated areas where remedial measures should be taken. Goovaerts (1997) presents two models of local uncertainty; the multi-Gaussian and the indicator-based algorithms. These tools can be used to classify test locations as “clean” or “contaminated”. The local uncertainty approach may not be appropriate for certain applications. For example, the probability of occurrence of a string of low or high values requires modelling of spatial uncertainty (Goovaerts, 1997). This may require simulation.

Conditional simulation

In recent years the application of kriging has shifted from mapping to conditional simulation, also called *stochastic imaging*. Conditional simulation enables modelling of spatial uncertainty. The theory behind a number of such simulation techniques is presented by for example Cressie (1993), Goovaerts (1997), and Deutsch and Journel (1998).

Other geostatistical techniques

Numerous additional geostatistical considerations affect uncertainty in environmental applications, for example anisotropy, spatial drift or trend, multivariate analysis, mixed or overlapping populations, concentration-dependant variances, and specification of confidence limits (Flatman and Yfantis, 1996). There are also methods for correction of sample data from badly designed sampling plans, such as preferential sampling (judgement sampling), where sample locations are more clustered in certain areas. Techniques to correct for this include *polygonal declustering* and *cell declustering* (Goovaerts, 1997; Isaaks and Srivastava, 1989).

Markov chains

In addition to geostatistics, Markov chains can be used as a spatial statistic tool. It is a non-parametric technique that can be used for spatial probability estimation. The properties at unknown points are estimated from properties of surrounding known points. The property of interest is divided into a number of states. The probability for a specific state can be calculated at unknown points. The methodology can be combined with Bayesian statistics, as demonstrated by for example Rosén (1995) and Rosén and Gustafson (1996).

2.5 Modelling uncertainty

2.5.1 The deterministic approach

Sturk (1998) points out that in the traditional deterministic approach a lot of trust is put in engineering judgement for the assessment of uncertainty. Often, a hedge against uncertainty is built-in in predictions and deterministic models. In the field of geotechnics this is performed by using safety factors, which may result in building things too strong (Alén, 1998).

A common way to evaluate the uncertainty associated with deterministic models is by *sensitivity analysis*. The values of the input parameters are varied in a systematic way and the change in model result is studied. A review of these techniques is beyond the scope of this report.

2.5.2 The probabilistic approach

Probability is often used as the measure of uncertain belief (Morgan and Henrion, 1990). The probabilistic approach is getting more and more common in risk assessment of contaminated land. For hydrogeological decision analysis, Freeze et al. (1990) provide a detailed discussion of how uncertainty is handled in a probabilistic approach. The uncertainty in empirical quantities is expressed by means of stochastic variables. The stochastic variables are assigned as probability density functions (PDFs) that describe the uncertainty of the quantity. The problem then becomes how to select the appropriate PDF for a stochastic variable. The probabilistic approach is often based on Bayesian statistics (section 2.4.2).

For spatial contamination problems, the heterogeneity is described by *stochastic process theory*. Important keywords are e.g. autocorrelation function, correlation length, stationarity, and ergodicity. The acceptance of the ergodic hypothesis is a required step if stochastic process theory is to be used (Freeze et al., 1990). Stochastic process theory is complex and a description is outside the scope of this report. Presentations can be found in e.g. Freeze et al. (1990) and Gorelick et al. (1993).

An important difference between a probabilistic model and a deterministic algorithm is that the statistic model provides an error of estimation, which the deterministic approach does not (Myers, 1997).

2.5.3 The possibility approach

The possibility approach, also called the fuzzy approach (Guyonnet et al., 1999), is based on fuzzy set theory and fuzzy logic, which are a generalisations of classical set theory and Boolean logic (Mohamed and Côte, 1999). In fuzzy set theory the parameter uncertainty is incorporated by assigned a degree of membership to each element of a set. A specific type of fuzzy set is the fuzzy numbers. An example of a fuzzy number is (Mohamed and Côte, 1999):

approximately $5 = \{2/0, 3/0.33, 4/0.67, 5/1, 6/0.5, 7/0\}$

The numbers on the right hand side of the slashes indicate the degree of membership to the set, where 1 indicates total membership and 0 non-membership (Mohamed and Côte, 1999). Fuzzy numbers can also be illustrated graphically as symmetric or asymmetric triangular or trapezoidal possibility functions similar to, but not equal to, probability functions. Calculations like addition, subtraction, multiplication and division can be performed with the possibility functions.

Fuzzy set theory has been used to capture the uncertainty in transport parameters, although its use has been restricted to analytical solutions or simple 2-d and steady state problems (James and Oldenburg, 1997). Mohamed and Côte (1999) developed a risk assessment model for human health based on fuzzy set theory. The model includes four transport models; (1) groundwater transport, (2) run-off erosion for contaminated soils, (3) soil-air diffusion and air dispersion, and (4) sediment diffusion-resuspension.

Guyonnet et al. (1999) compared the possibility approach to the probabilistic approach for addressing the uncertainty in risk assessments. A conclusion was that the possibility approach is of more conservative nature than the probabilistic approach. Guyonnet et al. (1999) argue that the possibility approach may be better in some circumstances, since PDFs are difficult to develop when data are sparse, and environmental hazards are often perceived by the general public in terms of possibilities rather than probabilities.

Kumar et al. (2000) used fuzzy set theory in combination with neural networks (see section 6.3.4) for subsurface soil geology interpolation. The possibility of occurrence of different geologic classes was calculated.

3 PRE-SAMPLING UNCERTAINTY

3.1 Introduction

Pre-sampling uncertainty is considered to be uncertainty that evolves before any sampling takes place. Uncertainty is introduced by á-priori knowledge, during development of conceptual models, when the conceptual models are transformed into some kind of quantitative or predictive model, and when the quantitative model are used for prediction. It is important to note that the model used for prediction does not always have to be very complicated. However, it is important to consider that even if very simple techniques are used for predicting the future, the predictions are still based on a model (the person making the prediction may not even be aware of this fact).

Model uncertainty is included in this chapter because models are often used prior to sampling when data are scarce. However, they are also used in post-sampling activities. Therefore, the section on model uncertainty in this chapter could as well have been placed in chapter 5.

3.2 Uncertainty in á-priori knowledge

Á-priori knowledge is information that is known prior to the field investigation, i.e. before any samples have been taken. This information may include both “hard” and “soft” data (see section 2.4.2). At the early stages of a project information is sparse and the soft data may be the most important information.

To make use of the soft information it is often necessary to quantify it in some way. A common method to make quantitative use of prior soft information is to use it for estimating the value of some kind of parameter, for example the thickness of the soil layer at a contaminated site. In this process it is inevitable that the estimated parameter will be associated with some uncertainty. The higher quality of the prior information, the less uncertain will the parameter value be (provided that the person performing the estimating has the professional skill). Koltermann and Gorelick (1996) put it this way: “...*an estimation of any parameter will always carry uncertainty with it, and the estimates of uncertainty will themselves be uncertain*”. This expression pinpoints the important concept of *parameter uncertainty* mentioned in section 2.3.

A common way of including á-priori knowledge is to define prior distributions according to section 2.4.2. A less stringent methodology is presented by Bosman (1993), who suggests defining *guess-fields* of estimated contaminant concentrations. The guess-field should be based on historical information but the reliability of such information is often poor (Bosman, 1993). Therefore, an estimate of the reliability of each guess-field point must be given. Ferguson (1993) recommended a similar approach, which Ferguson and Abbachi (1993) used for hot spot detection, as described in section 4.10.

3.3 Uncertainty in conceptual models

Gustafson and Olsson (1993) define a model as an application of a theory to a specific problem (Figure 4). A model can be conceptual or mathematical. In this section uncer-

tainty in conceptual models will be described, while mathematical models are discussed in section 3.5.

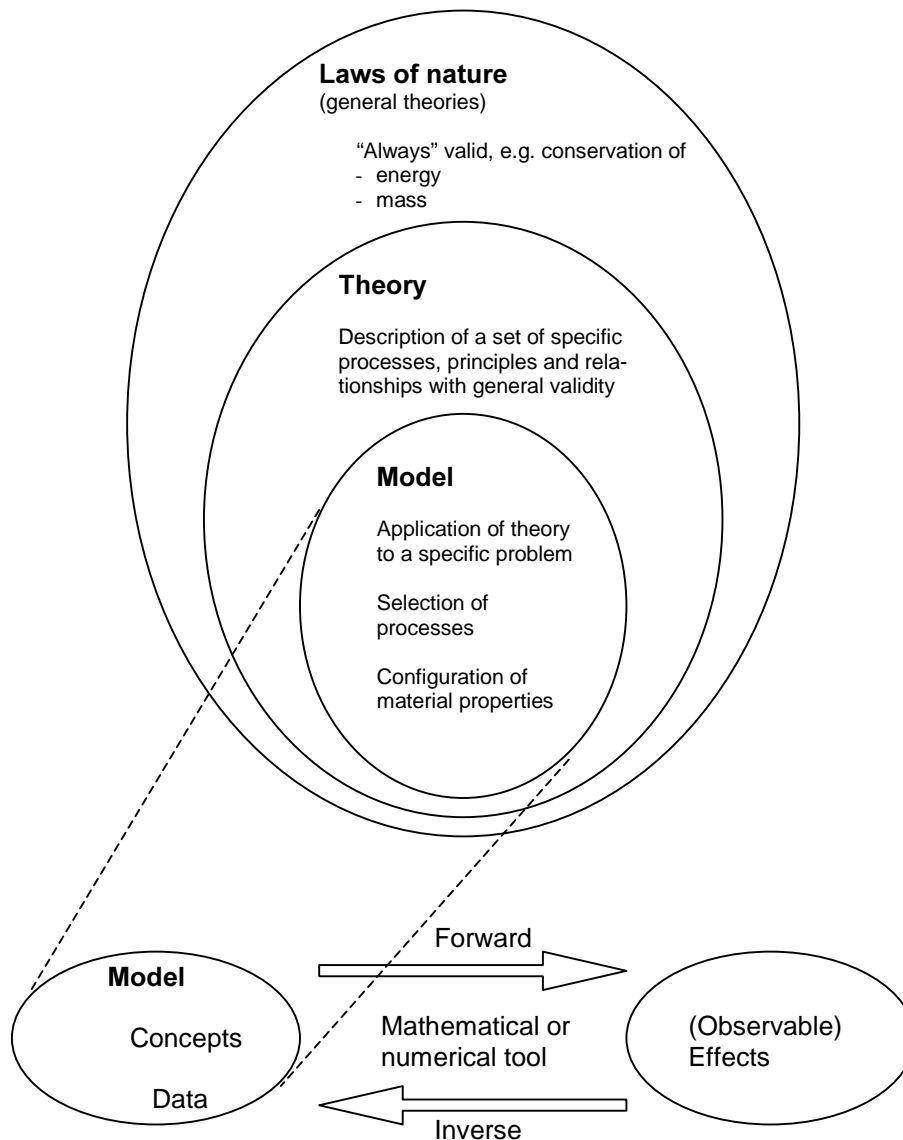


Figure 4. Schematic hierarchy of theories and models (after Gustafson and Olsson, 1993).

Conceptual models can be of many different types depending on the problem. For contaminated sites ASTM (1995) has developed a standard guide for *conceptual site models*. ASTM defines a conceptual site model as "...a written or pictorial representation of an environmental system and the biological, physical, and chemical processes that determine the transport of contaminants from sources through environmental media to environmental receptors within the system". The conceptual model is used to integrate all site information and to determine whether information is missing, and whether additional information needs to be collected (ASTM, 1995). Six basic activities are identified for developing a conceptual site model:

1. identification of potential contaminants
2. identification and characterisation of the sources of contaminants
3. delineation of potential migration pathways through environmental media
4. establishment of background areas of contaminants for each contaminated medium
5. identification and characterisation of human and ecological receptors
6. determination of the limits of the study area or system boundaries

Uncertainties in the conceptual site model need to be identified so that efforts can be taken to reduce them. This is especially important for early versions of conceptual models based on limited and incomplete information (ASTM, 1995). However, ASTM do not mention how this should be achieved in practice.

To perform some of the activities listed above, especially activity 3 and 6, it is necessary to take the geologic and hydrogeological conditions into account. A common way to do this is to develop conceptual geologic or hydrogeologic models. A conceptual hydrogeological model describes qualitatively how a groundwater system functions (Koltermann and Gorelick, 1996). Poeter and McKenna (1995) emphasise the importance to work with a range of different interpretations of the subsurface to take the uncertainty into account. Similarly, a set of different conceptual models is often considered when uncertainty are handled for radioactive waste disposal studies (Äikäs, 1993). Yuhr et al. (1996) points out the importance of considering geologic uncertainty in site characterisation and ways to reduce it.

Uncertainty arises from the choice of conceptual model, including the assumptions about relevant physical processes (James and Oldenburg, 1997). These uncertainties can be as important as the estimates of contaminant levels themselves. Whenever a model based on limited data is used, model assumptions take the place of data. This can increase the subjectivity at the loss of objectivity (Borgman et al., 1996a).

James and Oldenburg (1997) have investigated the uncertainty in transport parameter variance and site conceptual model variations for a large-scale 3-D finite difference transport simulation of trichloroethylene concentrations using FOSM and Monte Carlo methods. To transform the actual site history and conditions into a numerical simulation system a conceptual model was defined. A set of four conceptual models was defined with the purpose to take the *conceptual model uncertainty* into account; one base case and three variations of the base case. For each of the four conceptual models the same set of parameter uncertainties (variances) were used for uncertainty analysis. The result shows that large uncertainties in calculated contaminant concentrations arise from both parameter uncertainty and choice of conceptual model. Especially important was uncertainty about the subsurface heterogeneity. Because of the large uncertainties, a conclusion is that predicted contaminant concentrations should always include estimates of uncertainty. James and Oldenburg (1997) point out that uncertainty in the numerical simulation model was not considered in the study, only parameter uncertainty and conceptual model uncertainty.

3.4 Combining hard and soft information

A conceptual model are qualitative in nature but can be developed into some kind of quantitative model by combining hard and soft information. Depending on the problem

many different types of quantitative models are used, e.g. for hydrogeological simulations or contaminant transport simulations.

Koltermann and Gorelick (1996) reviewed techniques for incorporating geological information in models for generating maps of hydraulic properties in sedimentary deposits. They distinguish three different classes of models for this purpose: (1) structure-imitating methods, (2) process-imitating methods, and (3) descriptive methods. *Structure-imitating methods* try to imitate the geometry of spatial patterns in the geologic media by techniques like spatial statistics, correlated random fields, probabilistic rules etc. *Process-imitating methods* try to model the important processes. This can be achieved by using aquifer models and geologic process models. *Descriptive methods* combine site-specific data and regional data with a conceptual geologic model. The aquifer is divided into zones and parameter values are assigned to each zone depending on the conceptual geologic model and available measurements. A consequence of this is that some of the descriptive methods are *data-driven*, i.e. they are heavily based on site-specific measurement. Others are *interpretation-driven*, i.e. they are primarily based on qualitative geologic information (Koltermann and Gorelick, 1996). Some models can incorporate soft information fairly well, while others are unable of doing that. As an example, indicator-based methods allow great flexibility to include hard and soft data. Including soft information is often critical because hard data are often sparse (Koltermann and Gorelick, 1996).

Bayesian methods have been increasingly used for site characterisation during the last decade. McLaughlin et al. (1993) applied these techniques to characterise a coal tar disposal site. The strategy they used consists of four steps. First, a probabilistic description of the natural heterogeneity in the subsurface is made. Second, prior statistics describing the geological properties, hydraulic variables, and contaminant concentrations are derived. The prior statistics are used in a solute transport model to make a prior estimate. Third, a procedure is developed to update the prior statistics with field data. Finally, a sequential program of data collection and updating is carried out, resulting in better and better characterisation of the subsurface.

Freeze et al. (1990) transform a conceptual model into a *geological uncertainty model* and a *parameter uncertainty model* by incorporating the uncertainties. These models are used together with a *hydrogeological simulation model* to take the uncertainties in geology and in parameter values into account by Bayesian updating.

Johnson (1996) used a similar approach and used Bayesian methodology to integrate soft information with hard data. Soft information of where contamination is likely to be, is used to develop a conceptual model. This image is updated with new sample data by indicator kriging.

Rosenbaum et al. (1997) discuss different probabilistic models for estimating lithology. They include geostatistics (indicator kriging and indicator cokriging), conditional simulation, and Bayes-Markov simulation.

3.5 Uncertainty in quantitative models

A type of uncertainty that may be very important but often is overlooked is the uncertainty in the model itself, i.e. the model structure is the source of the uncertainty (see section 2.3). As a typical example, this type of uncertainty could be the inherent uncertainty in the hydrogeological simulation model due to the fact that the model is a simplification of reality. This uncertainty may be very difficult to quantify or separate from other sources of uncertainty. A common name for it is *model uncertainty* (Morgan and Henrion, 1990) and this is also the name used in this report. Both qualitative and quantitative models are affected by model uncertainty. Qualitative models are often equivalent to conceptual models and the uncertainty associated with these models has already been discussed. The quantitative models, on the other hand, can be of different nature. Typical examples include *mathematical models*, *numerical models*, and *spatial statistical models*.

However, some authors define model uncertainty in other ways. Wagner (1999) and Lacasse and Nadim (1996) use model uncertainty in the context of uncertainty in model predictions. The latter define model uncertainty as “...*the ratio of the actual quantity to the quantity predicted by a model*”. This is an important difference in definition since the uncertainty in model predictions also includes all introduced uncertainty in data input (an ideal model that models reality correctly may still give imprecise predictions if data are uncertain). With the definition according to Lacasse and Nadim (1996) the model uncertainty can be expressed with the concepts of bias and precision by calculating the ratio of the actual quantity to the quantity predicted by the model. A mean value different from 1.0 expresses bias in the model, while the precision can be expressed by the standard deviation of the model predictions.

Lacasse and Nadim (1996) argue that it is absolute more rational to include model uncertainty than to ignore it. Model uncertainty is generally large but it can be reduced. They describe how model uncertainty can be included in a geotechnical calculation model. Examples of how to evaluate model uncertainty is to compare model tests with deterministic calculations, pooling of expert opinions, results from literature, and engineering judgement. For a qualitative understanding of model uncertainty they identify and list factors of influence on model uncertainty.

Sturk (1998) recognises three causes for modelling uncertainty; (1) *professional uncertainty*, (2) *simplifications*, and (3) *gross errors*. Professional uncertainty arises because of lack of knowledge of the studied phenomenon. Simplifications are introduced in models to make them less complex. Gross errors result because of omissions or lack of competence.

Russell and Rabideau (2000) evaluated the importance of e.g. model complexity in a risk-based decision analysis framework, see section 6.4.4

In practice, model uncertainty and parameter uncertainty often overlap (Taylor, 1993). Sometimes, it is possible to construct a model in such a way that model uncertainty is converted to uncertainty about parameter values. Such an approach often simplifies the analysis. In situations when a phenomenon is modelled by different models, Morgan and Henrion (1990) argue that it is inappropriate to assign probabilities to the different models. Rowe (1994) comes to the same conclusion and states that probability has no meaning here.

Knopman and Voss (1988) developed a methodology for comparing different models (*model discrimination*) based on error in contaminant transport model predictions. The error is quantified by regression analysis. They define a model error vector as:

$$E = E_S + E_R$$

where E_S is systematic error and E_R is random error. The systematic error is introduced when the physical system is described with an incorrect mathematical description. The source of random errors is measurement errors and inability of the model to capture stochasticity in the physical system. The model error may be in three forms; (1) over-specification, (2) underspecification, and (3) misspecification of the mathematical structure approximating the true system.

Geostatistical modelling includes several techniques, such as variogram analysis and a number of different kriging methods. Estimation errors may be introduced in different ways such as; insufficient number of samples, sample data of poor quality that do not represent the actual concentrations, and poor estimation procedures (Myers, 1997). Kitanidis, as referred by Koltermann and Gorelick (1996), notes that *variograms* are often calculated and used without regard for the uncertainty in their parameters, i.e. the variograms are considered deterministic. To take the uncertainty of a variogram into account a single, deterministic variogram should not be used; instead a set of variogram models that fit the experimental data should be used.

Deutsch and Journel (1998) discuss the uncertainty about uncertainty models used in stochastic simulation. They recommend sensitivity analysis to be used for model parameters, especially decision variables, but also parameters such as the variogram range.

Dagdalen and Turner (1996) conclude that *kriging* is likely to lead to misinterpretation of the extent and degree of contamination at a site when it is applied without consideration of geological complexity. They discuss some aspects of this problem.

4 SAMPLING UNCERTAINTY

4.1 Introduction

The result of a site-investigation depends on samples of good quality. However, it is often extremely difficult to obtain quality and reproducible samples and the requirements to achieve sampling correctness are not widely known (Myers, 1997). Therefore, major sampling errors often occur. Maps of spatial contamination are often based on the assumption that the available data are reliable, which is often not the case, so one is simply mapping an illusion provided by the available data (Myers, 1997). This may, of course, affect the remedial decision and have a negative impact on both the economics and the quality of the site cleanup. There are different ways of approaching this problem. One is by using sampling theory developed by Pierre Gy to assist in the search and extraction of mineral resources. This theory, and other ways of handling sampling uncertainty for environmental applications, will be described in this chapter.

Typically, the error arising from field sampling is much larger than that associated with post-sampling activities, such as sample preparation and chemical analysis (Crépin and Johnson, 1993). Sampling is an integral part of the measurement process that is often neglected in terms of the errors it generates (Ramsey et al., 1995). In practice there are two questions that generally arise:

1. How many samples should be taken?
2. Where should the samples be located?

Uncertainties related to these and other sampling questions are addressed in this chapter.

4.2 Sampling key terms

When the sampling literature is reviewed it becomes clear that an agreed terminology is lacking. An author may use the word “sample” in one context, while other authors may have a different definition. As an example, the soil volume analysed at a laboratory may have at least the following different names: *sample*, *analysis sample*, *subsample*, *test portion*, *aliquot*, or *split*. Pitard (1993) and Myers (1997) present extensive lists of key word explanations related to sampling. These are based on the particulate sampling theory presented in section 4.7. Other important sampling terms can be found elsewhere, e.g. in Keith (1996). The European Committee for Standardization (CEN) also defines several sampling key terms during its present standardisation work.

It is very important to understand the concepts of *target population*, *sampled population* (*sample domain*), and *sample*. If they are not clearly defined and related to the objective of the investigation, the data collection may contain little useful information (Myers, 1997). The target population is the real population of interest. When the target and sampled populations are not the same, there is a potential for biased results (Chai, 1996), i.e. sampling from mixed domains must be avoided.

Myers (1997) makes a distinction between *sampling strategy* and *sampling plan*. Sampling strategy is a standardised set of goals and conditions that provide for correct sample design, correct sample collection, and correct spatial assessment. A sampling plan on the other hand, is a unique set of goals and conditions developed for a particular site. Therefore, a sampling plan can never be standardised, in contrast to a sampling strategy.

4.3 Sampling objectives

The literature about soil sampling is extensive but Huesemann (1994) identifies at least two major limitations regarding it. Either the sampling publication is too sophisticated to be understood by an average scientist or engineer, or the document is too concerned with generalities such as QA/QC procedures (see section 4.4). Another limitation that is addressed is the difficulty to find a sampling publication that matches one's *sampling objective* and specific needs. In the literature it is relatively rare to find discussions about different sampling objectives. However, before measurements can be made, the concept of the problem to be solved and the model to be followed for its solution must be reasonably clear (Taylor, 1996). Koerner (1996) states that to collect a representative sample successfully, one must first clearly define the objectives of the sampling exercise. The most commonly mentioned sampling objectives appear to be:

1. to determine the average site contamination level (mean concentration)
2. to classify the soil in different concentration classes during, or prior to, remediation
3. to locate "hot spots"
4. to delineate the contaminated area/plume
5. to create a contour map (isopleth) over contaminant concentrations
6. to forecast the contamination level during excavation (i.e. the concentration that can be expected as excavation proceeds)
7. to determine which chemical substances are present
8. to monitor concentration changes over time

Scientific or statistically sound methods to reach several of these, and similar sampling objectives, are described in the environmental literature. In practice, several other, and sometimes less stringent, sampling objectives exist. One is to get a rough estimate about the concentration level at a site (Huesemann, 1994). This objective is a rather subjective one, but in practice it is used quite commonly. Such vague sampling objectives make data evaluation problematic. Taylor (1996) mentions a question that is commonly asked but cannot be answered by sampling: *Are all members of the population within acceptable limits?* To answer this and similar questions the whole population has to be analysed.

Several authors, e.g. Shefsky (1997), emphasise the importance of clearly stating the sampling objective in quantitative terms. Action or decision concentrations should be specified with demands on the level of precision that is required.

In the following sections (4.4-4.8) some aspects that are common to several sampling objectives will be described. They include; quality aspects and uncertainties in the planning phase, which sampling pattern to use, sampling theories, especially the particulate sampling theory of Pierre Gy, and composite sampling. After that, uncertainty associated with some of the eight sampling objectives listed above will be described (section 4.9-4.11). These include:

- Sampling to estimate mean concentrations. This is relevant for sampling objective number 1 and 2 above.
- Sampling to detect hot spots (sampling objective 3).

- Sampling to delineate contaminated areas. This is relevant for sampling objective 4 and 5.

Sampling objective 6-8 will not be addressed further. Sampling objective 2 (classification of soil in concentration classes) will not be discussed explicitly. Such methods are described by for example Naturvårdsverket (1997).

4.4 Planning phase (DQOs and QA/QC)

Data Quality Objectives (DQOs) and *Quality Assurance/Quality Control* (QA/QC) systems consist of protocols and procedures for sampling and analysis that should be followed to keep the uncertainty at acceptable level. Many of the procedures in DQOs and QA/QC are qualitative in nature, although quantitative advice sometimes is presented. These concepts and procedures are developed by U.S. EPA and are presented in a number of guidelines and reports. A thorough discussion of QA/QC samples is given by for example van Ee et al. (1990), who also provide QA methods and equations for quantification of total measurement error.

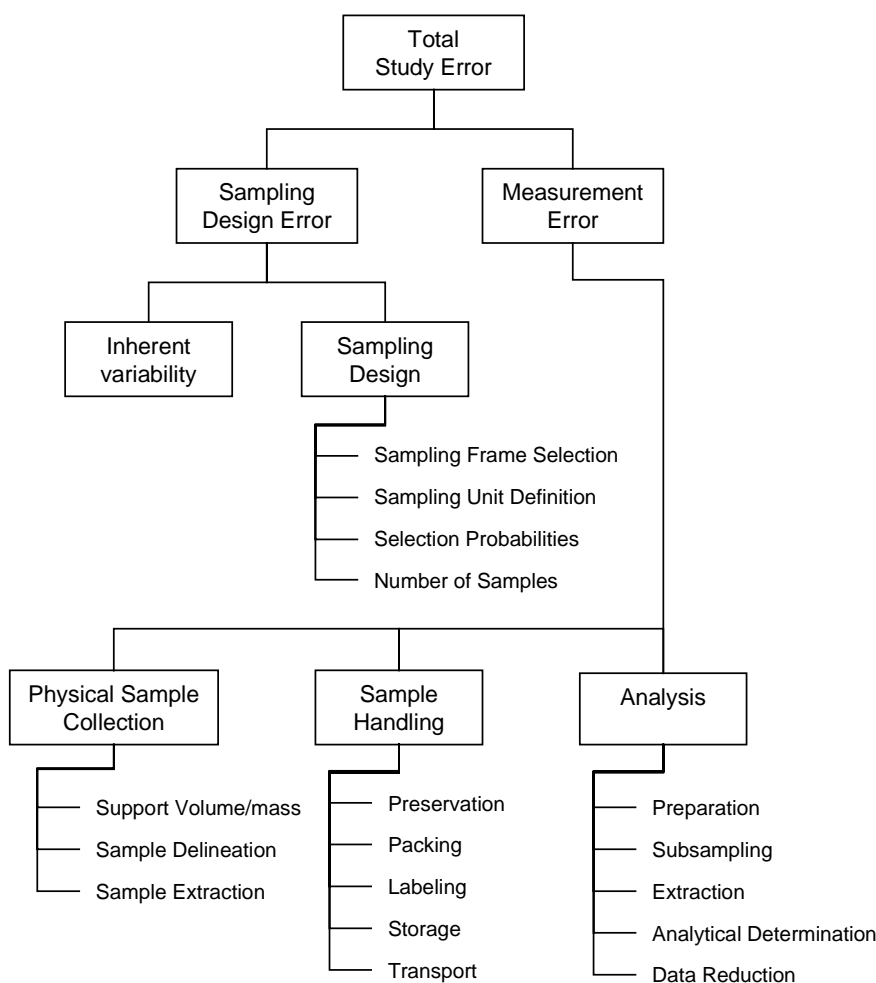


Figure 5. An example of how Total Study Error can be broken down by components (after U.S. EPA, 1994).

The DQO process (U.S. EPA, 1994) is a planning approach to develop sampling designs for data collection activities that supports decision-making. The system should be used on a project level but it is not mandatory to follow the DQO in the U.S. EPA quality system. The process is iterative. The DQO enables evaluation of two opposing conditions, such as if a site should be regarded as contaminated or not. The outcome of the DQO process is a sampling design for collecting data, including the number of samples to collect, where and how to collect them, and acceptable limits on the probability of decision errors. The DQO process can be summarised as (Robbat, 1996):

1. State the problem
2. Identify the decision to be made
3. Identify the inputs to the decision
4. Define the boundaries of the study
5. Develop decision rules
6. Specify tolerable limits on decision errors
7. Optimise the conceptual model

In step 5, the choice of a statistical decision rule is encouraged. Two types of decision errors are considered in step 6; false acceptance and false rejection. In this step the Total Study Error is broken down into sampling design errors and measurement errors according to Figure 5.

Computer software related to the DQO process is presented in the appendix. A framework for integration of the DQO process with *particulate sampling theory* (section 4.7) and geostatistics is presented by Myers (1997).

4.5 Sampling pattern design

A great number of different sampling designs (sampling patterns) are described in the literature. Here, the most common ones will be described briefly, mainly organised according to Gilbert (1987) (other authors uses a somewhat different classification). A detailed presentation of sampling designs is also given by U.S. EPA (2000a). In this section, our emphasis is on spatial sampling, although many of the described strategies can also be applied on temporal sampling. Four basic ways of selecting sampling locations can be distinguished:

1. Haphazard sampling
2. Judgement sampling
3. Search sampling
4. Probability (statistical) sampling

In addition, *hybrid sampling plans* (Taylor, 1996) is sometimes used, i.e. combinations of the ones listed above. Taylor (1996) identifies a fifth type of sampling design that is commonly used: *protocol sampling*. This type of sampling follows specified legal or contractually mandated plans. A common feature of the mentioned sampling strategies is that they use a fixed sampling pattern. A different approach is the *adaptive sampling* strategy (e.g. Cox, 1999; Johnson, 1996; Robbat, 1996), where the sampling is performed in stages and the information from one stage is used to locate the sampling

points for the next stage. This approach usually accounts for patterns of variability by geostatistics. Examples of adaptive sampling are presented in chapter 6. An innovative and relatively unknown sampling design is the *ranked set sampling* (U.S. EPA, 2000a). It combines simple random sampling with the field investigators professional judgement. The advantages are cost-efficiency and better estimates of mean concentrations.

In on-going work on sampling techniques for waste, the European Committee for Standardization (CEN) uses a somewhat different classification of sampling designs. CEN differentiate between *probabilistic sampling* and *methodological sampling*. Probabilistic sampling is statistically based, while methodological sampling is not. Two types of methodological sampling can be distinguished, *informative* and *non-informative*. Informative methodological sampling is probabilistic sampling from a subpopulation, resulting in partial representative samples. With non-informative sampling there is no way of judging how representative a sample is.

During *haphazard sampling* the samples are taken with no systematic method in mind and the method should not be used (Huesemann, 1994). The philosophy behind haphazard sampling is that “any sampling location will do”, but this assumption is highly suspect in environmental pollution studies (Gilbert, 1987). Thompson and Ramsey (1995) use the term *convenience sample* for a sample taken with no regard to representativeness. Several authors have pointed out the uselessness of haphazardly taken soil samples.

Judgement sampling is subjective sampling of population units selected by an individual (Gilbert, 1987). Taylor (1996) calls this strategy *intuitive sampling* while U.S. EPA (2000b) uses the term *authoritative sampling*. If the individual is skilful this strategy can be successful but the result depends heavily on the personal judgement of the surveyor. However, there is no way to verify the result, except through additional random sampling, and no estimation of error is possible (Crépin and Johnson, 1993; U.S. EPA, 2000b). Therefore, judgement sampling should not be used to characterise the average contamination level at a site (Huesemann, 1994). Despite its drawbacks judgement sampling is often used in practice. Often, judgement sampling does not refer to using prior information to design a sampling strategy. The use of prior information and expert opinion is clearly needed in many situations (Gilbert, 1987).

One type of sampling that often can be regarded as judgmental is *preferential sampling*. Goovaerts (1997) defines preferential sampling as when sampling locations “...are neither regularly nor randomly distributed over the study area”. Some reasons for preferential sampling may be; (1) restrictions in accessibility, (2) denser sampling in critical areas, or (3) clustered samples to characterise short-range variability. Any preferential sampling is likely to impact sample statistics (Goovaerts, 1997).

Search sampling is a sampling strategy that is used to locate “hot spots”, plumes, or pollution sources. A sampling pattern is designed to find an object, of defined size and shape, with a specified probability of a hit. The method is based on search theory, which can be used to find the optimum spacing between sampling points and to minimise the number of samples. Often, prior information and expert judgement are used to design the search-sampling plan. The validity of such a procedure depends on the accuracy of prior information and on accurate measurements (Gilbert, 1987). A simplification in

search theory is that the search objects is approximated as having simple shapes, often circular or ellipsoid shapes.

Probability sampling refers to the use of a specific method of random selection (Gilbert, 1987) and is based on probability theory. It can be used to estimate the average contamination level at a site. Another common name for this type of strategy is *statistical sampling*. In practice, random sampling is achieved only by following strict rules so that all individual samples in a population are selected without bias. The most commonly used probability sampling designs include:

- Simple random sampling
- Stratified random sampling
- Multistage sampling
- Cluster sampling
- Systematic sampling
- Systematic random sampling
- Double sampling

Four of these sampling designs are illustrated in Figure 6.

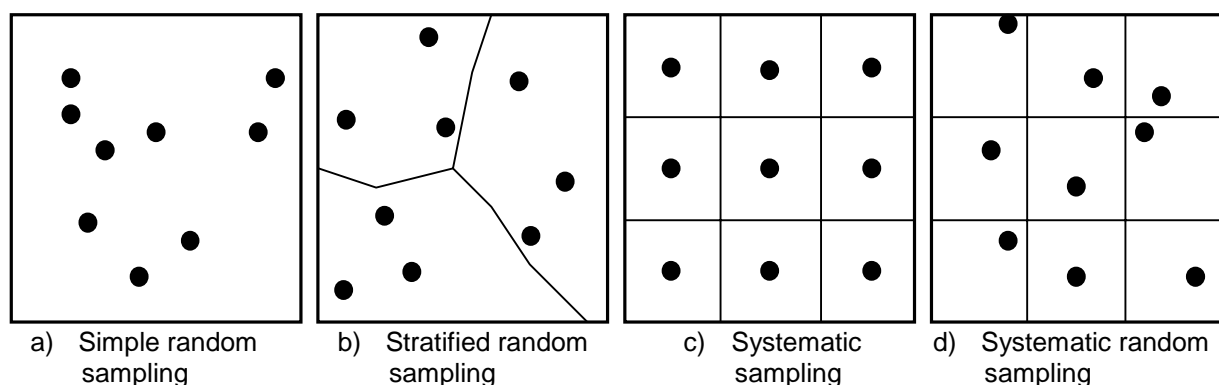


Figure 6. Illustration of some probability sampling-designs, applied to spatial sampling.

In *simple random sampling* (or strict random sampling) the sample position is located randomly, i.e. all personal bias relating to landscape position is avoided and each location has an equal probability of selection (Figure 6a). Simple random sampling is appropriate for estimating means and totals for the entire population, but only if the population does not contain major trends (Gilbert, 1987). For small (< 0.5 ha) and uniformly affected sites, as few as five to ten samples may be sufficient, while larger sites may need up to 25 samples. If a sufficient confidence level cannot be achieved with 25 samples, the site should be divided into less variable units (Crépin and Johnson, 1993). It is important to note that random sampling is not equivalent to picking sampling locations haphazardly (Gilbert, 1987). A major problem with simple random sampling is that clusters tend to arise, which can lead to non-representative sampling (Myers, 1997). This can be avoided if systematic random sampling is used (see below).

In *stratified random sampling* the site is divided into a number of non-overlapping parts or subareas called strata, and within each strata sampling positions are located randomly (Figure 6b). The site can be divided into different strata based on expected differences in contamination levels (Huesemann, 1994) or on geological grounds. The method of stratified sampling can be used to (1) make statements about each strata separately, or (2) to increase the precision of population estimates by distributing the samples over the entire area. An advantage of the stratified random sampling is that the increased homogeneity in each strata, increases sampling precision. Stratified sampling also makes it possible to apply a number of statistical techniques (Crépin and Johnson, 1993).

In *multistage sampling* the population is divided into primary units and a set of primary units is selected by simple random sampling (Gilbert, 1987). Then, each sample is randomly subsampled (two-stage sampling). An example of two-stage sampling in contaminated land investigations is the random collection of soil samples, followed by random selection of aliquots (test portions) from each soil sample.

Another definition of multistage sampling is when an area is sampled several times with data evaluation performed between each stage. This strategy is similar to the adaptive sampling strategy. It results in better spatial definition of contaminants for a given sample size (Ferguson, 1993). If two-stage sampling is performed, about half of the total number of samples should be assigned to each stage, as a rule of thumb (Ferguson, 1993). A more formal solution can be derived by Bayesian statistics.

Cluster sampling is where a cluster of individual units are chosen at random, and all units in the cluster are measured (Gilbert, 1987). This sampling design is of little interest at contaminated sites and will not be discussed further.

In *systematic sampling* (or grid sampling) the sampling points follow a simple pattern and are separated by a fixed distance, usually in a regular grid (Figure 6c). Locating sampling points in the field is easier with systematic sampling than with simple random sampling. The first sampling point has to be located randomly and the rest of the grid should be laid out according to that point. Random sampling can always be replaced by systematic sampling and tools like geostatistics can be applied (Crépin and Johnson, 1993). Statistical studies indicate that systematic sampling may be preferred over other sampling plans for estimating means, totals and patterns of contamination (Gilbert, 1987). Systematic sampling can also be used with a temporal pattern, i.e. samples can be taken at specified time intervals.

One drawback of the systematic sampling is the loss in precision when the population has a periodic trend that corresponds to the sampling interval. One such example is when the contaminant has been transported along a drain parallel to the sampling grid. Also, the variance of estimated means and other quantities are more complicated to calculate than for simple random sampling (Gilbert, 1987).

Systematic random sampling (or random sampling within blocks or segments) is a combination of random sampling and systematic sampling. It has the advantage of both methods and in many cases it is the method of choice (Huesemann, 1994). Random samples are taken within a systematic rectangular or square grid area (Figure 6d), i.e. the study area is better covered compared to simple random sampling. Systematic ran-

dom sampling also has advantages over systematic sampling if systematic contamination patterns are present at the site.

4.6 Sampling theories

Borgman et al. (1996b) distinguishes between two types of sampling theories:

1. design-based sampling (classical finite sampling theory), and
2. model-based sampling

The main distinction between the two is that in model-based sampling a model is used to account for patterns of variability within the population. One example of the model-based approach is when a geostatistical model is used to design the sampling. Model-based sampling is usually more effective than design-based sampling because it makes more complete use of information about the population. In design-based (classical) sampling no assumption of the underlying population is made, which makes it fundamentally objective (Borgman et al., 1996b). The theories to determine mean concentrations (section 4.9) and to detect hot spots (section 4.10) are examples of design-based sampling.

A third sampling theory is available, except from the two mentioned: The particulate sampling theory of Pierre Gy. This theory will be described below.

4.7 Particulate sampling theory

4.7.1 Introduction

Although several articles and books about sampling have been written there is only one sampling theory that can be regarded as general and comprehensive. This sampling theory was developed by the French mining-engineer Pierre Gy in the 1960s, 1970s and 1980s for the mining industry. The theory is applicable to the sampling of particulate materials and fluids (Borgman et al., 1996b). Other theories only cover small parts of the sampling problem but Gy's theory is the only theory of sampling of particulate material that is accepted and undisputed world-wide (Pitard, 1993). However, it is not evident that the theory is applicable to all aspects of environmental sampling, for example sampling of volatile organic compounds (VOC) (see section 4.7.7).

The theory has connections with both classical statistics and geostatistics (Borgman et al., 1996b). Geostatistical aspects of Gy's theory are addressed by Flatman and Yfantis (1996) and Myers (1997). A improvement upon classical statistics is that the various error sources are directly addressed. The separation of error sources is an important aspect because some errors may represent bias, not simple random variation (Borgman et al., 1996b). The various components of variance of this sampling theory sound trivially obvious when pointed out, but they are easily overlooked (Flatman and Yfantis, 1996). Also, sampling theory shows that seemingly harmless assumptions may result in substantial sampling errors (Myers, 1997).

Gy has presented his work in a number of French and English publications but the high complexity has restricted the use of his sampling theory by engineers and scientists. A more accessible presentation of the theory is presented by Pitard (1993). Shorter pres-

entations of the theory, with the focus on environmental sampling, are available in for example Myers (1997), Shefsky (1997), Borgman et al. (1996b), Flatman and Yfantis (1996), and Mason (1992).

4.7.2 Sampling errors

According to the sampling theory by Pierre Gy there are seven basic sampling errors:

1. The Fundamental Error (FE)
2. The Grouping and Segregation Error (GE)
3. The Long-Range Heterogeneity Fluctuation Error (CE_2)
4. The Periodic Heterogeneity Fluctuation Error (CE_3)
5. The Increment Delimitation Error (DE)
6. The Increment Extraction Error (EE)
7. The Preparation Error (PE)

These errors, together with the *analytical error* AE , are components of the *overall estimation error* OE . For convenience the first six errors can be combined to shorten mathematical formulas; the *short-range heterogeneity error* $CE_1 (=FE+GE)$, the *continuous selection error* $CE (=CE_1+CE_2+CE_3)$, and the *materialisation error* $ME (=DE+EE)$. The following error sum relationships can be formulated:

Sampling Selection Error, SE : $SE = CE + ME$

Total Sampling Error, TE : $TE = SE + PE$

Overall Estimation Error, OE : $OE = TE + AE$

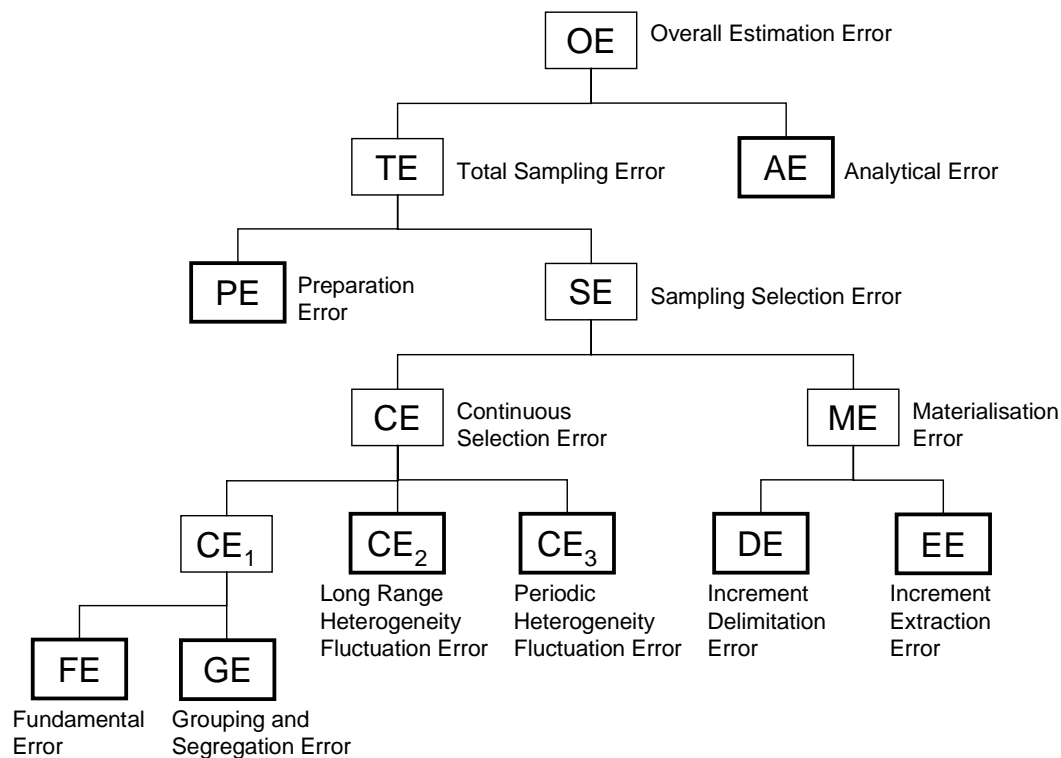


Figure 7. Illustration of the relationships between different types of errors, according to Pierre Gy's sampling theory.

The relationship between the different errors is illustrated in Figure 7. The meaning of the basic sampling errors is not self-evident and is therefore explained below.

A key word in the sampling theory is sampling *correctness*. A sample methodology is considered unbiased and correct if all of the particles in the sampling unit have exactly the same probability of being included in the sample (Shefsky, 1997).

The Fundamental Error (FE)

The fundamental error is an important error because it still remains when a sampling operation is perfect. Since the fundamental error affects the measured concentration it is critical to manage this error, and it is the only error that never cancels out. It is also the only error that can be assessed independently of the sampling method (Borgman et al., 1996b).

The fundamental error is a statistical consequence of the particulate nature of geologic samples (Shefsky, 1997), see also the Representative Elementary Volume in Figure 8. As defined by Gy, the fundamental error is the relative error in estimating the grade (grade of ore may be analogous to the percentage of some chemical present) (Borgman et al., 1996b). It is caused by the range of particle sizes in the medium and the fact that often only certain sized particles contain the pollutant of interest (Flatman and Yfantis, 1996). Also, the contaminant itself can exist as particles such as lead shots or paint chips, which may result in large fundamental errors as illustrated by Shefsky (1997). On the other hand, if all the contaminant particles are very small, the fundamental error will also be small. Pitard (1993) provides equations and sampling nomographs for quantification of the fundamental error as a function of some physical properties. The most important of these properties is the maximum particle size. Mason (1992) provides a practical guide for determining the maximum and minimum variance of the fundamental error. A statistical basis of the fundamental error, with illustrative explanations, is presented by Shefsky (1997). He also provides practical advice on how to avoid or reduce fundamental error and bias in soil sampling.

It is important to note that a fundamental error is introduced at each stage of sampling. Therefore, repeated subsampling leads to accumulation of fundamental errors, which quickly can exceed 50-100 % (Myers, 1997). Because only a small mass of soil usually is used in laboratory analysis, the stage of subsampling for the analytical sample is the stage most apt to unacceptable magnitude of the fundamental error (Flatman and Yfantis, 1996). Shefsky (1997) states that if “...we subsample without any regard to homogenisation or particle properties, the result will be analytical disaster”.

If a large fundamental error is present, the sample will not represent what is in the field and it may be a waste of time to spend money to make other errors small (Borgman et al., 1996a). The magnitude of the fundamental error increases when the constitution heterogeneity (see section 4.7.3) increases (Flatman and Yfantis, 1996).

The Grouping and Segregation Error (GE)

The grouping and segregation error results from the distribution heterogeneity (see section 4.7.3). The heterogeneity may be in density, particle size, shape, adhesion, cohesion, magnetism, affinity for moisture etc. so that the particles come together by groups during movement or vibration (Flatman and Yfantis, 1996). Under normal conditions

this error has a non-zero mean. The variance of the grouping and segregation error can be written as:

$$s^2(GE) = s^2(CE_1) - s^2(FE) = \gamma_s \xi s^2(FE)$$

where γ_s is the grouping factor and ξ is the segregation factor. In practice there are three ways of minimising $s^2(GE)$:

1. To minimise the variance of the fundamental error $s^2(FE)$.
2. To minimise the grouping factor γ_s . In practice, this may be the most effective way to minimise the variance $s^2(GE)$.
3. To minimise the segregation factor ξ by homogenisation techniques. However, Myers (1997) warns that homogenisation of heterogeneous material is often wishful thinking and may instead promote segregation, e.g. by the gravity force. To avoid such errors, mechanical sampling splitting devices could be used (Shefsky, 1997).

Theoretically, this error grows larger without bounds if the size of the analytical sample approaches the size of the grains in the sample. The consequence of this is that when subsampling is performed in the laboratory, the chemist can turn the analytic equipment into a random number generator if the sample material has not been prepared correctly (required fineness and correct subsampling procedure) (Flatman and Yfantis, 1996).

The Long-Range Heterogeneity Fluctuation Error (CE₂)

This error is due to the long-range heterogeneity h_2 described in section 4.7.3. It is generated by local trends. The variance of this error, $s^2(CE_2)$, can be quantified by the variogram of geostatistics.

The Periodic Heterogeneity Fluctuation Error (CE₃)

This error is introduced by cyclic phenomena. It is a non-random error generated by heterogeneity h_3 (see section 4.7.3).

The Continuous Selection Error (CE)

The continuous selection error is the sum of the four errors above ($FE+GE+CE_2+CE_3$). It is regarded as the result of the total heterogeneity contribution h (see section 4.7.3) and is the sum of errors generated by the process of material selection during sampling.

The Materialisation Error (ME)

Many of the errors introduced by heterogeneity are typically random errors. Materialisation errors, on the other hand, are almost always biased and are probably the major cause of all biases in sampling (Myers, 1997). The materialisation error is the sum of the *increment delimitation error (DE)* and the *increment extraction error (EE)*.

The *increment delimitation error* results from an incorrect shape of the volume delimiting the increment (in soil sampling an increment is the physical piece of soil removed from the ground and from which the sample is taken). Incorrect shape of the delimitation tool (sampling device) can lead to unequal probability for material to be part of the increment, i.e. the concept of equiprobability in all directions is violated. A shovel with a round shape is an example of a sampling device that causes a delimitation error. The shovel will preferentially collect material from the top of the lot, with less material extracted from the bottom. This introduces a delimitation error and a bias.

The *increment extraction error* results from an incorrect extraction of the increment. The extraction is said to be correct if all particles with their centre of gravity within the boundaries of the increment will belong to the increment. Depending on the construction of the sampling device and how the sampling is performed, these particles may not be extracted as a part of the increment, or particles outside the domain of the increment may be collected. This increment extraction error is often different from zero and is therefore an important source of sampling bias.

Myers (1997) points out that most of the sampling devices available for environmental sampling are incorrect. Examples of such devices are augers, thieves, and triers. If a cylinder could be cut out exactly in the material by a laser and then taken out intact by levitation as in science fiction, these two errors could be avoided (Flatman and Yfantis, 1996).

The Sampling Selection Error (SE)

The sum of *CE* and *ME* is the *sampling selection error*. The mean of *SE* is what usually is called sampling bias. If the mean of *SE* is zero the sample is said to be unbiased, although this case is never encountered in practice (Pitard, 1993). A sample is said to be accurate when the mean of *SE* is smaller than a certain standard of accuracy.

The Preparation Error (PE)

All the previously described errors are selective while the preparation error is non-selective. The preparation can consist of a number of stages, such as comminution (grinding, crushing and pulverising), screening, mixing, drying, filtration, weighing, packing etc. (Pitard, 1993). Different types of errors can be introduced during preparation:

- *Contamination errors*, due to improper procedures, contaminated equipment, dust problems etc.
- *Loss errors*, especially for volatile organic compounds if improper sampling, storage, and handling procedures are employed.
- *Alteration of chemical composition*. This error is related to loss errors and includes for example degradation of chemical substances in the sample.
- *Unintentional mistakes*. Common examples of unintentional mistakes include dropping samples, mixing labels, mixing fractions from different samples, contamination etc. This error can be reduced if the personnel are skilled.
- *Sabotage and fraud*. This type of error may be rare but the possibility for it still exists.

The Analytical Error (AE)

This error is not part of the sampling error and is described in section 5.3. It does not include the selection of the subsample to be analysed. The error in subsampling should be a part of the total sampling error *TE*.

The Overall Estimation Error (OE)

The overall estimation error can be quantified as the sum of the variances of all sampling and analytical errors:

$$\sigma_{OE}^2 = \sigma_{CE}^2 + \sigma_{ME}^2 + \sigma_{PE}^2 + \sigma_{AE}^2$$

Usually, several stages of sampling (subsampling) and preparation are performed prior to analysis. If all types of errors are taken into account the overall estimation error can be expressed as (Pitard, 1993):

$$OE = AE + \sum_{n=1}^N (PE_n + FE_n + GE_n + CE_{2n} + CE_{3n} + DE_n + EE_n)$$

where N is the number of sampling or preparation stages. There are some important aspects of this expression; (1) there are many different types of errors to consider, (2) errors are introduced in all stages of sampling, preparation and analysis, and (3) the errors are additive. Because sampling is performed in stages, most of the errors can enter the problem many times and the stage with the greatest error will form the lower bound on the overall estimation error (Borgman et al., 1996b).

4.7.3 Heterogeneity

Myers (1997) and Pitard (1993) comment that homogeneity is in fact an illusion. Instead, heterogeneity constitutes a fundamental part of sampling theory. Two basic types of heterogeneity can be distinguished: *Constitution Heterogeneity* and *Distribution Heterogeneity*. The constitution heterogeneity is caused by the fundamental properties of the particles, i.e. the inherent variability in composition of each particle. This type of heterogeneity is fixed and homogenisation has no effect on it. The only way to reduce the constitution heterogeneity is by comminution where the properties of the particles are changed. Constitution heterogeneity is the cause of the fundamental error FE.

A lot, such as a soil volume, can be visualised as separate groups of particles making up the lot. Each group is composed of a certain number of particles with certain characteristics. The *Distribution Heterogeneity* is related to how the individual particles are represented within the groups. It depends on three factors: (1) the constitution heterogeneity, (2) the spatial distribution of particles, and (3) the shape of the lot. The distribution heterogeneity is responsible for the grouping and segregation error GE. It is also closely related to the *support* described in section 4.7.5. The distribution heterogeneity gets smaller when the support volume is increased and it is always smaller or equal to the constitution heterogeneity (Pitard, 1993).

For practical applications, a model of heterogeneity h with three different kinds of heterogeneity is suggested by Pitard (1993):

$$h = h_1 + h_2 + h_3$$

where:

h_1 = the *short-range heterogeneity*, typically random and discontinuous. It is influenced by the constitution heterogeneity and the distribution heterogeneity. The short-range heterogeneity h_1 is responsible for the sampling errors FE and GE. The variance of the short-range heterogeneity is equal to $\sigma^2(CE_1)$.

h_2 = the *long-range heterogeneity* (or *large scale segregation*) reflects local trends, usually non-random and non-cyclic. Its variance is equal to $\sigma^2(CE_2)$.

h_3 = the *periodic heterogeneity* reflects cyclic phenomena. This heterogeneity is most common in flowing streams of material, such as water. Its variance is equal to $\sigma^2(CE_3)$.

The effect of short-range and long-range heterogeneity is illustrated in Figure 8. Myers (1997) presents methods to characterise the heterogeneity h and to estimate optimum sample weights.

A concept closely related to heterogeneity, but not addressed in particulate sampling theory, is *anisotropy*. A medium is said to be anisotropic with respect to a certain property if that property varies with direction (Bear, 1988). This concept is very important for a property like hydraulic conductivity, but also for contaminant concentrations in soil or groundwater. Anisotropic contamination at a site can be handled by geostatistics, which are not included in the particulate sampling theory. However, Myers (1997) integrates sampling theory and geostatistics into the concept of Geostatistical Error Management (GEM).

4.7.4 Representativeness

One objective in a typical site-investigation is to collect “representative” samples. However, the term *representative* is always loosely defined in qualitative terms, i.e. it is a qualitative parameter (Myers, 1997). Several definitions of representativeness exist. Koerner (1996) gives the following definition: “*A representative environmental sample is one that is collected and handled in a manner that preserves its original physical form and chemical composition and that prevents changes in the concentration of the materials to be analysed or the introduction of outside contamination*”. Chai (1996) defines representative sampling as “*...a process by which a set of samples is obtained from a target population to collectively mirror or reflect certain properties of the population*”. Taylor (1996) states that it is virtually impossible to demonstrate representativeness. In fact, “representative samples” are often used to make decisions even though no real evidence is presented to verify that the sample represents anything but itself.

Sampling theory goes one step further and provides a quantitative definition of the representativeness of a sample. A sample is regarded as representative when the mean square error of the sample selection error, $r^2(SE)$, is smaller than a specified standard r_0^2 (Myers, 1997):

$$r^2(SE) = m^2(SE) + \sigma^2(SE) \leq r_0^2$$

4.7.5 Sample Support

When characterising a geologic or environmental area the physical size of the sampling unit is very important and must be specified. This concept is known as *support* (Koltermann and Gorelick, 1996; Myers, 1997). The support in soil sampling is defined as the size, shape, and orientation of the physical sample taken at a sample point. The concept is closely related to the nugget effect in geostatistics (Starks, 1986). Flatman and Yfantis (1996) define support as a larger volume that the sample is supposed to represent, often the remediation unit. The support can be illustrated with the concept of *Representative Elementary Volume* (REV) used in the field of hydrogeology (Figure 8).

The change in measured value when the volume increases is a result of heterogeneity at different scales (see section 4.7.3), and the REV is the volume where the large fluctuations cease (Bear, 1979).

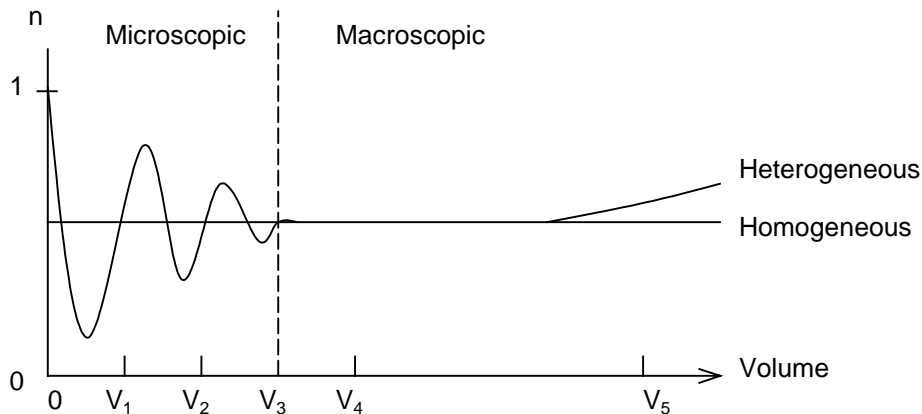


Figure 8. Illustration of the Representative Elementary Volume (REV) applied to porosity at the microscopic and macroscopic scale (after Hubbert in Freeze and Cherry, 1979). The REV is equal to the volume V_3 .

If the support is not specified, a statement of a population of measured concentrations has no meaning. Myers (1997) gives the following example of a statement that is meaningless unless the support is specified: “The concentrations of lead at the site follow a lognormal distribution, with a mean of 1743 ppm and a standard deviation of 587 ppm.” This statement is only valid for one specific sample support. If the support is changed the statement will most probably be incorrect.

Another way of describing the concept of support is as follows: A sample point is seldom located exactly. Often, the sampling team may miss the point by several meters. The support should be chosen in such a way that error variance due to misses will be small (Starks, 1986). Samples with smaller support volumes have higher variances and to reduce the variance the support can be increased. One way of increasing the support is to increase the diameter of the sample (core), but this is often impractical. A better method is to take several samples in the vicinity of the designed sample point. Starks (1986) suggests taking the samples in a square grid around the sampling point, while Jenkins et al. (1996) used a 20 cm wide “sampling path” of circular shape. Each sample is then analysed, or compositing (see section 4.8) of all samples could be performed to save cost.

The support volume has important implications for estimation of the volume of contaminated material for remediation. It is tempting to estimate the volume based on the sampling distribution of concentrations but such an approach can lead to serious estimation errors and high unexpected cost. Myers (1997) describes how the support vol-

ume influences the variance of the sampling distribution curve. Any change in support must be checked using a nomogram made up for the specific site (Flatman and Yfantis, 1996). If the support volume increases the variance decreases. Myers (1997) presents two methods to correct for small sample support; the *affine correction* and the *indirect lognormal correction*.

4.7.6 Dimension of lots

An important concept in the particulate sampling theory is the dimension of the lot to be sampled. Lots can be classified as zero-, one-, two- or three-dimensional lots (Pitard, 1993). One example of a *zero-dimensional lot* is randomly distributed coloured balls in a jar, i.e. a lot that is considered a statistical population. A *one-dimensional lot* could be a stream of water or material on a conveyor belt. The one-dimensional lot always has a linear or chronological behaviour (Pitard, 1993). An example of a *two-dimensional lot* is a piece of flat contaminated land where the depth of the lot is small compared to the areal extent. *Three-dimensional lots* are for example piles of material, truckloads, or soil units where the thickness is not negligible.

It is possible to perform statistically correct sampling only from zero- and one-dimensional lots. Acceptable sampling is possible for two-dimensional lots under favourable conditions but it is practically impossible to collect correct samples for three-dimensional lots, except in liquids (Myers, 1997). However, it is always possible to divide a three-dimensional lot into a series of two-dimensional strata.

Derived equations in the sampling theory (Myers, 1997; Pitard, 1993) are primarily applicable to zero- and one-dimensional lots. Two- and three-dimensional lots require geostatistical techniques.

4.7.7 Application to environmental problems

One of the key concepts of Gy's theory is that one can work with the particle size (by grinding) and the sample weight in order to reduce the variance to an acceptable level (Mason, 1992).

The theory was not developed for environmental problems but many aspects of the theory are still relevant for such problems. However, for certain environmental problems the theory may still be incomplete. For quantification of the fundamental error one important parameter is the maximum particle size of the medium. If the pollutants are heavy metals grinding will (probably) not alter the chemical composition but for volatile or semi-volatile chemicals this technique is inappropriate. However, most soils do not require grinding unless gravel, cobbles, or other large components are present (Mason, 1992).

The extension of the theory to volatile or semi-volatile chemicals is an undeveloped part of environmental sampling (Flatman and Yfantis, 1996). Further discussion on this matter can be found in Mason (1992).

4.8 Composite sampling

The process of compositing is a physical averaging of the material used to form the composite sample. Samples are taken from different locations and then mixed to in-

crease sample support (Shefsky, 1997). The measured value for a composite sample should be similar to the arithmetic mean of the measured values from the individual samples forming the composite sample (Fabrizio et al., 1995). For bulk populations, such as soil and water, Gilbert (1987) provides statistical methods for composite sampling. Fabrizio et al. (1995) describe procedures for formation of composite samples from segmented populations, i.e. populations with identifiable units such as fish etc.

G. Lovision et al. identified five stages of composite sampling (Fabrizio et al., 1995):

1. identification of the target population
2. statistical sampling
3. compositing (the actual formation of composite samples)
4. subsampling
5. measurement

Compositing is an effective way to reduce intersample variance caused by the heterogeneous distribution of contaminants (Jenkins et al., 1996). Another advantage of compositing is the cost-savings since only the composite sample need to be analysed. Myers (1997) points out that composite samples may save money but that they effectively wipe out any understanding of the long-range heterogeneity. They may be useful for calculating average concentrations but not for spatial mapping of contaminants. Gilbert (1987) and Garner et al. (1996) provide equations for cost-effective strategies of detecting hot spots using composite sampling. U.S. EPA (2000a) provides practical information about composite sampling, as well as equations.

A problem when composite samples are used to estimate the mean concentration is that a certain amount of homogeneity must be assumed, otherwise the result cannot be trusted. Naturvårdsverket (1997) present a rule of thumb of when homogeneity in the soil can be assumed (regarding contaminant concentrations). If the coefficient of variation ($CV = \text{standard deviation}/\text{mean}$) is less than 1, composite samples can generally be used.

4.9 Mean concentration – classical approaches

4.9.1 Random sampling uncertainty

The spatial heterogeneity at the sampling location is often quantified as *sampling precision*, although this may not be in accordance with the sampling theory in section 4.7. It can be quantified by taking duplicate samples at some proportion of the sampling locations, typically 10 % (Ramsey and Argyraki, 1997). This increases the sampling cost but enables quantification of sampling precision (but not sampling bias). Ramsey et al. (1995) found duplicate sampling to be a useful method for estimation of sampling precision.

The random components of sampling uncertainty, expressed as variances, are additive (Taylor, 1996):

$$s_{\text{sampling}}^2 = s_1^2 + s_2^2 + \dots s_n^2$$

where s_{sampling}^2 denotes the total random components of sampling variance, and $s_1^2 \dots s_n^2$ are the components from sources 1 ... n, respectively. Taylor (1996) suggests that it

may be difficult and time-consuming to quantify all random components of sampling uncertainty but that the overall value could be quantified by suitable experiments. He suggests taking at least seven replicate samples in a narrowly defined sampling area where the population variability is expected to be negligible. The equation above does not take population variability into account. To do this, Taylor (1996) suggests the following relationship:

$$s_{sample}^2 = s_{sampling}^2 + s_{population}^2$$

Jenkins et al. (1996) studied the short-range variability of contaminants at a sampling location. The site was contaminated by explosives. Seven discrete samples were collected at each sample location according to Figure 9. The sampling precision was described by classical statistic methods and the heterogeneity was found to be enormous at the site. The conclusion was that random grab sampling without consideration of uncertainty may be totally inadequate for remedial decisions, although this strategy may be appealing for its low cost (Jenkins et al., 1996). Composite samples on the other hand resulted in good estimates with low standard deviation (see section 4.8).

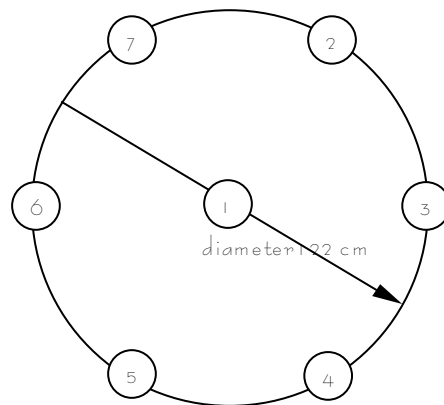


Figure 9. Illustration of sampling scheme for a short-range heterogeneity study performed by Jenkins et al. (1996). Seven discrete samples were collected at each sample location.

Most studies of spatial variability are performed at scales with sampling points ranging from meters to hundreds of meters apart. In contrast, Schumacher and Minnich (2000) studied the extreme short-range variability in VOC-contaminated soil. They found that soil properties (TCO, sand content and clay content) and contaminant concentrations generally varied between 1 and 4 times at a distance of 15 cm (vertical direction). However, extreme cases with concentration differences up to 43 times were noted. Their findings are closely related to the concept of sample support, as described in section 4.7.5.

4.9.2 Systematic sampling uncertainty

Systematic uncertainty, or bias, can be defined as the difference between the expected value of a statistic and a population parameter. The following relationship can be formulated to illustrate bias (U.S. EPA, 2000b):

$$E = \mu + b$$

where E denotes the expected value of a sample average x , μ denotes the true value of interest, and b is the bias. Similar formulations of bias are given by Chai (1996). He distinguishes between three types of bias: (1) sampling bias, (2) measurement bias, and (3) statistical bias. The statistical bias is further divided into selection bias, statistical bias in estimators, and bias in distribution assumptions.

Biased samples result from non-random sampling and from discriminatory sampling (Taylor, 1996). This happens if certain individuals in the sampled population are excluded, i.e. the rule of equiprobability is violated. Mason (1992) divides factors that cause sampling bias into two categories: *intentional* influences and *accidental* influences. One problem is that the bias may change over time. In sampling there is no such thing as a constant bias (Pitard, 1993). Even when sampling is carried out in an ideal way there will always be some sampling bias due to the particulate structure of most material. The bias may be negligible or very small but it is never strictly zero (Pitard, 1993). Chai (1996) notes that sampling bias often cannot be measured because the true value μ is typically not known.

Ramsey and Argyraki (1997) present four methods of estimating the *measurement uncertainty* at a contaminated site:

1. Single sampler/single protocol
2. Single sampler/multiple protocol
3. Multiple sampler/single protocol
4. Multiple sampler/multiple protocol

In each method the uncertainty of the estimated mean concentration at the site is calculated. In the multiple sampler methods (3 and 4), several different persons/organisations sample the site independently of each other and the variation in result is analysed. In the multiple protocol methods (2 and 4), different sampling plans (sampling pattern, number of samples etc.) are used at the site. In Methods 2-4, the *sampling bias* is estimated to varying extent, but these methods demand extensive sampling. With method 2 the bias between different sampling protocols can be estimated. In method 3 the bias introduced by the sampler is estimated. Method 4 combines method 2 and 3. Squire et al. (2000) applied the multiple sampler/single protocol approach to estimate sampling bias at a synthetic reference sampling target.

The detection limit for sampling bias varies according to the *geochemical variability* (see section 4.9.3) of the contaminant at the site (Ramsey et al., 1995). The greater the real variation in concentration, the higher the detection limit of sampling bias.

Ramsey et al. (1999) demonstrated how sampling bias can be estimated by using a reference sampling target, analogous to the use of a reference material for the estimation of analytical bias. However, it appears that their approach to sampling bias excludes bias introduced by the sampling equipment and bias inherent in the sample according to Pierre Gy's sampling theory.

Van Ee et al. (1990) present a methodology to quantify sampling bias by quality assessment (QA) samples (see section 4.9.5).

Chang et al. (1998) demonstrates how sampling bias can be included in the estimation of total sampling uncertainty through the use of kriging and co-kriging techniques at a site where anisotropy exists. However, it appears that sampling bias introduced by the sampling technique is not considered in their work.

Barcelona (1996) points out that elements of the sampling operation may cause serious errors which cannot be treated strictly by statistics. Sources of such systematic errors may be sampling devices and handling operations. To identify and control such errors, documentation of sampling procedures in protocols is suggested. Koerner (1996) provides a discussion of the effect of the sampling equipment.

Usually, the effort of reducing uncertainty in sampling is aimed at reducing the random errors. Therefore, the systematic errors in sampling are not widely known. Pitard (1993) emphasises this with the following words: “*The desire of controlling accuracy without controlling sampling correctness is certainly the worst judgement error that a person can make. It is a direct departure from logic.*” He also points out that it can be meaningless to put much effort in estimation of sampling accuracy if there is a large bias. This is illustrated by the following words: “*Many are those who are tempted to test an incorrect sampling system for its accuracy*”.

4.9.3 Geochemical variability (population variability)

A type of uncertainty that is closely related to sampling uncertainty is the large scale variability, also called *geochemical variability* (Ramsey and Argyraki, 1997), *environmental variability* (Clark et al., 1996), or the more statistically sound term *population variability* (Crépin and Johnson, 1993) (see also section 2.3). As is often the case with empirical quantities in the geologic environment, a scale problem is present. A contaminant is never evenly distributed on a site, especially in soil. In the next section some models are presented where this type of uncertainty is one separable part of the total uncertainty.

4.9.4 Models of total random uncertainty

The total uncertainty is the sum of the contributions from random uncertainty and systematic uncertainty (Taylor, 1996). Taking more samples, i.e. replicates, can reduce the random uncertainty. The systematic component of uncertainty on the other hand (see section 4.9.5), is independent of the number of replicates (Taylor, 1996). The total uncertainty includes not only sampling uncertainty but also other types of uncertainty, e.g. uncertainty in laboratory analysis (see chapter 5).

Ramsey and Argyraki (1997) use the term *measurement error* for the combined error of sampling and analyses, including both random and systematic errors (note that measurement error sometimes refers only to analytical error). Crépin and Johnson (1993) use the term *measure variability*, in which they include sampling uncertainty, handling, transport and preparation uncertainty, subsampling uncertainty, lab uncertainty and between-batch uncertainty.

A detailed formulation of the measurement variability σ_m^2 is given by van Ee et al. (1990):

$$\sigma_m^2 = \sigma_s^2 + \sigma_h^2 + \sigma_{ss}^2 + \sigma_a^2 + \sigma_b^2$$

where σ_s^2 = sampling variability,
 σ_h^2 = handling, transportation and preparation variability
 σ_{ss}^2 = subsampling variability
 σ_a^2 = analytical variability
 σ_b^2 = between batch variability

It is assumed that the data are normally distributed or that a normalising data transformation has been performed. The total variability is presented as:

$$\sigma_i^2 = \sigma_m^2 + \sigma_p^2$$

where σ_p^2 is the population variability. Similar equations to the ones above are quite common in the literature. One such formulation of the total variance is given by Ramsey and Argyraki (1997):

$$s_{total}^2 = s_{geochem}^2 + s_{samp}^2 + s_{anal}^2$$

Ramsey et al. (1995) estimated the relative importance of measurement errors (the sum of sampling and analytical errors) and geochemical variability in relation to the total variance for a few site-investigations (Figure 10).

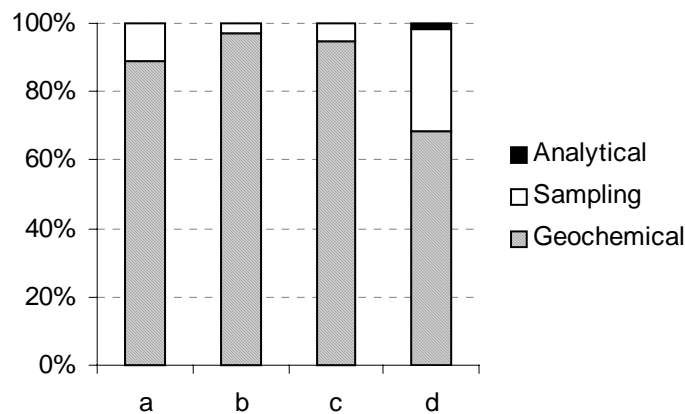


Figure 10. The relative importance of analytical variance, sampling variance, and geochemical variance in a site-investigation of lead (Pb) and copper (Cu) concentrations in soil (after Ramsey et al., 1995). The analytical variance is less than 0.1 % for cases a, b, and c.
 (a) Pb data from four sampling protocols combined.
 (b) Pb data from five-fold composite samples at each sampling point.
 (c) Pb data from single samples on a regular grid.
 (d) Cu data from four sampling protocols combined.

The geochemical variance constitutes the major part of the total variance. The relative importance of the sampling variance varies considerably whereas the analytical uncertainty constitutes a minor part of the total variance. For characterisation of contaminated soil, several other authors have come to the same conclusion, see for example Jenkins et al. (1996).

The classical statistic method ANOVA (analysis of variance) can be used to separate the sampling uncertainty, analytical uncertainty and geochemical variability, e.g. as demonstrated by Jenkins et al. (1996). ANOVA is rather sensitive to outliers in the data set. Therefore, a method less sensitive to outliers has been developed, called Robust ANOVA. The classical ANOVA method is available in statistical software packages, while Robust ANOVA is not (Ramsey, 1998).

A similar model to the one by Ramsey and Argyraki (1997) is presented by Huesemann (1994) and Barnard (1996):

$$\sigma_x^2 = \frac{\sigma_c^2}{n} + \frac{\sigma_s^2}{n \cdot m} + \frac{\sigma_a^2}{n \cdot m \cdot k}$$

where σ_x^2 is the variance of the mean concentration, σ_c^2 corresponds to the geochemical variance (including sampling variance), σ_s^2 is the subsampling variance within a single sampling core, σ_a^2 is the analytical variance, n is the number of sampling cores, m is the number of subsamples within a core, and k is the number of subsample analyses. The equation shows that if the number of analyses is large the analytical variance can be neglected. Also, if the number of subsamples is large the subsample variance can be ignored. It is most likely that the largest source of sampling uncertainty is the large-scale soil contaminant heterogeneity σ_c^2 (Huesemann, 1994).

There is an important difference between the models by Ramsey and Argyraki (1997) and Huesemann (1994). In the previous model s_{samp}^2 represents small scale heterogeneities at the sampling location while σ_s^2 in Huesemann's model represents small scale heterogeneities in the sample core.

If the sample is homogenised properly prior to analysis, the subsampling uncertainty can generally be neglected and the equation reduces to (Huesemann, 1994):

$$\sigma_x^2 = \frac{\sigma_c^2}{n} + \frac{\sigma_a^2}{n \cdot k}$$

In many cases σ_a is much smaller than σ_c . In these cases the equation reduces to this simple model (Huesemann, 1994):

$$\sigma_x^2 = \frac{\sigma_c^2}{n}$$

4.9.5 Models of total systematic uncertainty

Systematic components of uncertainty from various sources are algebraically additive (Taylor, 1996):

$$B_{total} = B_1 + B_2 + \dots + B_n$$

where B_{total} is the total systematic uncertainty, and $B_1 \dots B_n$ are the components of uncertainty from sources 1 ... n. Conceptually, sources of bias can be identified but quantifying their contribution may be difficult. Taylor (1996) suggests that a bias “budget” is developed and the bounds of each component are estimated. If possible, corrections for bias should be made.

In soil sampling studies, van Ee et al. (1990) distinguish between bias introduced in; (1) sample collection, (2) handling and preparation, (3) subsampling, and (4) the laboratory analytical process. Each of these biases can be derived from either (a) contamination or (b) some other source. The *total measurement bias* is defined as the sum of these eight (4×2) different components of bias, similar to B_{total} in the equation above. Van Ee et al. (1990) present equations and methods to quantify the different components of bias by quality assessment (QA) samples.

4.9.6 Required number of samples

In every site-investigation the question arises of how many samples to collect and how certain the result will be. The answer to these questions depends on the objectives of the investigation, the criteria that must be met (confidence levels etc.) and how the sampling strategy is designed. Here, the question of the required number of samples will be addressed for the objective to calculate mean concentration.

For economic reasons it is often necessary to limit the number of collected samples. However, there is no empirical procedure to decide the minimum sampling effort that should be performed in advance, at least not in waste material (Pitard, 1993). To determine the required number of samples a decision must be made whether or not the samples should be regarded as spatially correlated or not, i.e. if the variables are *random* or *spatial* variables. Correlation is a statistical measurement of the fact that samples taken close together are more similar in value than samples taken further apart (Flatman and Yfantis, 1996). Random variables are field samples that are further apart than the range of correlation, and spatial variables are field samples closer together than the range of correlation (Flatman and Yfantis, 1996).

Uncorrelated data

Classical statistics can be used to estimate the required number of samples if spatial correlation is ignored. The number of samples required for a given level of accuracy of the mean concentration can be expressed as (Crépin and Johnson, 1993):

$$n = \left(\frac{t \cdot s}{D} \right)^2$$

where

n = the number of needed samples.

- t = the Student's t-value for a chosen level of precision, for example a confidence level of 95 % ($\alpha=0.05$).
- s = the standard deviation. It can be known from previous investigations or, for example, estimated as $s = R/4$, where R is the estimated range of concentration likely to be encountered in sampling (approximately ± 2 standard deviations at the 95 % confidence level).
- D = the variability in mean estimation we are willing to accept (given in the same unit as s).

The equation is based on the assumption of random sampling from a normal distribution (Box et al., 1978). Crépin and Johnson (1993) give the following example of the use of the equation: A site is contaminated with PCB, with a concentration ranging from 0 to 13 ppb. A confidence level of 95 % is chosen for the estimated mean to be within 1.5 ppb of the true mean. The number of sampling points is initially assumed to be 10 and the corresponding t-value is 2.262. This gives:

$$n = \left(\frac{2.228 \cdot 3.25}{1.5} \right)^2 = 23$$

A new t-value is derived, based on $n = 23$ and the calculation is repeated. The result is $n = 21$. The calculations must be repeated (iteration) until n corresponds to the t-value used in the calculation.

Keith et al. (1996) describe a similar approach for calculating the required number of samples based on mean concentration. They also provide equations to calculate the required number of samples to reach a specified confidence level of success, where the success depends on some kind of decision rule.

Gilbert (1987) presents three approaches to determine the number of samples; (1) pre-specified variance, (2) pre-specified margin of error, and (3) pre-specified relative error. Calculation examples for each case are also given.

U.S. EPA (1994) presents equations for calculation of the required sample size for testing mean concentration versus an action level. Two cases are considered; (1) standard deviation is known, and (2) standard deviation is unknown. The approach is based on random samples taken from a normal distribution.

Some authors, for example Huesemann (1994), emphasise the importance of performing a small field study prior to determining the necessary number of field samples. The purpose of such a study is to roughly estimate the geochemical and analytical variance so that the necessary number of samples, and the number of analysis per sample to meet the specified sampling objectives, can be calculated. From equations the necessary number of soil samples can be calculated by solving for n . If only the large scale heterogeneity is considered (σ_c) the simple model presented in the previous section can be used (coefficients of variation can be used instead of standard deviations in the equations):

$$n = \left(\frac{S_c}{S_x} \right)^2$$

where S_c and S_x are coefficients of variation (relative standard deviations). From the equation follows that if the sampling objective is $S_x \leq 10\%$ and a preliminary field study has given $S_c = 40\%$, the approximate number of required samples are 16. Further iterations may be required since the value S_c may change when additional samples are taken (Huesemann, 1994).

Correlated data

Some authors have studied how correlated data affect the required number of samples. Gilbert (1987) provides equations and examples on how time and space correlated data affect the required number of samples. A conclusion is that if sample units are positively correlated, and this correlation is not taken into account, the required number of samples will be greater than if no correlation exists (the opposite is true for negative correlation). The reason for this is that each sample unit contains less information when sample units are correlated (Gilbert, 1987). On the other hand, when Webster and Burgess (1983) took the correlation into account with geostatistical techniques they found that the required number of samples was much smaller than if correlation was ignored. This is because the correlation contains information that can be handled by geostatistics.

Myers (1997) presents the concept of decline curves to determine the number of samples at a site. A decline curve is a graph showing how the statistical error (mean estimation) declines when the number of sample increases. The statistical error is plotted on the y-axis and the number of samples on the x-axis. Each sampling program has its own decline curve. Kriging can be used to develop a decline curve for a particular site and sampling program.

Chang et al. (1998) have demonstrated the use of semivariograms, kriging and cokriging techniques to determine the minimum number of samples for reliable concentration estimates. In a case study, 28 samples were determined to be a minimum. They conducted duplicate analyses on 10% of the samples and used certified reference materials to verify the analytical procedure. The kriging/cokriging approach was compared to the robust ANOVA technique (see section 4.9.4) and the conclusion was that the latter requires more sample duplicates to achieve a reliable measure of the sampling precision if spatial anisotropy exists.

4.10 Detection of hot spots

Search theory, also called *probabilistic targeting* or *geometric probability*, has been developed in the effort to discover geologic ore deposits. J.D. Savinskii, D.A. Singer, F.E. Wickman, and L.J. Drew made great contributions in the 1960-70s. Today, the literature is quite extensive. Here, only a few papers and books with relevance for detection of hot spots at contaminated sites will be mentioned.

A hot spot can be defined as “...a local area where the concentration of one or more contaminants could lead to unacceptable risks to human health or to the environment” (Ferguson, 1993). Frequently asked questions when searching for hot spots are (Myers, 1997):

- What grid spacing is needed to locate a hot spot with a specified confidence?

- For a given grid spacing, what is the probability of hitting a hot spot of a specified size?
- What is the probability that a hot spot exists when no hot spots were found?

Gilbert (1987) addresses these questions. He presents procedures to determine grid spacing, the size of a hot spot likely to be hit, the probability of not hitting a hot spot, how prior information can be taken into account, and the probability that a hot spot exists when none has been found. A number of nomographs for different cases are presented. An important design parameter for the sampling plan is the critical hot spot size, i.e. the largest hot spot that may go undetected.

Drew (1979) evaluated different drilling patterns when searching for elliptical shaped targets. Rhombic, rectangular, triangular, and square patterns were evaluated. Situations when particular patterns are more efficient than others were identified. It was found that the optimum hole spacing is a function of (1) drilling cost, (2) the value of a target, (3) the target shape, and (4) the target occurrence probability. The optimum hole spacing corresponds to the maximum of a gross drilling return (GDR) function. Different cases are considered; single target, multiple targets, and the case when target occurrence is uncertain. Also, a multistage pattern-drilling scheme was proposed, where prior target occurrence probabilities are updated using Bayes' rule.

Ferguson (1992) presents a sampling grid, the herringbone, that is demonstrated to be better than the square grid in detecting hot spots, especially elongated hot spots. He also demonstrates the effect of target shape, target size, target orientation, and the number of existing targets, on the required number of samples to locate targets. Equations and nomographs are provided for several cases. In his presentation it is assumed that all parts of a site have equal probability of containing a hot spot, i.e. prior information is not taken into account. The drawback of this assumption is that the required number of samples may become very large.

In addition to the equiprobable sampling design discussed above, Ferguson and Abbachi (1993) demonstrates how expert judgement about likely hot spot locations can be incorporated in the sampling design. One approach is to partition the site into subareas and then assign scores to each subarea, reflecting the strength of belief as to where the target is likely to be. Fewer samples is required from subareas with lower scores (lower expectation that a hot spot exist in the area) to achieve a given success probability. By a provided nomograph it is possible to determine the probability that a hot spot does not exist when no hot spot was located, given specified priori probabilities that a hot spot exists. In this way it is possible to determine if the sampling was successful or not. Ferguson and Abbachi (1993) recommend multistage sampling so that geostatistical techniques can be used to estimate contaminant concentrations at unsampled locations. In a first stage the sampling objective is to locate hot spots but a second stage aims at estimation of contaminant concentrations by kriging. A method is described to determine the required number of samples for specified estimation error.

The methodology described above has been further developed into a software (Nathanail et al., 1998; Tucker et al., 1996) described in the appendix.

UK Department of the Environment (1994) has presented a practical guide for sampling strategies, aiming at detection of hot spots. Simple equations and nomographs for estimating the required number of samples to detect a hot spot are provided.

Warrick et al. (1998) presented a probabilistic method for evaluating whether a monitoring system is capable of detecting a subsurface contaminant plume. In contrast to Drew (1979) and Gilbert (1987), prior information of contaminant release is taken into account. Warrick et al. (1998) provide a discussion on how to expand the theory to account for varying degrees of hit reliability (not simply “hit “ or “no hit”), probability of multiple hits, and how the algorithm can be incorporated in flow and transport simulation packages for calculation of detection probabilities.

Ross and Vieux (2000) developed a probabilistic methodology for determining the minimum number of monitoring points necessary to detect subsurface leakage through vertical barriers. Their method combines hydraulic signature (head distribution) assessment and search theory as presented by Singer and Wickman in 1969 and Gilbert (1987).

A similar concept to hot spot detection is anomaly characterisation in the field of geotechnics, as described by Halim and Tang (1993). They applied search theory and Bayesian updating to the problem of finding a geologic anomaly.

4.11 Delineation of contaminated areas

Delineation of contaminated areas, e.g. hot spots or areas with concentration exceeding a threshold level, is basically the same as creating a simple contour map. A practical rule for such a case is that geostatistical methods should be considered. Flatman and Yfantis (1996) discuss different aspects of this situation such as sampling grids, grid orientation etc. Several authors have applied kriging techniques to delineate a contaminated area. One such example is presented by Juang and Lee (1998), who applied simple indicator kriging to delineate metal contaminated areas of a paddy field. Probabilities of false positive and false negative classification (see section 5.4) were obtained.

Squire et al. (2000) used some very simple methods for delineation of a hot spot; triangulation and linear interpolation.

5 POST-SAMPLING UNCERTAINTY

5.1 Erroneous samples

Crépin and Johnson (1993) list a number of post-sampling activities that can introduce errors. Other authors also mention the following error sources:

- handling
- transport
- preparation
- subsampling
- lab analysis
- human error

Several of these types of errors are generally systematic in nature (see section 4.9.2). Often, the errors can be assessed by use of different types of quality assessment (QA) and quality control (QC) samples. They include sample duplicates, replicates, spikes, blanks, and splits (Shefsky, 1997). Lewis (1996) identifies two major groups of blanks; *field blanks* and *laboratory blanks*. Field blanks are used to assess the errors (contamination) introduced during sample collection, storage, and transport (other definitions exist). Laboratory blanks are used to provide information of contamination occurring in the laboratory. Van Ee et al. (1990), Lewis (1996), and Black (1996) describe different types of blanks and their use. Lewis (1996) also provides some equations for assessing the effectiveness of blanks. Van Ee et al. (1990) present systematic ways for quantifying bias in handling, transport etc. by using QA samples.

5.2 Uncertainty in laboratory analyses plans

A number of uncertainties are involved when a sampling activity is planned. Most of these uncertainties are discussed elsewhere in this chapter. Here, two important uncertainties related to the planning of laboratory analyses will be described briefly. The first one is whether or not the relevant chemical substances are analysed at all. The risk assessment of a site is restricted to the substances that has been analysed, but there may well be other substances of great concern present. Few authors have addressed this uncertainty specifically. This question, and the one mentioned below, is closely related to the uncertainty in risk assessment. However, the risk assessment issue is not addressed in this report.

A second uncertainty associated with laboratory analyses can be illustrated by the following question: Is the selected analysis method the right type of analysis for the contaminant? Even if a standardised analysis method is used, the result may not be applicable, e.g. for risk assessment purposes, because guideline values may have been developed based on a different analysis method. An example of this uncertainty is whether or not the result from a certain leaching test is applicable on a particular contaminated soil problem.

5.3 Field and laboratory analyses

Traditionally, the uncertainty in analytical techniques has been the focus of research. However, as have been previously described, the analytical error constitutes only one part of the overall error.

Ramsey and Argyraki (1997) distinguish two ways of estimating the uncertainty in chemical analysis: The “bottom up” approach and the “top down” approach. Both approaches have their advantages and drawbacks. In the “bottom up” approach the random error from each procedural step of a laboratory method is quantified separately as a standard deviation (s_i), and the corresponding variance (s_i^2) is calculated. The overall analytical uncertainty (s_a^2) is quantified by summing the variances for all procedural steps of the method (Huesemann, 1994):

$$s_a^2 = \sum_{i=1}^n s_i^2$$

where n is the number of procedures in the analysis method. Generally, the analytical procedures is weighing, extraction, instrumental analysis etc. For most environmental analyses the analytic uncertainty is around or less than 10 % (coefficient of variation) but analytical methods with numerous procedural steps results in larger values of s_a (Huesemann, 1994). Generally, the standard deviation increases drastically when the concentration approaches the detection limit.

There are developed methods for estimating analytical precision (random error) and bias (systematic error). The precision is estimated with duplicate analyses of a sample and the bias by use of certified reference materials (Ramsey et al., 1995).

The “bottom up” approach is used by individual laboratories to estimate the uncertainty of a laboratory method but it tends to give over-optimistic estimates of the uncertainty (Ramsey and Argyraki, 1997).

The “top down” approach uses inter laboratory trials to estimate the uncertainty of a measurement. The same field sample is analysed by a number of selected laboratories ($n > 8$) by the same analytical method. The scatter results from all the laboratories are used as an estimate of the overall analytical uncertainty (Ramsey and Argyraki, 1997).

As an alternative, one can use the Horwitz equation to estimate the analytical variance, which is an empirically given equation (Albert and Horwitz, 1996; Clark et al., 1996):

$$CV(\%) = 2^{(1-0.5 \log C)} \approx 2C^{(-0.1505)}$$

where $CV(\%)$ is the among laboratory coefficient of variation in percent and C is the concentration (mass/mass). The intention of the Horwitz equation is to estimate the among laboratory variance for multi-laboratory studies. However, it can be expected that within laboratory variance usually is smaller than among-laboratory variance. Therefore, the Horwitz equation offers a high-end estimate of within laboratory variance for a particular analytical method (after conversion from coefficient of variation to variance). The Horwitz equation is particularly valuable in early stages of a project, before

sampling has been conducted, since it offers an estimate of analytical variance when no data exists (Clark et al., 1996).

In contrast to sampling uncertainty, analytical uncertainty is not site specific, i.e. analytical uncertainty estimates can be applied to any site where that particular analytical method is being used. A special problem of analytical uncertainty evolves when the measured value is close to the detection limit of the analysis method. Taylor (1996) states that the limit of quantification is about 3 times the limit of detection, which makes decision-making foolish if it is based on data obtained at the limits of capability of the methodology.

5.4 Data evaluation

Additional errors can be introduced when data are analysed and evaluated. Van Ee et al. (1990) identify a problem when data are analysed: whether a transformation of data is needed to stabilise the variance or not. The dependence of measurement error variances on sample concentration is frequently encountered (van Ee et al., 1990).

Myers (1997) describes two types of errors in data evaluation; *estimation error* and *misclassification error*. He illustrates the errors with a simple example: Consider a block area where the true concentration is 22 ppm and where a threshold for remediation of 25 ppm has been determined. If the concentration is estimated to be 29 ppm the *estimation error* equals 7 ppm. In this case there is also a *misclassification error* since the estimated concentration exceeds the threshold concentration while the true concentration is below the threshold. With the baseline condition as a null hypothesis, two types of misclassification errors can occur (Myers, 1997; U.S. EPA, 1994):

1. Type I Error: False rejection decision error (false positive or overestimation)
2. Type II Error: False acceptance decision error (false negative or underestimation)

In the example above there is a Type I Error because the concentration is falsely estimated to exceed the threshold. If the true concentration is above the threshold while the estimated concentration is below there would be a Type II Error. Type I Errors lead to excessive remediation costs while Type II Errors can bring human health and ecological risks. Therefore, Type II Errors are sometimes called *Consumer's Risk* (Gilbert, 1987). The different decision errors can be illustrated by the *decision space* in Figure 11.

Of course, other uncertainties exist when data are evaluated and decisions are made. However, it is beyond the scope of this report to discuss them all. One example of such uncertainties is the uncertainty associated with guideline values used for risk assessment.

In cases where a strict threshold level is used for classification purposes, a small estimation error can lead to misclassification (Myers, 1997), resulting in increased remediation cost or environmental risk.

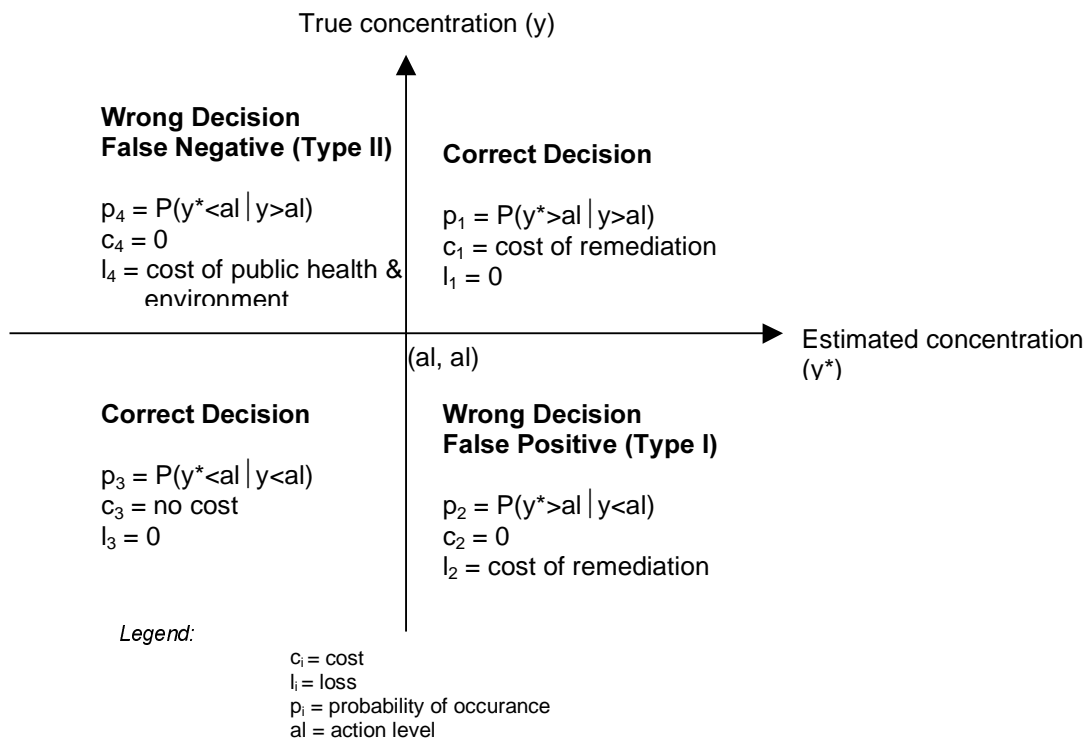


Figure 11. The decision space for classification of contaminated land (after Flatman and Englund, 1991).

6 DATA WORTH ANALYSIS AND COST EFFECTIVENESS

6.1 Introduction

During a data collection program there are two primary issues of concern (James and Gorelick, 1994):

1. where to take the next sample, and
2. when to cease sampling (a stopping rule).

Closely connected to the latter question is the concept of *data worth*, i.e. the worth of additional information. How the concept of data worth is defined depends on the philosophy of the decision-maker. In a decision analysis framework, the purpose is often to find a cost-efficient solution to a problem (e.g. to maximise an objective function). In a remediation project, data collection such as sampling, will continue until a *stopping rule* terminates the sampling. Different decision frameworks will have different stopping rules, and therefore the worth of data can be defined differently. One common definition of data worth is that it equals the reduction in project cost (or risk) that the new data will result in. Another definition is that additional data only has value if it can alter the choice of decision to make (Hammitt, 1995).

In section 6.4 some of the most common or most interesting frameworks will be presented. Many are based on sequential sampling or adaptive sampling, i.e. the sampling is performed in stages. This may require several visits to the contaminated site for the sampling team.

6.2 Decision theory

In *decision theory* there is a key term for the worth of additional data: the *Expected Value of Information* (EVI). The value of information is an integral part of any study of decision-making (Heger and White, 1997). Another term for EVI in sampling problems is EVSI, the expected value of sample information (Dakins et al., 1995). The EVI depends on the set of alternative decisions that are considered, and how the payoff depends on the decision and the uncertain parameters. EVI can be described as the difference between the expected payoff if one selects the optimal decision based on posterior information, and the expected payoff for the optimal decision based on prior information (Hammitt and Shlyakhter, 1999). In many situations the EVI is an increasing function of prior uncertainty, i.e. additional data are worth more at the early stages of a project when uncertainty is large. This may be intuitively suspected but because of the complexity of the dependence of EVI by several factors, there is no general relationship between uncertainty and EVI (Hammitt and Shlyakhter, 1999). However, Hammitt (1995) emphasises that an overly narrow prior distribution (overconfidence) may reduce the assessed value of information.

Of importance in environmental risk assessments is the probability of a surprise, e.g. that an unsuspected chemical at a site is the most important from a risk perspective. Hammitt and Shlyakhter (1999) states that the EVI is likely to be more sensitive to the probability of surprising outcomes than is the optimal decision under uncertainty. This suggests that thoughtful analysis of potential surprises is beneficial. Hammitt and Shlyakhter (1999) discuss how the prior probability distributions influence EVI. They conclude that the probability for surprise is often underestimated.

Another important concept in decision analysis is the *Expected Value of Perfect Information*, EVPI (Morgan and Henrion, 1990). It is the same as the EVI for new and perfect information that reduces the posterior uncertainty to zero. In other words, the EVPI is the upper bound for EVI. For sampling EVPI is equivalent to EVI when perfect sampling is carried out, removing all uncertainty. Therefore, EVPI is equal to the maximum justifiable exploration budget (James et al., 1996b). Other common terms for EVPI are *Expected Opportunity Loss* (EOL) and *expected regret*. James et al. (1996b) describe EVPI with the concept of expected regret.

Generally, better decisions will be made when uncertainty is taken into account. The question is how much better the decisions will be if uncertainty is considered. How much better of you will be by including uncertainty, can be quantified and the concept for this is the *Expected Value of Including Uncertainty*, EVIU. The EVIU is defined as the expected difference in value of a decision based on probabilistic analysis and a decision that ignores uncertainty (Morgan and Henrion, 1990).

In practice, decision theory is often applied to single objective problems. However, there are ways of applying the theory also for multiple objectives. Haimes and Hall (1974) applied multi-objective decision theory to a water resource problem with two decision variables and three objective functions.

6.3 Tools for data worth analysis

6.3.1 Payoff tables

A payoff table is a table showing the economic outcomes for different decision alternatives or events. Freeze et al. (1992) demonstrate the use of payoff tables and how they are used to structure information for data worth analysis.

6.3.2 Decision trees and event trees

The most commonly used tool for data worth analysis is the decision tree (Heger and White, 1997). Decision Trees have been used for a long time and most risk analysts are familiar with them. Freeze et al. (1992) use decision trees to analyse worth of additional data in the risk-based framework described below. Most decision analysis texts include material about decision trees and their application. Decision trees are horizontal structures, which proceed with time from left to right. In a decision tree, a node represent a decision (decision node), an uncertain event (chance node), or an outcome (terminal node) (Treeage Software, 1999). A decision tree without decision nodes is called an event tree. An example of an event tree is given in Figure 12.

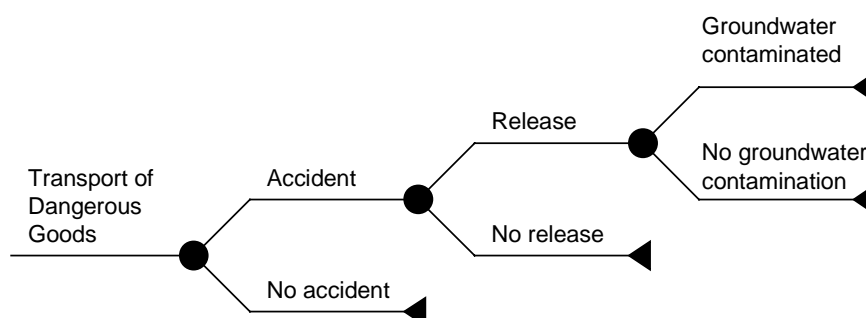


Figure 12. Example of an event tree for transport of dangerous goods.

6.3.3 Influence diagrams

A tool with growing applications in decision analysis is the influence diagram, also called relevance diagram. Heger and White (1997) argue that influence diagrams are more effective than decision trees for data worth analysis. There are two main advantages of influence diagrams compared to decision trees:

1. The size of the influence diagram is a linear function of the number of variables, whereas the size of the decision tree is an exponential function.
2. Influence diagrams provide a better visual representation of the relationship among variables.

Another advantage is the ease of understanding for the decision-maker even for problems with many variables, in contrast to decision trees that easily become difficult to understand. Other authors have also compared decision trees and influence diagrams, for example Call and Miller (Heger and White, 1997). They also describe the Decision Programming Language (DPL) as a tool that captures the useful features of both decision trees and influence diagrams.

An influence diagram consists of chance nodes (ovals), decision nodes (rectangles), value nodes (diamonds), and sometimes deterministic nodes. Typical functions assigned to value nodes include cost-benefit-risk or simple cost functions (Heger and White, 1997). The nodes of the influence diagram are inter-connected by arcs. Two types of arcs exist; information arcs and conditional arcs. Together they show the flow of information and uncertainty through the decision-making process. Figure 13 shows an example of an influence diagram.

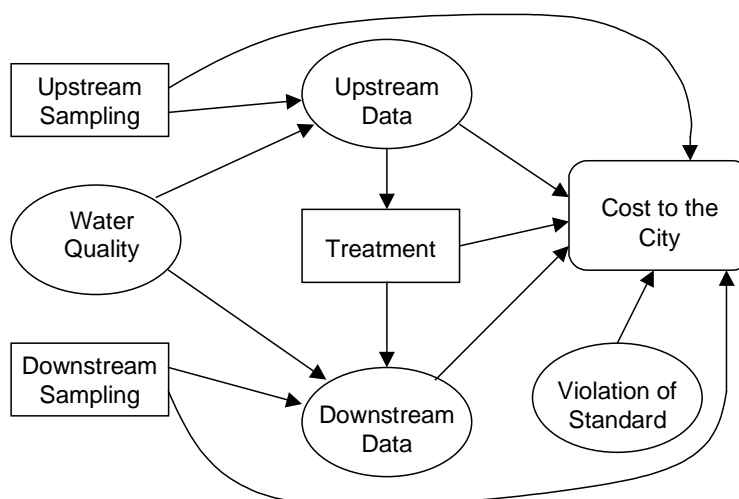


Figure 13. Example of an influence diagram illustrating the monitor-and-treat decision problem. Sampling is performed upstream and downstream of a discharge point in a river to study arsenic concentrations regarding sampling of river water (after Heger and White, 1997).

Attoh-Okine (1998) explains the influence diagram tool for application on contaminated land problems. Heger and White (1997) used an influence diagram to calculate the worth of historical data of water quality in a river (Figure 13). However, no calculation of the worth of additional, uncollected data was performed.

6.3.4 Expert systems

Expert systems have only been used to a limited extent for data worth and decision analysis in problems related to contaminated land. The three main paradigms for expert systems are; rule-based systems, neural networks, and Bayesian networks (Jensen, 1998). A short presentation of these will be given based on Jensen (1998).

Rule-based systems try to model the expert's way of reasoning by means of a set of rules. In rule-based systems the uncertainty is often treated by fuzzy logic (see section 2.5.3), certainty factors etc.

A *neural network* consists of one layer of input nodes, one layer of output nodes, and normally 1-2 hidden layers. The network performs pattern recognition based on training results from known input and output values. It is not possible to read the uncertainty in the conclusion from a neural network. As an example, neural networks have been used for soil geology interpolation (Kumar et al., 2000) and estimation of hydrogeological parameters (Mukhopadhyay, 1999).

A *Bayesian network* consists of nodes and arcs connecting the nodes. The arcs reflect cause-effect relationship and the strength of an effect is given as a probability. Bayesian networks can be updated when additional information becomes available. Other names for Bayesian networks are Bayes nets, belief networks, Bayesian belief networks (BBNs), causal probabilistic networks (CPNs), or causal networks. Similar to influence diagrams, they are useful in showing the structure of a decision problem. In fact, a Bayesian network is equivalent to an influence diagram consisting of only chance nodes. So far, Bayesian networks have not been much applied to contaminated land problems. Software for Bayesian networks is presented in the appendix.

6.4 Strategies for data worth analysis

6.4.1 Traditional approach: Fixed cost or fixed uncertainty

Two strategies have traditionally been used for sampling; (1) to minimise the sampling cost for a specified level of accuracy (usually variance), or (2) to minimise uncertainty for a given sampling budget. These are also the strategies that Gilbert (1987) and Huesemann (1994) mention as the most commonly used in practice. In these strategies additional sampling have worth until uncertainty has been reduced to a specified level, or until there is no more money available for sampling. The first strategy often originates from some kind of regulatory framework where the cost-efficiency is not the primary goal.

Gilbert (1987) and Borgman et al. (1996a) provide cost functions that can be applied for both strategies when the population mean should be determined. Cost functions are presented for a number of sampling methods described below.

Stratified random sampling

The cost function presented is applicable when the main objective is to estimate the overall population mean (all strata included), or population total (such as total amount of a contaminant). With the cost function the optimum number of samples in each strata can be calculated. Equations for calculating the total number of samples are presented for (1) prespecified fixed cost, (2) prespecified variance, and (3) prespecified margin of error (Gilbert, 1987). Both Gilbert (1987) and Borgman et al. (1996a) provide examples of the calculations.

Two-stage sampling

In two-stage sampling the target population is divided into primary units (first stage) and the primary units into sub-units (second stage). An example of two-stage sampling is when soil samples are divided into subsamples (aliquots) chosen for measurement. Gilbert (1987) provides a cost function and equations to calculate the optimal number of subsamples.

Composite sampling

Gilbert (1987) presents a cost function for composite sampling. The number of batches, composites in each batch, and subsamples from each composite has to be selected (a batch is a group of population units). The idea is to select values for the number of batches, composites and subsamples that will minimise the variance (or the cost if the variance is prespecified).

Garner et al. (1996) present a number of models for cost-effective composite sampling using the concept of relative cost factors. One of the models is a multivariate model where more than one contaminant can be detected.

Double sampling

Double sampling can be used when two different techniques are available for measuring the contaminant concentration. An example is when one field screening technique and one laboratory analysis method is used in combination. For a cost-efficient solution only a small number of samples are analysed with both techniques but many samples are analysed with the cheaper of the two methods (note that this cost-efficiency is different from the cost-efficiency concept in a decision analysis framework). Gilbert (1987) presents a cost function and some equations to determine the optimum number of samples for each of the two measurement techniques. He also presents a case study.

6.4.2 Misclassification cost and loss functions

To classify an area as “contaminated” when it in reality is not, is a *misclassification*. The opposite situation is also a misclassification, i.e. a “clean” area is classified as “contaminated”. As described in section 5.4 these errors are generally called *false positive* (overestimation or overclassification) and *false negative* (underestimation or underclassification) respectively. Each misclassification that is made leads to some kind of cost. Aspie and Barnes (1990) present a simple cost function for the total cost of misclassification ($\$_c$):

$$\$_c = \$_o \cdot A_o + \$_u \cdot A_u$$

where S_o and S_u are the costs of overclassification and underclassification respectively (per unit area). A_o and A_u represent the total area that is overclassified and underclassified respectively.

Myers (1997) presents a cost optimisation graph (Figure 14) for optimising the number of samples to be taken. A large error in estimated concentration leads to a large misclassification cost. This often occurs when only few samples are taken, i.e. the sampling cost is low. When more samples are taken the sampling cost increases but the misclassification cost decreases. The cost of false positive classification can quite precisely be estimated in terms of engineering cost (Myers, 1997). An investment in sampling will probably reduce this cost. The optimum number of samples occurs where the total cost is at a minimum, i.e. at the cost optimum. At the cost optimum, marginal cost equals marginal benefit.

The loss function in Figure 14 is a simple and symmetric one. It only considers false positive decision errors (overestimation). Flatman and Englund (1991) present the symmetrical least-squares loss function, which incorporates both false positive and false negative classification errors. With this function the loss (cost) increases with the square of the size of the error. Usually, asymmetric loss functions are more realistic than symmetrical ones. Flatman and Englund (1991) as well as Myers (1997) present asymmetric loss functions that also incorporate false negative errors (underestimation). One is the *step function*. If there is a false positive decision there will be a fixed cost of remediation and if there is a false negative error a fixed estimated cost due to health effects will appear.

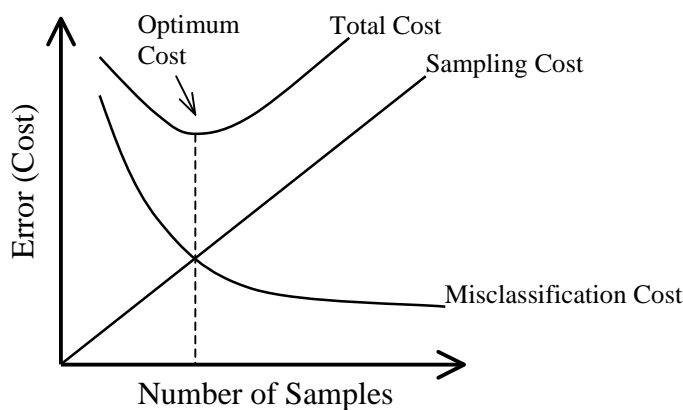


Figure 14. A cost optimisation graph for sampling. After J-M Rendu in Myers (1997).

An improvement of the step function is the *V curve loss function* (Figure 15). In this loss function the costs are not fixed, instead they increase when the decision error increases. The positive portion of the graph reflects false negative errors, bringing health effect costs. The negative part of the graph corresponds to remediation costs due to false positive errors. Myers (1997) also presents a *hybrid step-V curve*, where the false positive errors result in a fixed remediation cost, while the right hand side of the graph is identical with Figure 15.

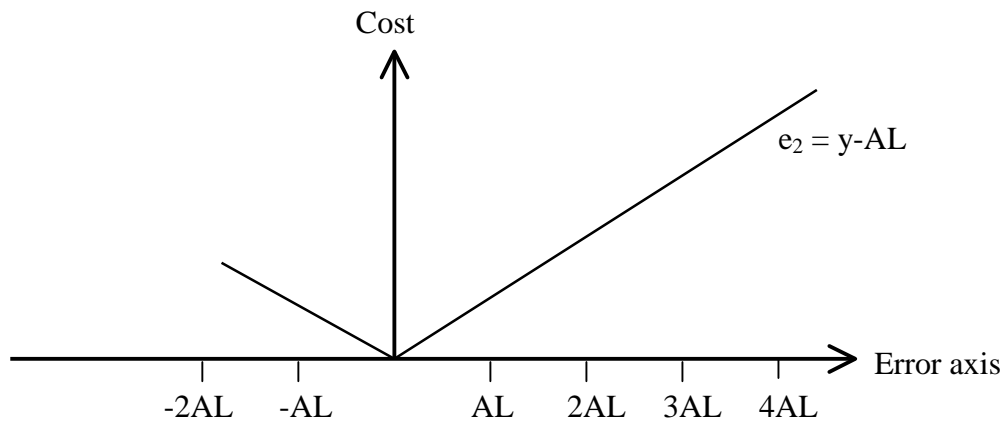


Figure 15. The V curve loss function. Error is defined as the difference between the true concentration and the action level (AL) for all incorrect decisions. The expected loss is a function of concentration (after Flatman and Englund, 1991).

Flatman and Englund (1991) argue that the loss function must be defined along with other goals and objectives before a sampling design is determined.

Thompson and Fearn (1996) present cost and loss functions similar to the ones above. The loss function expresses the expected loss (cost) as a function of uncertainty (variance), and cost of sampling and analysis. It expresses how the loss to an end-user increases with the magnitude of the measurement error in some way (the error is defined as the difference between the estimated and the true concentration). The true concentration is usually not known but the optimal division of resources between sampling and analysis can still be found by minimising the loss function. For simple loss functions minimisation can be made algebraically but this may not be possible for more complicated functions. Thompson and Fearn (1996) present solutions for some simple symmetrical loss functions as the quadratic loss function, the V curve loss function, and an empirical asymmetrical step function. For practical use the cost function must be designed for the particular problem and for example penalty costs could be incorporated.

The described loss functions consider misclassification of the whole area and do not address local misclassification. A *probability-based cost function* proposed by Aspie and Barnes (1990) manages this. It can be derived from the simple cost function presented in the beginning of this section, based on concept of *expected cost-of-errors*. Aspie and Barnes (1990) present an infill-sampling strategy for situations when a set of samples has been taken (known location and known value) and another sampling round is planned. The optimal locations are determined out of a set of possible new sample locations. The procedure is based on the *expected cost-of-errors* criterion using simple kriging. The probabilities of a false positive or false negative can be obtained for each point in a block model by using ordinary, simple, indicator, or probability kriging (Myers, 1997).

6.4.3 Uncertainty-based stopping rules

In this section a selection of sampling approaches with uncertainty-based stopping rules will be presented. These stopping rules do not take sampling cost, other costs or benefits, or the economical worth of data into account specifically. However, it appears that these approaches can be slightly modified to incorporate the aspects of cost and data worth.

Johnson (1996) has demonstrated an *adaptive sampling* strategy, i.e. the sampling is performed in subsequent stages. It is based on the objective to minimise uncertainty about the extension of contamination and uses a Bayesian/geostatistical methodology. An initial conceptual model based on soft information is used when sampling begins. Hard and soft data are combined and updating is performed by indicator kriging. New sample locations are selected from a set of potential sampling points based on the objective of maximising the area classified as clean at the 80 % certainty level. Data worth are not considered in the approach but could probably be incorporated relatively easily. Johnson (1996) states that his method for handling uncertainty “...leads naturally to measures of benefit one might expect from additional data collection”.

Chiueh et al. (1997) present a decision support system for probability analysis in soil contamination. It is based on a database, a model with an indicator kriging algorithm and a Geographic Information System (GIS). The whole study site is divided into blocks in a square grid. The user must specify a threshold level of contaminant concentration. Also, probabilistic criteria must be specified for when to remediate and when to take additional samples. These are based on the probability of exceeding the threshold level, and the probability of a false positive or negative classification. If the criteria are not fulfilled, additional sampling may have to be performed. The system has no cost function integrated, so cost calculations have to be made by the user.

Robbat (1996) has developed a strategy for adaptive sampling based on the DQO process. Field analysis techniques are proposed for cost-effectiveness. Several sampling rounds are made until confidence in the conceptual model is obtained. An illustrative flow chart of the process is presented. However, the stopping rule for sampling is expressed rather vague: “*Does quantitative analysis verify site screening?*” It is proposed that sampling should be directed by geostatistical techniques. Costs for different field analysis techniques are presented in an appendix.

Stenback and Kjartanson (2000) present another geostatistical approach for adaptive sampling. Samples are located to regions where there is high uncertainty regarding if the concentration exceeds a specified threshold level. Samples are spaced according to the *range* of a variogram calculated from sample data. A probability contour bounds the contaminated area and the sample collection ceases when there is low probability for the area outside the boundary to exceed the threshold. The probability map could be developed by indicator kriging, ordinary kriging, or a less well-known third method (Stenback and Kjartanson, 2000).

6.4.4 Bayesian decision analysis

Early frameworks

Several early frameworks for evaluation of data worth were reported during the 1970's. Davis and Dvoranchik (1971) demonstrated a framework to determine the value of ad-

ditional information about annual peak flow in a stream for a bridge construction problem. They applied the concept of expected opportunity loss (EOL) to calculate the worth of data. Gates and Kisiel (1974) applied the EOL concept on a groundwater model. Error in model prediction was coupled to a loss function and reduced by more information.

Maddock (1973) presented a management model for an irrigated farm with the purpose to maximise the expected profit of the farm. The analysis results in a decision on cropping and water pumping patterns over a design period, a choice of groundwater model, the ranking of data worth for different types of data, and the ranking of priority of further data collection. The concept of *expected regret* is used to measure the profit loss due to non-optimal parameter values in the model. Only direct monetary costs are considered in the model, not cost of failure since no failure criterion is defined.

Grosser and Goodman (1985) used loss functions and Bayesian decision theory to determine the optimal sampling frequency for chloride in a public water supply well. Three different loss functions were used for different chloride concentrations. The optimum number of samples was based on calculations of expected loss.

Ben-Zvi et al. (1988) demonstrated how preposterior analysis in a risk-based Bayesian framework can be used as a tool to assess the value of data before they become available. The methodology was applied to a problem of aquifer contamination. Three different states of contaminant intrusion into the aquifer were assumed to be possible and three different courses of action (decisions) was considered. Since decisions taken under uncertainty may lead to losses, loss functions are discussed. They define risk as “*expected loss*” or as “*the average loss of a consistent decision-maker*”. The value of additional information (data worth) is defined as the expected reduction in risk, which is equal to the difference between the prior risk and the preposterior risk. The preposterior risk is calculated as a weight average of the minimal posterior risks, where the weights are the prior probabilities for the different states.

The value of hydrogeological information for risk-based remedial action decisions was studied by Reichard and Evans (1989). They defined the total social cost (TSC_i) of an action a_i as:

$$TSC_i = C_i + C_r(1 - e_i)R_u$$

where

C_i = economic cost of remedial action a_i

C_r = cost per unit risk (assumed constant)

e_i = efficiency of action a_i in reducing risk ($0 \leq e_i \leq 1$)

R_u = uncontrolled risk level ($R_u \geq 0$)

The concept of expected opportunity loss (EOL) is used to calculate the value of groundwater monitoring. The computation is performed in a general three-step framework.

Freeze et al.

The framework presented by Freeze et al. (1992; 1990) is a framework for decision analysis. It is based on an objective function Φ of the form (Freeze et al., 1990):

$$\Phi = \sum_{t=0}^T \frac{1}{1+i} [B(t) - C(t) - R(t)]$$

where $B(t)$ = the *benefits* in year t , $C(t)$ = the *cost* in year t , and $R(T)$ = the *risk* in year t . All of these parameters are expressed in monetary terms. The optimal decision from a risk perspective is achieved when the objective function is maximised. At this optimum the risk level is called optimal risk. The risk is defined as the *probability of failure* multiplied with the *cost of failure*. The *failure criterion* has to be defined in the individual case. Note that benefits, cost, and cost of failure, are all economic terms. All other information, such as sample data, geological information, engineering considerations etc., are combined into one single term: the probability of failure.

In this framework, an additional measurement has worth only if the risk reduction it provides exceeds the cost of obtaining it. Each new sample or measurement brings an additional cost to the site-investigation, but it also provides additional information so that the probability of failure is reduced. The question is how much new information an additional measurement provides in relation to the additional cost. Freeze et al. (1992) define data worth in a precise way as “...*the increase in the expected value of the objective function between the prior analysis and the preposterior analysis due to the availability of the proposed additional measurements*”. The prior and preposterior analyses refer to the different stages of Bayesian updating (see section 2.4.2).

In the framework, data worth analysis is a tool for the rational design of a field investigation program. Its purpose is to reduce (1) the uncertainty about the natural system, (2) the cost of the site-investigation programme and (3) the associated risks. According to Freeze et al. (1992) data worth analysis can be used in two ways: (a) to compare alternative data-collection programmes, or (b) to be used as a stopping rule to decide when no further measurements should be made.

Freeze et al.(1992) applied the concept of data worth on a hypothetical hydrogeological problem. Some of the methods they propose to determine the probability of failure includes search theory, stochastic simulation, and indicator kriging. They also demonstrate the concept of *regret*. Regret is defined as the price that must be paid for selecting the non-optimal alternative given perfect knowledge.

James and Freeze

James and Freeze (1993) developed a Bayesian decision framework for addressing questions of hydrogeological data worth. The framework has much in common with the frameworks based on optimisation (see section 6.4.5) but uses an objective function identical to the one presented by Freeze et al. (1990). An indicator simulation algorithm incorporates hard and soft data regarding the continuity of an aquitard. In a preposterior analysis the *expected value of sample information* (EVSI) and the *expected value of perfect information* (EVPI) is evaluated. James and Freeze (1993) suggest that the best sample location is not necessary the point of greatest uncertainty. The best sample location may also depend on both uncertainty and on the decision being made.

Dakins et al.

Dakins et al. (1994) presented a decision framework for remediation of PCB-contaminated sediments. Instead of a fixed failure cost they use a loss function, where

the loss (cost) is a function of the difference between the true contaminated area and the remediated area. Their analysis includes the *expected value of including uncertainty* (EVIU) and EVPI. In a later paper Dakins et al. (1995) extended their analysis to include EVSI. They utilised Bayesian Monte Carlo methods to calculate EVSI in a pre-posterior analysis. A conclusion was that the calculated EVSI should be regarded as an upper bound due to a number of simplifying assumptions in the analysis.

James and Gorelick

James and Gorelick (1994) presented a Bayesian data worth framework for aquifer remediation programs. Bayesian decision analysis is used to handle contaminant transport in a spatially heterogeneous environment. The framework consists of three modules: (1) A module with numerical modelling and *Monte Carlo simulation* to predict the probable location of the contaminant plume, (2) a module for estimation of remediation cost based on *capture zone theory*, and (3) a geostatistical module with *indicator kriging* to combine prior information with sample information. The framework evaluates the monetary worth of spatially correlated samples (at observation wells) when delineating a contaminant plume. Measurements are made one by one in a stepwise manner and the optimal number of measurements is estimated, as well as the best location for the next sample. Prior, preposterior and posterior analyses are performed (Figure 16). Collection of data ceases when the cost of acquisition is greater than the reduction in remediation cost the next measurement would bring. The evaluation of data worth is performed during the preposterior analysis.

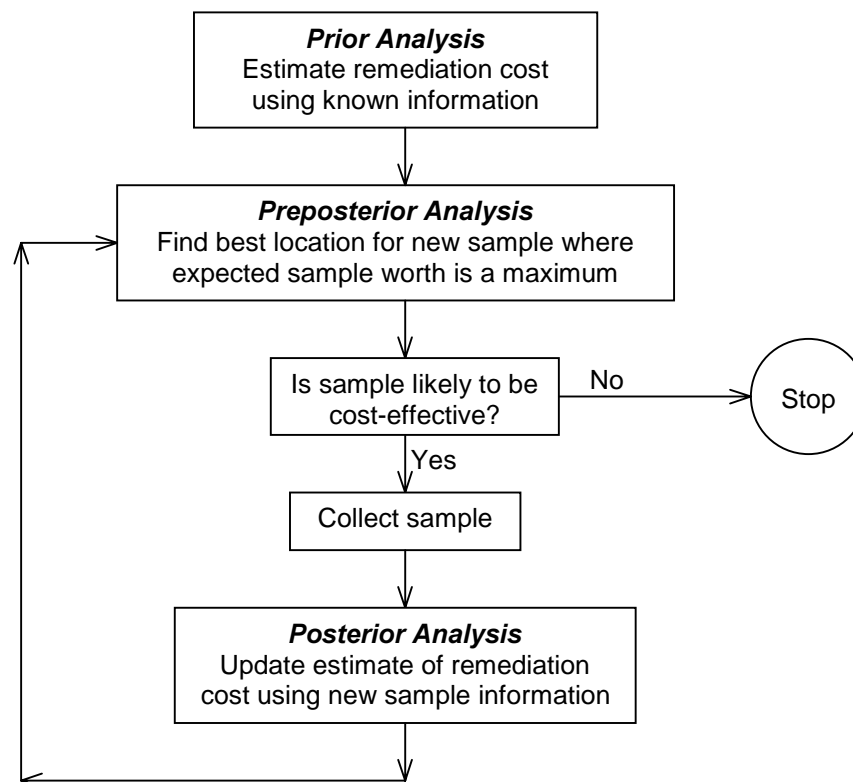


Figure 16. Outline of the steps in the data worth framework using prior, preposterior, and posterior analyses (after James and Gorelick, 1994).

An interesting conclusion made by James and Gorelick (1994) is that even if the worth of a single sample is less than its unit cost (which would indicate that the sample should not be taken), it may still be worthwhile to collect it. The information in the single sample may not be enough but combined with additional samples it may still be cost-effective to collect the sample.

James et al.

James et al. (1996a; 1996b) present risk-based decision analysis frameworks for remediation decisions of contaminated soil and groundwater. These are stripped-down approaches of the framework presented by Freeze et al. (1992; 1990). They illustrate how EVPI can be calculated based on the concept of expected regret. The quality of a sampling program is described by the *sample reliability*. The reliability ranges from 0 for measurements of no value to 1 for perfect sampling. The sample worth is calculated by a simplified data worth equation (assuming no false detections of failure):

$$\text{Sample worth} = \text{Sample reliability} \times \text{EVPI}$$

It is pointed out that there will be significant uncertainty in the calculated sample worth but that good estimates of the cost-effectiveness of a sampling program still can be made.

BUDA

Abbaspour et al. (1996) present a data worth model under the acronym BUDA, Bayesian Uncertainty Development Algorithm. The purpose of the data worth model is to analyse alternative sampling schemes in projects where decisions are made under uncertainty. The procedure in BUDA begins with problem definition, conceptualisation, definition of a goal function, definition of spaces etc. Four different problem spaces are identified: (1) the space of alternative decisions, (2) the space of states of nature, (3) the space of potential experiments to be performed, and (4) the space of samples (number and location). Uncertainty is divided into natural and informational uncertainty. Propagation of uncertainty is performed by Latin Hypercube and Monte Carlo methods. The failure criterion must be defined individually for each project. The EVSI is evaluated for each sampling scheme. In addition to EVSI, “the highest return value” of additional samples is analysed, i.e. the number of samples that maximises the return is identified. This number of samples should be collected in the first sample round. The data worth model has been applied to a landfill leachate plume (Abbaspour et al., 1998).

Smart Sampling

The Smart Sampling approach is based on the risk-based decision framework presented by Freeze et al. (1990). The basis is an economic objective function (Kaplan, 1998):

$$\text{Total Cost} = \text{Characterisation Cost} + \text{Treatment Cost} + \text{Failure Cost}$$

In principle, this function is identical to the objective function presented by Freeze et al. (1990). The expected failure cost is the probability of failure multiplied by the cost of failure. The Smart Sampling approach requires explicit decision rules, such as whether or not a piece of ground is contaminated or uncontaminated. In this example the failures are identical to the type I and type II errors described in section 5.4:

Type I error: Removal of uncontaminated soil, if not required.

Type II error: Failure to remove contaminated soil, if required.

Both types of failures involve a cost that is taken into account by the objective function.

The Smart Sampling process is designed to be iterative. This means that samples are taken few at a time, the information is analysed, the objective function updated, and a decision to take more samples is based on predictions of the worth of additional samples towards minimising the total cost of the remediation (Kaplan, 1998). Since spatial correlation of contaminant concentrations usually exists between sample locations at a site (Flatman and Yfantis, 1996), geostatistics is used as a tool. The distribution of contaminant concentrations is studied in a histogram and the spatial correlations in a variogram. Before additional samples are taken a *preposterior analysis* is performed. This is carried out by geostatistical simulation, typically kriging. The purpose of the simulation is to calculate how much the probability of failure will decrease when the additional information from sampling will be supplied. Since the probability of failure will change, so will the value of the objective function. This makes it possible to calculate the worth of additional sampling before it has been undertaken. When additional sampling has been performed a new analysis will be carried out. This iterative procedure continues until additional samples will no longer reduce the objective function.

The Smart Sampling approach, as presented by Kaplan (1998), develops graphical representations for decision makers, site owners, and the general public so that they may better understand the costs and consequences of achieving varying levels of cleanup at a contaminated site. Smart Sampling is developed at Sandia National Laboratories in USA. The purpose is to automate the decision framework into a user-friendly software tool.

A bibliography of projects where the Smart Sampling approach has been used is available at the internet link <http://www.nwer.sandia.gov/sample/bibliography.html>.

Russell and Rabideau

Russell and Rabideau (2000) evaluated the uncertainty in a risk-based decision analysis framework for Pump-and-Treat design. In total, 27 different remediation design alternatives were studied and for each alternative 36 calculations of the net present value (the objective function) were made, with the following variations in the framework:

- Two degrees of aquifer heterogeneity
- Two definitions of system failure
- Three definitions of cleanup standard
- Three failure costs

Based on these calculations, decision histograms were created, showing how many times each design alternative was favourable. It was found that the failure cost was the most important source of uncertainty in the analysis. Also, the need of a clear definition of failure was identified.

6.4.5 Optimisation approaches

The goal of frameworks based on optimisation has often been to minimise sampling costs while estimating some quantity (James and Gorelick, 1994). Here, a few works based on optimisation approaches will be described in short.

Tucciarelli and Pinder (1991) presented an optimal data acquisition strategy for groundwater remediation. The algorithm finds the total minimum for measurement cost and pumping cost. They use a deterministic simulation model but a stochastic optimisation approach. It was found that the optimisation algorithm was computationally intensive.

Wagner et al. (1992) used a stochastic optimisation model with the purpose to minimise total cost of installation and operation of pumping wells for containment of a contaminated groundwater plume. The value of information about hydraulic conductivity is calculated for different cases. It depends on when in the decision-making process the information is obtained and to what extent this information can affect further decisions.

Wagner (1999) presents a methodology for evaluation of data worth for groundwater management. The methodology consists of four steps: (1) The optimal groundwater management strategy is determined by a *management model*. The goal is to identify the least-cost remediation strategy (optimal pumping rates) that ensures water quality standards are met at a specified reliability level. This is performed by *non-linear maximum likelihood estimation theory*. (2) A monitoring *network design model* is used to identify the sampling strategy that will minimise uncertainty in model predictions for a specified data collection budget. The network design model is formulated as an *integer-programming optimisation* problem. (3) The management model is resolved based on the updated model prediction uncertainty from the network design model. The purpose is to determine the reduction in management cost if the optimal sampling strategy was implemented. (4) The worth of the sampling strategy is determined by comparing the reduction in management cost (worth of sample information) with the cost of data collection. The strategy is justified only if the worth of sample information exceeds sampling cost. Steps 2-4 can be repeated for a series of data collection budgets to identify the least-cost alternative.

Freeze and Gorelick (1999) compare the two approaches decision analysis and stochastic optimisation for aquifer remediation. A conclusion is that both strategies are well suited to assess the worth of additional data. A literature review is provided of research in optimisation and decision analysis for aquifer restoration. It was found that data worth within an optimisation framework has only been addressed in a few papers.

6.4.6 Rule-based adaptive sampling

An interesting but very different approach to sampling is the adaptive sampling approach as described by Cox (1999). It does not require any specific statistical model for the spatial distribution of contaminants, but instead constructs an increasingly accurate non-parametric approximation to it as sampling proceeds (Cox, 1999). In other words, the system “learns” from data as sampling goes on. *Decision rules* must first be developed and a vector of decision parameters describes them. The decision rules are created for the specific problem at hand, but are typically rules that promote cost-effective sampling and cleanup. When the decision parameters have been determined the adaptive

sampling strategy can be optimised by *simulation-optimisation* for identifying approximately optimal adaptive sampling decision rules. The simulation is performed using existing sample data. In the simulation, different combinations of values for the decision parameters are used and the combinations of parameter values that meet the performance criteria best are considered optimal.

Cox (1999) provides the following example of the adaptive sampling approach applied to contaminated land. Four decision parameters are used:

N = initial number of random samples to be taken from the area of concern. These samples are in addition to already collected samples.

S = number of soil samples to take from each newly sampled property (the area of concern is divided into a number of properties; in the example, each property is either to be remediated in its entirety or not at all).

K = maximum allowed length of a “search list”. This is a list of properties with the highest average sample concentration seen so far.

T = action threshold. A property is to be remediated if, and only if, the mean concentration of its S sample values exceeds T.

The decision parameter vector (N, S, K, T) is used in the simulation-optimisation to find the optimal combination of values. Then, the procedure of the adaptive sampling follows (a property is called *fathomed* when S samples have been taken from it and from each of its immediate neighbours):

1. Randomly sample N previously unsampled locations in the area of concern.
2. Place the K properties with the highest mean concentrations so far on the search list.
3. Fathom each property on the search list by taking S samples at the property and its neighbours, and remove the property from the list when it has been fathomed.
4. Whenever a property is found with a mean concentration greater than the values for properties not included on the initial search list (step 2), place that property on the search list.
5. Continue steps 3 and 4 until the search list is empty, i.e. until both the initial properties on the list and the ones added later has been removed. This terminates the adaptive sampling phase.
6. Remediate those properties with a mean concentration exceeding T.

Cox (1999) calls this procedure *worst first adaptive search*, in that the search concentrates on properties with high soil contamination levels. He also presents a strategy for the simulation-optimisation of the decision parameter vector.

The technique of adaptive sampling provides a practical approach to decide where to sample next, when to stop, and what to do to optimise an overall performance objective (Cox, 1999). The performance objective can be of different types, such as to minimise the number of misclassification errors, or expected utility. It can also be applied with risk-cost-benefit criteria.

One drawback in the example presented above is that the sampling may require many visits to the area of concern due to the multi-stage nature of adaptive sampling. However, it is possible to use only a few stages (e.g. a two stage sampling procedure) if the set of decision parameters is modified.

Note that the term adaptive sampling can be used for sampling approaches that are very different from the one presented here, e.g. those described in section 6.4.3.

6.4.7 Sampling for model discrimination

Another different approach to the worth of sample information is presented by Knopman and Voss (1988). The purpose of their study was to develop a methodology for choosing between different contaminant transport models based on errors in model predictions. They presented a methodology to compare different sampling designs in their efficiency of acquiring information for discrimination among models. Better sampling designs maximise the difference in predictions between models. The model predictions are compared by a defined objective function (a sum of squares objective function). The objective function is calculated for each sampling design. A sampling design with a larger value of the objective function supplies more information than a design with a smaller objective function value.

7 DISCUSSION

7.1 General

This literature review covers several topics related to sampling. As shown, the literature on the subject is quite extensive and covers much more than is presented in this report. There is a number of aspects about the material that is worth some comments. In this chapter a discussion is given with the purpose to identify research needs and areas of the sampling problem that could be developed further.

7.2 Prior information and expert knowledge

Many of the techniques and methodologies for sampling strategies are based on the assumption that no prior information about a site is available. Examples of such methods include classical sampling statistics, search theory etc. Some work has been done to include prior information like site history, geological conditions etc. in a quantitative analysis but in practice prior information is often included only in a qualitative way. More and better methods to include prior information quantitatively are needed. This is fundamental because prior information can be as important as the sampling information.

7.3 Models

The amount of data needed to understand the behaviour of a physical system may exceed the amount of data needed to make a good decision (Maddock, 1973). The physical model that uses the data may be far more complex than can be handled by the decision process. This may lead to collection of excessive data and overly sophisticated models (Maddock, 1973). Some of the models that have been encountered in this literature review have been quite complex, while some have been fairly simple. In many cases a simple model will supply sufficient information at a low cost. In other situations more advanced models will be needed. A conclusion is that there is a need for a methodology how to select a model of optimal complexity level. The methodology should be risk-based, taking the model uncertainty and the cost of application into account.

7.4 Particulate sampling theory

The presentation of the particulate sampling theory is described relatively detailed in this report compared to other topics presented. The reason is that the sampling theory is relatively unknown among practitioners, at least in Sweden. As mentioned previously, the theory was developed to promote mineral extraction but has also been applied to environmental problems. However, it is not evident that all aspects of the theory are applicable in all situations. For example, in the theory the *particle mass* is important because it relates to the grade in mineral exploration. In some contamination problems it may instead be the *surface area* of particles that is of primary concern, since the contaminant may cover the particle surface, not the particle volume. Probably, some aspects of the theory could be modified to include such aspects for environmental problems.

7.5 Sampling errors

Several of the different types of sampling errors presented in this report can be difficult to quantify. Practical experience of the size of each error can rarely be found in the literature. Information on how to quantify these errors is appreciated. As an example, the variability of contaminant concentrations in soil around a sampling point (short range), has only been addressed in a small number of papers but the conclusions have been that this type of uncertainty can be very large. However, these conclusions are only valid for these specific studies. The short-range variability will most likely depend on the site history, the geology, and the type of contaminant present. In situations where the variability is large, a single sample may randomly indicate “small contamination” or “large contamination”. In this case, there will be a potential for making wrong decisions if this type of sampling error is not taken into account. Another example of sampling error is the bias that can be introduced by the sampling equipment. Generally, a methodology to handle systematic sampling errors would be appreciated.

7.6 Sampling strategies

The question on how many samples to collect is always encountered in practice. One answer is to use data worth analysis as discussed below. Today, other methods are being used based on limited economic budgets or a desired accuracy of estimated properties. The literature on this matter is often restricted to certain types of problems, often sampling to estimate mean concentrations. Normally distributed data without any spatial correlation is often assumed. There is definitely a need for a review article or a publication on how to determine the necessary number of samples in a structured way under different assumptions, taking for example different sampling objectives, non-normally distributed data, spatial correlation, prior knowledge, and sampling with multiple objectives into account. No such well-structured presentation has been found in the literature. The perfect guide to sampling is yet to be published.

A common misunderstanding regarding the systematic sampling pattern requires a comment. The misunderstanding can be found among consultants and in rare cases even in the literature. Contrary to what many believe, sampling according to a systematic grid does not imply that randomness therefore is missing and that statistical or geostatistical techniques cannot be used. As long as all sampling locations have the same probability of selection (as is the case for systematic sampling if the first sampling point is selected randomly) the sampling can be regarded as random. Therefore, samples collected according to a systematic sampling pattern can very well be analysed statistically or with geostatistical techniques (provided that the grid itself is laid out randomly).

7.7 Data worth

The literature review has shown that data worth approaches have been used infrequently for three decades. There are probably a number of reasons why the concept has not been applied to real world problems more often. One reason is that the concept is relatively unknown. Another is that some methods are very computational intensive. The main reason is however that much of the previous work on data worth analysis has been quite complicated and difficult to grasp for an average consultant or researcher working in the field of contaminated land. This is especially true for the combination of data worth analysis and optimisation techniques. The complexity has restricted the practical use of

the data worth approach although the potential for its application is huge. Data worth analysis could for example be applied not only contamination problems, but also to classical geotechnical problems.

An exception to the complex papers discussed above is the work by James et al. (1996b) who present a stripped-down approach to data worth analysis. For broad picture analysis this kind of approach could be sufficient, for example to determine whether a sampling program is cost-effective or not. The methodology could be developed further to handle common problems encountered in contamination problems in an effective but understandable way. For in-depth analysis more sophisticated methods will probably be needed. Here, the Bayesian-Markov chain analysis is such a method that could be applied to data worth analysis of contaminated land on a relatively detailed level. Of course, more traditional geostatistic tools can also be applied, such as kriging techniques.

Preferable, a range of methods of different complexity should be available for data worth analysis. This will make it possible to select a simple approach when needed, for example in early stages of a project. If such an analysis turns out to be insufficient, more complex methodologies could be applied.

In cases when data worth analysis has been applied to sampling problems, the sampling uncertainties have often been neglected. Some of the uncertainties may have been considered, such as natural variability, while others have been ignored. If a practical data worth methodology is developed, it must be able to handle the different types of uncertainty that evolve in a site-investigation.

In the risk-based decision analysis approach a cost arises when failure occurs, i.e. when the failure criterion is met. The importance of the choice of failure criterion needs to be evaluated. How will the result of the analysis change if a different failure criterion is used? Also, the criterion for when to start remediation must be chosen properly (this is identical or closely related to the failure criterion). To begin remediation if the mean concentration exceeds an action level may not be an appropriate approach in all situations. The choice to start remediation must also be coupled to the different exposure routes for humans and how the contaminants affect the environment. A basis for the choice of failure and remediation criterion should be the source-path-target concept. Therefore, the data worth analysis will also have connections with the risk assessment methodology.

REFERENCES

- Abbaspour, K.C., Schulin, R., Schläppi, E. and Flühler, H., 1996. A Bayesian approach for incorporating uncertainty and data worth in environmental projects. *Environmental Modeling and Assessment*, 1996(1): 151-158.
- Abbaspour, K.C., Schulin, R., van Genuchten, M.T. and Schläppi, E., 1998. Procedures for Uncertainty Analyses Applied to a Landfill Leachate Plume. *Ground Water*, 36(6): 874-883.
- Albert, R. and Horwitz, W., 1996. Coping with Sampling Variability in Biota: Percentiles and Other Strategies. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 653-669.
- Alén, C., 1998. On Probability in Geotechnics. *Random Calculation Models Exemplified on Slope Stability Analysis and Ground-Superstructure Interaction*, Volume 1. Dissertation Thesis, Chalmers University of Technology, Göteborg, 242 pp.
- Aspie, D. and Barnes, R.J., 1990. Infill-Sampling Design and the Cost of Classification Errors. *Mathematical Geology*, 22(8): 915-932.
- ASTM, 1995. *Standard Guide for Developing Conceptual Site Models for Contaminated Sites*, ASTM Standards on Environmental Sampling. American Society for Testing and Materials, pp. E 1689-95, 1-8.
- Attoh-Okine, N.O., 1998. Potential Application of Influence Diagram as a Risk Assessment Tool in Brownfields Sites. In: K.B. Hoddinott (Editor), *Superfund Risk Assessment in Soil Contamination Studies: Third Volume*, ASTM STP 1338. American Society for Testing and Materials, West Conshohocken, pp. 148-159.
- Ball, R.O. and Hahn, M.W., 1998. The Statistics of Small Data Sets. In: K.B. Hoddinott (Editor), *Superfund Risk Assessment in Soil Contamination Studies: Third Volume*, ASTM STP 1338. American Society for Testing and Materials, West Conshohocken, pp. 23-36.
- Barcelona, M.J., 1996. Overview of the Sampling Process. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 41-61.
- Barnard, T.E., 1996. Extending the Concept of Data Quality Objectives To Account for Total Sample Variance. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 169-184.
- Bear, J., 1979. *Hydraulics of Groundwater*. McGraw-Hill, New York, 569 pp.
- Bear, J., 1988. *Dynamics of Fluids in Porous Media*. Dover Publications, New York, 764 pp.

- Ben-Zvi, M., Berkowitz, B. and Kesler, S., 1988. Pre-Posterior Analysis as a Tool for Data Evaluation: Application to Aquifer Contamination. *Water Resources Management*, 1988(2): 11-20.
- Black, S.C., 1996. Defining Control Site and Blank Sample Needs. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 139-154.
- Borgman, L.E., Gerow, K. and Flatman, G.T., 1996a. Cost-Effective Sampling for Spatially Distributed Phenomena. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 753-778.
- Borgman, L.E., Kern, J.W., Anderson-Sprecher, R. and Flatman, G.T., 1996b. The Sampling Theory of Pierre Gy: Comparisons, Implementation, and Applications for Environmental Sampling. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 203-221.
- Bosman, R., 1993. Sampling Strategies and the Role of Geostatistics in the Investigation of Soil Contamination. In: F. Arendt, G.J. Annokkée, R. Bosman and W.J. van den Brink (Editors), *Contaminated Soil '93, Fourth International KfK/TNO Conference on Contaminated Soil*. Kluwer Academic Publishers, Berlin, pp. 587-597.
- Box, G.E.P., Hunter, W.G. and Hunter, J.S., 1978. *Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building*. John Wiley & Sons, New York, 653 pp.
- Chai, E.Y., 1996. A Systematic Approach to Representative Sampling in the Environment. In: J.H. Morgan (Editor), *Sampling Environmental Media*. American Society for Testing and Materials, ASTM, West Conchohocken, pp. 33-44.
- Chang, Y.H., Scrimshaw, M.D., Emmerson, R.H.C. and Lester, J.N., 1998. Geostatistical analysis of sampling uncertainty at the Tollesbury Managed Retreat site in Blackwater Estuary, Essex, UK: Kriging and cokriging approach to minimise sampling density. *The Science of the Total Environment*, 1998(221): 43-57.
- Chiueh, P.-T., Lo, S.-L. and Lee, C.-D., 1997. Prototype SDSS for Using Probability Analysis in Soil Contamination. *Journal of Environmental Engineer*, 123(5): 514-520.
- Christian, J.T., Ladd, C.C. and Baecher, G.B., 1994. Reliability Applied to Slope Stability Analysis. *ASCE, Journal of Geotechnical Engineering*, 120(12): 2180-2207.
- Clark, M.J.R., Laidlaw, M.C.A., Ryneveld, S.C. and Ward, M.I., 1996. Estimating Sampling Variance and Local Environmental Heterogeneity for both Known and Estimated Analytical Variance. *Chemosphere*, 32(6): 1133-1151.
- Cox, L.A., 1999. Adaptive Spatial Sampling of Contaminated Soil. *Risk Analysis*, 19(6): 1059-1069.

- Crépin, J. and Johnson, R.L., 1993. Soil Sampling for Environmental Assessment. In: M.R. Carter (Editor), *Soil Sampling and Methods of Analysis*. Canadian Society of Soil Science/Lewish Publishers, pp. 5-18.
- Cressie, N.A.C., 1993. *Statistics for Spatial Data*. John Wiley & Sons, New York, 900 pp.
- Dagdelen, K. and Turner, A.K., 1996. Importance of stationarity for geostatistical assessment of environmental contamination. In: S. Rouhani, R.M. Srivastava, A.J. Desbarats, M.V. Cromer and A.I. Johnson (Editors), *Geostatistics for Environmental and Geotechnical Applications*. ASTM, West Conshohocken, PA, pp. 117-132.
- Dakins, M.E., Toll, J.E. and Small, M.J., 1994. Risk-Based Environmental Remediation: Decision Framework an Role of Uncertainty. *Environmental Toxicology and Chemistry*, 13(12): 1907-1915.
- Dakins, M.E., Toll, J.E., Small, M.J. and Brand, K.P., 1995. Risk-Based Environmental Remediation: Bayesian Monte Carlo Analysis and the Expected Value of Sample Information. *Risk Analysis*, 16(1): 67-79.
- Davis, D.R. and Dvoranchik, W.M., 1971. Evaluation of the Worth of Additional Data. *Water Resources Bulletin*, 7(4): 700-707.
- Decisioneering, 1996. *Crystal Ball version 4.0, users manual*, Denver.
- Deutsch, C.V. and Journel, A.G., 1998. *GSLIB, Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 369 pp.
- Drew, L.J., 1979. Pattern Drilling Exploration: Optimum Pattern Types and Hole Spacings When Searching for Elliptical Shaped Targets. *Mathematical Geology*, 11(2): 223-254.
- Englund, E.J. and Sparks, A.R., 1991. *GEO-EAS 1.2.1 Geostatistical Environmental Assessment Software User's Guide*. EPA 600/4-88/033, U.S. Environmental Protection Agency, Las Vegas.
- Fabrizio, M.C., Frank, A.M. and Savino, J.F., 1995. Procedures for Formation of Composite Samples from Segmented Populations. *Environmental Science & Technology*, 29(5): 1137-1144.
- Ferguson, C.C., 1992. The Statistical Basis for Sampling of Contaminated Land. *Ground Engineering*(June 1992): 34-38.
- Ferguson, C.C., 1993. Sampling Strategy Guidelines for Contaminated Land. In: F. Arendt, G.J. Annokée, R. Bosman and W.J. van den Brink (Editors), *Contaminated Soil '93, Fourth International KfK/TNO Conference on Contaminated Soil*. Kluwer Academic Publishers, Bering, pp. 599-608.

- Ferguson, C.C. and Abbachi, A., 1993. Incorporating Expert Judgement into Statistical Sampling Designs for Contaminated Sites. *Land Contamination and Reclamation*, 1: 135-142.
- Flatman, G.T. and Englund, E.J., 1991. Asymmetric loss functions for Superfund remediation decisions, *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pp. 204-209.
- Flatman, G.T. and Yfantis, A.A., 1996. Geostatistical Sampling Designs for Hazardous Waste Sites. In: L.H. Keith (Editor), *Principles for Environmental Sampling*. American Chemical Society, Washington, DC, pp. 779-801.
- Freeze, R.A., Bruce, J., Massman, J., Sperling, T. and Smith, L., 1992. Hydrogeological Decision Analysis: 4. The Concept of Data Worth and Its Use in the Development of Site Investigation Strategies. *Ground Water*, 30(4): 574-588.
- Freeze, R.A. and Cherry, J.A., 1979. *Groundwater*. Prentice-Hall, Englewood Cliffs, 604 pp.
- Freeze, R.A. and Gorelick, S.M., 1999. Convergence of Stochastic Optimization and Decision Analysis in the Engineering Design of Aquifer Remediation. *Ground Water*, 37(6): 934-954.
- Freeze, R.A., Massman, J., Smith, L., Sperling, T. and James, B., 1990. Hydrogeological Decision Analysis: 1. A Framework. *Ground Water*, 28(5): 738-766.
- Garner, F.C., Stapanian, M.A. and Williams, L.R., 1996. Composite Sampling for Environmental Monitoring. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 680-689.
- Gates, J.S. and Kisiel, C.C., 1974. Worth of Additional Data to a Digital Computer Model of a Groundwater Basin. *Water Resources Research*, 10(5): 1031-1038.
- Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York, 320 pp.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics Series. Oxford University Press, New York, 483 pp.
- Gorelick, S.M., Freeze, R.A., Donohue, D. and Keely, J.F., 1993. *Groundwater Contamination, Optimal Capture and Containment*. Lewis Publishers, Boca Raton, 385 pp.
- Grosser, P.W. and Goodman, A.S., 1985. Determination of groundwater sampling frequencies through Bayesian decision theory. *Civ. Engng Syst.*, 2(December): 186-194.

- Gustafson, G. and Olsson, O., 1993. The development of Conceptual Models For Repository Site Assessment, The Role of Conceptual Models in Demonstrating Repository Post-Closure Safety: Proceedings of an NEA Workshop. Nuclear Energy Agency, Organisation for Economic Co-operation and Development, Paris, pp. 145-158.
- Guyonnet, D., Côme, B., Perrochet, P. and Parriaux, A., 1999. Comparing Two Methods for Addressing Uncertainty in Risk Assessment. *Journal of Environmental Engineer*, 1999(July): 660-666.
- Haimes, Y.Y. and Hall, W., 1974. Multiobjectives in Water Resource Systems Analysis: The Surrogate Worth Trade Off Method. *Water Resources Research*, 10(4): 615-624.
- Halim, I.S. and Tang, W.H., 1993. Site Exploration Strategy for Geological Anomaly Characterization. *ASCE, Journal of Geotechnical Engineering*, 119(2): 195-213.
- Hammitt, J.K., 1995. Can More Information Increase Uncertainty? *Chance*, 8(3): 15-17, 36.
- Hammitt, J.K. and Shlyakhter, A.I., 1999. The Expected Value of Information and the Probability of Surprise. *Risk Analysis*, 19(1): 135-152.
- Harr, M.E., 1987. *Reliability-Based Design in Civil Engineering*. Dover Publications, New York, 291 pp.
- Heger, A.S. and White, J.E., 1997. Using influence diagrams for data worth analysis. *Reliability Engineering and System Safety*, 1997(55): 195-202.
- Hoffman, F.O. and Hammonds, J.S., 1994. Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability. *Risk Analysis*, 14(5): 707-712.
- Hoffman, F.O. and Kaplan, S., 1999. Beyond the Domain of Direct Observation: How to Specify a Probability Distribution that Represents the "State of Knowledge" About Uncertain Inputs. *Risk Analysis*, 19(1): 131-134.
- Huesemann, M.H., 1994. Guidelines for the Development of Effective Statistical Soil Sampling Strategies for Environmental Applications. In: E.J. Calabrese and P.T. Kosteki (Editors), *Hydrocarbon Contaminated Soils and Groundwater*. Lewis Publishers, Boca Raton, pp. 47-96.
- International Organization for Standardization, 1994. Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions. *ISO 5725-1*.
- Isaaks, E.H. and Srivastava, R.M., 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York, 561 pp.

- James, A.L. and Oldenburg, C.M., 1997. Linear and Monte Carlo uncertainty analysis for subsurface contamination transport simulation. *Water Resources Research*, 33(11): 2495-2508.
- James, B.R. and Freeze, R.A., 1993. The Worth of Data in Predicting Aquitard Continuity in Hydrogeological Design. *Water Resources Research*, 29(7): 2049-2065.
- James, B.R. and Gorelick, S.M., 1994. When Enough is Enough: The Worth of Monitoring Data in Aquifer Remediation Design. *Water Resources Research*, 30(12): 3499-3513.
- James, B.R., Gwo, J.-P. and Toran, L., 1996a. Risk-Cost Decision Framework for Aquifer Remediation Design. *Journal of Water Resources Planning and Management*, 122(6): 414-420.
- James, B.R., Huff, D.D., Trabalka, J.R., Ketelle, R.H. and Rightmire, C.T., 1996b. Allocation of Environmental Remediation Funds Using Economic Risk-Cost-Benefit Analysis: A Case Study. *Groundwater Monitoring & Remediation*, (Fall 1996): 95-105.
- Jenkins, T.F. et al., 1996. Assessment of Sampling Error Associated with Collection and Analysis of Soil Samples at Explosive-Contaminated Sites. Special report 96-15, US Army Corps of Engineers, Cold Regions Research & Engineering Laboratory.
- Jensen, F.V., 1998. A Brief Overview of the Three Main Paradigms of Expert Systems, (lecture notes). Dept. of Computer Science, Aalborg University. December 1, 2000, http://www.hugin.dk/huginintro/paradigms_pane.html.
- Johnson, R.L., 1996. A Bayesian/Geostatistical Approach to the Design of Adaptive Sampling Programs. In: S. Rouhani, R.M. Srivastava, A.J. Desbarats, M.V. Cromer and A.I. Johnson (Editors), *Geostatistics for Environmental and Geotechnical Applications*. ASTM, West Conshohocken, PA, pp. 102-116.
- Juang, K.-W. and Lee, D.-Y., 1998. Simple Indicator Kriging for Estimating the Probability of Incorrectly delineating Hazardous Areas in a Contaminated Site. *Environmental Science & Technology*, 32(17): 2487-2493.
- Kaplan, P.G., 1998. A Tutorial: A real-life adventure in environmental decision making. August 21, 2000, <http://www.doe.gov/riskcenter/>, U.S. Department of Energy, The Center for Risk Excellence.
- Keith, L.H. (Editor), 1996. *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, 848 pp.
- Keith, L.H., Patton, G.L., Lewis, D.L. and P.G., E., 1996. Determining What Kinds of Samples and How Many Samples to Analyze. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 3-40.

- Knopman, D.S. and Voss, C.I., 1988. Discrimination Among One-Dimensional Models of Solute Transport in Porous Media: Implications for Sampling Design. *Water Resources Research*, 24(11): 1859-1876.
- Koerner, C.E., 1996. Effect of Equipment on Sample Representativeness. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 155-168.
- Koltermann, C.E. and Gorelick, S.M., 1996. Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research*, 32(9): 2617-2658.
- Kumar, J.K., Konno, M. and Yasuda, N., 2000. Subsurface Soil-Geology Interpolation Using Fuzzy Neural Network. *Journal of Geotechnical and Geoenvironmental Engineering*, 126(7): 632-639.
- Lacasse, S. and Nadim, F., 1996. Uncertainties in characterising soil properties. In: C.D. Schackelford, P.P. Nelson and M.J.S. Roth (Editors), *Uncertainty in the Geologic Environment: From Theory to Practice*. American Society of Civil Engineers, Madison, WI, pp. 49-75.
- Lewis, D.L., 1996. Assessing and Controlling Sample Contamination. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 85-110.
- Maddock, T., 1973. Management Model as a Tool for Studying the Worth of Data. *Water Resources Research*, 9(2): 270-280.
- Mason, B.J., 1992. Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies. EPA/600/R-92/128, U.S. EPA, Center for Environmental Research Information, Cincinnati, OH.
- McLaughlin, D., Reid, L.B., Li, S.-G. and Hyman, J., 1993. A Stochastic Method for Characterizing Ground-Water Contamination. *Ground Water*, 31(2): 237-249.
- Mohamed, A.M.O. and Côte, K., 1999. Decision analysis of polluted sites - a fuzzy set approach. *Waste Management*(19): 519-533.
- Morgan, M.G. and Henrion, M., 1990. *Uncertainty, A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, 332 pp.
- Mukhopadhyay, A., 1999. Spatial Estimation of Transmissivity Using Artificial Neural Network. *Ground Water*, 37(3): 458-464.
- Myers, J.C., 1997. *Geostatistical Error Management. Quantifying Uncertainty for Environmental Sampling and Mapping*. Van Nostrand Reinhold, New York, 571 pp.

- Nathanail, C.P., Ferguson, C.C. and Tucker, P., 1998. The Use of Bayesian Statistics and Prior Information to Design Sampling Strategies. In: D. Moore and O. Hungr (Editors), Eight International Congress, International Association for Engineering Geology and the Environment. Balkema, Rotterdam, Vancouver, pp. 369-376.
- Naturvårdsverket, 1997. Åtgärdskrav vid efterbehandling, Vägledning för säkerställande av att acceptabla resthalter och restmängder uppnås - metoder och säkerhet. Rapport 4807, Stockholm.
- Olsson, L., 2000. Att bestämma subjektiva sannolikheter. Varia 488, Statens Geotekniska Institut, Linköping.
- Pitard, F.F., 1993. Pierre Gy's Sampling Theory and Sampling Practice. CRC Press, Boca Raton, 488 pp.
- Poeter, E.P. and McKenna, S.A., 1995. Reducing Uncertainty Associated with Ground-Water Flow and Transport Predictions. *Ground Water*, 33(6): 899-904.
- Ramsey, M.H., 1998. Sampling as a Source of Measurement Uncertainty: Techniques for Quantification and Comparison with Analytical Sources. *Journal of Analytical Spectrometry*, 13(February 1998): 97-104.
- Ramsey, M.H. and Argyraki, A., 1997. Estimation of Measurement Uncertainty from Field Sampling: Implications for the Classification of Contaminated Land. *The Science of the Total Environment*, 198(1997): 243-257.
- Ramsey, M.H., Argyraki, A. and Thompson, M., 1995. Estimation of sampling bias between different sampling protocols on contaminated land. *Analyst*, 120: 1353-1356.
- Ramsey, M.H., Squire, S. and Gardner, M.J., 1999. Synthetic reference sampling target for the estimation of measurement uncertainty. *Analyst*, 124: 1701-1706.
- Reichard, E.G. and Evans, J.S., 1989. Assessing the Value of Hydrogeological Information for Risk-Based Remedial Action Decisions. *Water Resources Research*, 25(7): 1451-1460.
- Robbat, A., 1996. A Guideline for Dynamic Workplans and Field Analytics: The Keys to Cost-Effective Site Characterization and Cleanup, U.S. EPA, Office of Site Remediation and Restoration and Office of Environmental Measurement and Evaluation, Medford, Massachusetts.
- Rosén, L., 1995. Estimation of Hydrogeological Properties in Vulnerability and Risk Assessments. Dissertation Thesis, Chalmers University of Technology, Göteborg.
- Rosén, L. and Gustafson, G., 1996. A Bayesian Markov Geostatistical Model for Estimation of Hydrogeological Properties. *Ground Water*, 34(5): 865-875.

- Rosenbaum, M.S., Rosén, L. and Gustafson, G., 1997. Probabilistic models for estimating lithology. *Engineering Geology*, 47(1997): 43-55.
- Ross, R.R. and Vieux, B.E., 2000. A Probabilistic Method for Estimating Monitoring Point Density for Containment System Leak Detection. *Ground Water*, 38(4): 533-540.
- Rowe, W.D., 1994. Understanding Uncertainty. *Risk Analysis*, 14(5): 743-750.
- Russell, K.T. and Rabideau, A.J., 2000. Decision Analysis for Pump-and-Treat Design. *Groundwater Monitoring & Remediation*, (Summer 2000): 159-168.
- Schumacher, B.A. and Minnich, M.M., 2000. Extreme Short-Range Variability in VOC-Contaminated Soils. *Environmental Science & Technology*, 34(17): 3611-3616.
- Shefsky, S., 1997. Sample Handling Strategies for Accurate Lead-in-Soil Measurements in the Field and Laboratory, International Symposium of Field Screening Methods for Hazardous Wastes and Toxic Chemicals. December 1, 2000, <http://www.niton.com/shef01.html>, Las Vegas, Nevada.
- Squire, S., Ramsey, M.H. and Gardner, M.J., 2000. Collaborative trial in sampling for spatial delineation of contamination and the estimation of uncertainty. *Analyst*, 125: 139-145.
- Starks, T.H., 1986. Determination of Support in Soil Sampling. *Mathematical Geology*, 18(6): 529-537.
- Stenback, G.A. and Kjartanson, B.H., 2000. Geostatistical Sample Location Selection in Expedited Hazardous Waste Site Characterization. In: P.W. Mayne and R. Hryciw (Editors), *Innovations and Applications in Geotechnical Site Characterization*. Geotechnical Special Publication. American Society of Civil Engineers, Denver, CO, pp. 155-168.
- Sturk, R., 1998. Engineering Geological Information - Its Value and Impact on Tunneling. Dissertation Thesis, Royal Institute of Technology, Stockholm, 191 pp.
- Sullivan, T.M., Gitten, M. and Moskowitz, P.D., 1997. Evaluation of selected environmental decision support software, U.S. Department of Energy, Office of Environmental Mangement.
- Taylor, A.C., 1993. Using objective and subjective information to develop distributions for probabilistic exposure assessment. *Journal of Exposure Analysis and Environmental Epidemiology*, 3(3): 285-298.
- Taylor, J.K., 1996. Defining the Accuracy, Precision, and Confidence Limits of Sample Data. In: L.H. Keith (Editor), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp. 77-83.

- Thompson, M. and Fearn, T., 1996. What Exactly is Fitness for Purpose in Analytical Measurements? *Analyst*, 121(March 1996): 275-278.
- Thompson, M. and Ramsey, M.H., 1995. Quality Concepts and Practices Applied to Sampling - An Exploratory Study. *Analyst*: 261-270.
- Treeage Software, 1999. Decision analysis. Data 3.5 user's manual, Williamstown.
- Tucciarelli, T. and Pinder, G., 1991. Optimal Data Acquisition Strategy for the Development of a Transport Model for Groundwater Remediation. *Water Resources Research*, 27(4): 577-588.
- Tucker, P., Ferguson, C.C., Abbachi, A. and Nathanail, P., 1996. A decision support system for designing sampling strategies for potentially contaminated sites. In: M.-J. Kraak, M. Molenaar and E.M. Fendel (Editors), *International Symposium on Spatial Data Handling - Advances in GIS research*. London Taylor & Francis, Delft, Netherland, pp. 65-76.
- U.S. EPA, 1994. Guidance for Data Quality Objectives Process. EPA QA/G-4. EPA/600/R-96/055, Office of Research and Development, Washington, D.C.
- U.S. EPA, 1997a. Data Quality Evaluation Statistical Toolbox (DataQUEST) User's Guide. EPA/600/R-96/085, Office of Research and Development, Washington, DC.
- U.S. EPA, 1997b. Guiding Principles for Monte Carlo Analysis. EPA/630/R-97/001, Risk Assessment Forum, Washington, D.C.
- U.S. EPA, 1997c. Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM). NUREG 1575, EPA 402-R-97-016.
- U.S. EPA, 2000a. Guidance for Choosing a Sampling Design for Environmental Data Collection, EPA QA/G-5S. Peer review draft, Office of Environmental Information, Washington, DC.
- U.S. EPA, 2000b. Guidance for Data Quality Assessment, Practical Methods for Data Analysis, EPA QA/G-9. EPA/600/R-96/084, Office of Environmental Information, Washington, DC.
- U.S. EPA, 2000c. User's Guide for the Data Quality Objectives Decision Error Feasibility Trials (DQO/DEFT) Software. EPA QA/G-4D, Washington, DC.
- UK Department of the Environment, 1994. Sampling Strategies for Contaminated Land. CLR 4, The Centre for Research into the Built Environment/The Nottingham Trent University, Nottingham.
- Wagner, B.J., 1999. Evaluating Data Worth for Ground-Water Management under Uncertainty. *Journal of Water Resources Planning and Management*, 125(5): 281-288.

- Wagner, J.M., Shamir, U. and Nemati, H.R., 1992. Groundwater Quality Management Under Uncertainty: Stochastic Programming Approaches and the Value of Information. *Water Resources Research*, 28(5): 1233-1246.
- van Ee, J.J., Blume, L.J. and Starks, T.H., 1990. A Rationale for the Assessment of Errors in the Sampling of Soils. EPA/600/4-90/013, U.S. Environmental Protection Agency / Environmental Research Center, Las Vegas.
- Warrick, A.W., Young, M.H. and Wierenga, P.J., 1998. Probabilistic Analysis of Monitoring Systems for Detecting Subsurface Contaminant Plumes. *Ground Water*, 36(6): 894-900.
- Webster, R. and Burgess, T.M., 1983. Spatial Variation in Soil and the Role of Kriging. *Agricultural Water Management*, 1983(6): 111-122.
- Vose, D., 1996. *Quantitative Risk Analysis. A Guide to Monte Carlo Simulation Modelling*. John Wiley & Sons, Chichester, 328 pp.
- Yates, S.R. and Yates, M.V., 1990. *Geostatistics for Waste Management: Users Manual for the GEOPACK (Version 1.0) Geostatistical Software System*. EPA/600/8-90/0004, Robert S. Kerr Environmental Research Laboratory / U.S. Environmental Protection Agency, Ada.
- Younes, H.L., 1998. *Current Tools for Assessing Intelligent Agents in Real-time Decision Making*. Master Thesis, Royal Institute of Technology, Stockholm, 45 pp.
- Yuhr, L., Benson, R.C. and Sharma, D., 1996. Achieving a Reasonable Level of Accuracy in Site Characterization in the Presence of Geologic Uncertainty. In: C.D. Schackelford, P.P. Nelson and M.J.S. Roth (Editors), *Uncertainty in the Geologic Environment: From Theory to Practice*. American Society of Civil Engineers, Madison, WI, pp. 195-209.
- Äikäs, T., 1993. *Conceptual Models: The Site Characterization Point of View, The Role of Conceptual Models in Demonstrating Repository Post-Closure Safety: Proceedings of an NEA Workshop*. Nuclear Energy Agency, Organisation for Economic Co-operation and Development, Paris, pp. 17-26.

APPENDIX: SOFTWARE PACKAGES

A1 INTRODUCTION

Using computer software can facilitate the practical application of several methods and techniques described in the literature review. Here, some of the most common software packages will be presented in short and the focus is on public domain software or relatively cheap packages, although some more expensive ones have been included. The presentation shows only a selection of available programs and it is by no means complete. The author has used only a few of the programs and no guarantee of the capabilities of the different software packages can be given. Further information about the software may be found at the listed web pages. A selection of other environmental decision support software can be found in Sullivan et al. (1997).

A2 SAMPLING DESIGN AND DATA EVALUATION SOFTWARE

A2.1 ASSESS

ASSESS is a public domain quality assessment program from U.S. EPA. It is based on the principles described by van Ee et al. (1990) and is a tool for statistically assessing measurement errors in the collection of soil samples. The *ASSESS* software is available from U.S. EPA on a CD (<http://www.epa.gov/crdlvweb/databases/datahome.htm>).

A2.2 DataQUEST

Data Quality Evaluation Statistical Toolbox (*DataQUEST*) is a public domain DOS software developed by U.S. EPA for data quality assessment purposes. The program is designed to provide a quick and easy way to review data and analyse assumptions even for those not familiar with standard statistical packages (U.S. EPA, 1997a). The software and manual (U.S. EPA, 1997a) can be downloaded from U.S. EPA (<http://www.epa.gov/quality/tools-prj.html>).

A2.3 DEFT

The Decision Error Feasibility Trials (*DEFT*) software is public domain software developed by U.S. EPA for use in the DQO process. *DEFT* was originally written for DOS but a beta version for Windows is now available (U.S. EPA, 2000c). The program uses hypothesis testing to calculate the required number of samples for not exceeding specified probabilities for decision errors (false positive and false negative) regarding a given action level (mean concentration). No account is made for spatial correlation between samples and normally distributed data values are assumed. Three different sampling designs can be handled: simple random sampling, stratified sampling, and composite sampling. The program also calculates sampling cost. The *DEFT* software and manual (U.S. EPA, 2000c) can be downloaded from U.S. EPA (<http://www.epa.gov/quality/tools-prj.html>).

A2.4 DQO-PRO

DQO-PRO is a public domain Windows-based software in the form of a calculator interface. It has three different functions: (1) to calculate the required number of samples for a specified confidence level (Success-Calc), (2) to calculate the number samples necessary to estimate an average concentration within a specified error (Enviro-Calc), and (3) to calculate the number of samples necessary to detect a hot spot (HotSpot-Calc) when no prior information is considered (Keith et al., 1996). The purpose of the software is to be a tool in data quality objectives (DQOs) and for quality control (QC). It can be used in conjunction with the commercial program *Practical Environmental QC Samples* (Keith et al., 1996) described below. The two programs are part of the commercial software package for environmental sampling and analysis, sold under the name of *Practical QC* (Keith et al., 1996). *DQO-PRO* is available from an internet link at the American Chemical Society (<http://www.instantref.com/dqopro1.htm>).

There appears to be a problem with the decimal point in the program, complicating its use on some computers.

A2.5 ELIPGRID-PC

ELPGRID-PC is a public domain DOS-program originally developed by D. A. Singer for locating geological deposits (Gilbert, 1987). The purpose of the current version is to search for hot spots of contaminants. The program calculates the probability of detecting a hot spot, sampling cost and the required size of the sampling grid. The theoretical basis for ELPGRID is summarised in Gilbert (1987). *ELPGRID-PC* can be downloaded from Oak Ridge National Laboratory (<http://homer.hsr.ornl.gov/ets/elipgrid.html>).

A2.6 Practical Environmental QC Samples

Practical Environmental QC Samples (Keith et al., 1996) is a low cost expert system that provides answers, explanations, and definitions for selecting quality control (QC) samples. It covers 15 types of QC samples, such as field blanks, trip blanks, equipment blanks, local control site background samples, replicate analyses, replicate samples etc. The software is available from Instant Reference Resources, Inc. (<http://www2.outer.net/iref/pracqc.htm>).

A2.7 Robust ANOVA

Ramsey (1998) describes a simple, non-commercial DOS-software for the statistical method Robust ANOVA. It can be used to separate different types of variability in the measurement process. The DOS executable file is called ROBCOOP4 and is available at the Journal of Analytical Atomic Spectrometry website (<http://www.rsc.org/suppdata/ja/1998/97/>).

A2.8 SimSite

SimSite is a public domain software for Windows 95. The software simulates a site-investigation to evaluate the need for remediation. The purpose of the software is educational and to develop site characterisation strategies. The program was developed by

the University of Texas at Austin
(http://www.ce.utexas.edu/prof/gilbert/public_html/SimSite.html).

A2.9 Site-ASSESS

Site-ASSESS (Assessment of Sampling Strategies Expert Support System) (Nathanail et al., 1998; Tucker et al., 1996) is a computer decision support system developed under contract for the UK Department of the Environment. Its purpose is to optimise sampling strategies for locating contamination hot spots and is centred on expert knowledge. It allows soft information such as prior knowledge acquired during a desk study to be incorporated by Bayesian statistics.

Prior information from maps, archive data on previous site use, interviews with former workers or residents, visible evidence of contamination, geophysical measurements etc. are incorporated in the analysis by a scoring system. The cells in an information grid covering the site are assigned scores depending on the prior information. The scores are converted to *a priori* probabilities. The number of samples required is expressed as a function of hot spot size and the area of the site. The sampling locations are generated by an optimisation technique, taking the prior information into account.

Site-ASSESS also allows the number of samples to be expressed as a function of hot spot size. The program makes it possible to compare sampling strategies designed on the basis of prior information with strategies based on the assumption of equiprobability. Unfortunately, Site-ASSESS is not available on a commercial basis (year 1999).

A2.10 Visual Sample Plan

Visual Sample Plan (VSP) is a new software for the planning of sampling activities. VSP has a graphic environment and calculates the number of samples to be collected, their location, and total sampling cost for a given sampling objective. The algorithms for detection of hot spots are the same as in ELIPGRID-PC. A free beta version of VSP is available from Oak Ridge National Laboratory (<http://etd.pnl.gov:2080/DQO/software/vsp/vspbeta.html>).

A3 GEOSTATISTICAL SOFTWARE

A3.1 Introduction

A multitude of different geostatistical software packages exists on the market. The ones presented here is only a subjective selection. A number of other geostatistical programs can be found at The Central Server for GIS & Spatial Statistics on the Internet (www.ai-geostats.org).

A3.2 DRILL GUIDE

DRILL GUIDE is a commercial software for geostatistical analysis by kriging. The software differs from many other geostatistical packages in that the variogram analysis is performed in the form of expert system algorithms. The variogram is multi-dimensional, taking the trends in data in different directions into account. *DRILL*

GUIDE is part of the Environmental Visualization System (EVS) and is sold by C Tech Development Corporation (<http://www.ctech.com>).

A3.3 ECOSSE

ECOSSE (Ecological Spatial Statistical Evaluation) is a Windows 95 product with 2D tools such as basic statistic analysis, variogram analysis, kriging and conditional simulation. It is available from Geostokos (Ecosse) Limited in Scotland (<http://www.geocosse.com>).

A3.4 GEOEAS

In 1987, U.S. EPA released the public domain software *GEOEAS* (Geostatistical Environmental Assessment Software) for geostatistical analysis of spatially correlated data (Englund and Sparks, 1991). It is probably the most widely used geostatistical software for environmental applications, despite its limitations. It is DOS-based (UNIX and Linux-versions also exist) and offers many useful geostatistical features such as basic statistics, scatter plots/linear regression, variogram computation and different options for kriging. Unfortunately, *GEOEAS* has not been upgraded since 1991. There are some peculiar errors in the program which may complicate the use at times. *GEOEAS* is available from U.S. EPA (<http://www.epa.gov/ada/geoeas.html>).

A3.5 GEOPACK

Another public domain software from U.S. EPA is *GEOPACK* (Yates and Yates, 1990). It is a geostatistical software system written for users with little experience in geostatistical techniques but with the requirements to also satisfy more advanced users. It is a menu-driven DOS-program with tools for basic statistics, variogram analysis, linear estimation (different kriging options) and nonlinear estimation. *GEOPACK* can be downloaded from U.S. EPA (<http://www.epa.gov/ada/geopack.html>).

A3.6 Geostatistical Toolbox

The *Geostatistical Toolbox* is a public domain DOS package from 1990 developed by Roland Froidevaux. It is one of the few geostatistical freewares with 1D, 2D and 3D tools and it contains several modules, e.g. univariate and bivariate statistics, variogram analysis and kriging. *Geostatistical Toolbox* is available from the same resource as *VarioWin* (manual included).

A3.7 GSLIB

Another public domain geostatistical software package is *GSLIB* (Geostatistical Software Library). It is a collection of geostatistical programs developed at Stanford University (Deutsch and Journel, 1998). *GSLIB* is a 3D package geared to professional users and runs under DOS (Linux-versions are also available). *GSLIB* offers advanced types of geostatistical analysis, such as cokriging, universal kriging, and conditional simulation. No manual is included in the public domain version of *GSLIB*. The software is available at the *GSLIB* web-site (<http://www.gslib.com>).

The original *GSLIB* inspired the writing of the well-known geostatistic book *GSLIB: Geostatistical Software Library and User's Guide* by Clayton and Journel (1998). The book is delivered with a CD-ROM with *GSLIB* source code.

A commercial supplement to the *GSLIB* package is available at a relatively low cost under the name *WinGslib*. It is a Windows 95/98/NT interface to *GSLIB* and related programs. *WinGslib* includes all tools in the *GSLIB* and some additional ones. *WinGslib* can be obtained from Statios (www.statios.com).

A3.8 Surfer

Surfer is a 3D and 2D surface modelling package for interpolation and with some geo-statistical features. The 7.0 version of *Surfer* offers e.g. variogram analysis and kriging options. *Surfer* is commercial available from Golden Software (www.goldensoftware.com).

A3.9 UNCERT

A more recent public domain package is *UNCERT*, developed under contract to the U.S. Bureau of Reclamation (Myers, 1997). It was developed for evaluating uncertainty in subsurface geology, hydraulic properties, and the migration of hazardous contaminants in groundwater flow systems. *UNCERT* runs under UNIX operating system, but work are in progress rewriting *UNCERT* into Java, making the software platform independent. The software contains a broad range of geostatistical tools, such as:

- modules to visualise 3D data
- 2D contouring and gradient analysis module
- 2D and 3D gridding module, using Inverse-Distance, Simple and Ordinary Kriging, and Trend-Surface Analysis
- semivariogram analysis in 1D and 2D

UNCERT and user's manual can be obtained from the Colorado Schools of Mines (<http://uncert.mines.edu/>).

A3.10 VarioWin

VarioWin is a software for spatial data analysis in 2D. It includes a number of graphical tools for variogram modelling and analysis. *VarioWin* was originally sold together with the book "VARIOWIN: Software for Spatial Data Analysis in 2D" by Yvan Pannatier (Springer-Verlag). However, the first edition of the book is out of print and the author has now made the program available free of charge. *VarioWin* can be obtained from the University of Lausanne (<http://www-sst.unil.ch/research/variowin/index.html>).

A4 PROBABILITY ESTIMATION SOFTWARE

A number of spreadsheets and software programs for assigning subjective probabilities or probability density functions (PDFs) are included as appendices in Olsson (2000). They include:

- a spreadsheet for assigning probabilities according to the eigenvector method (in Swedish)
- a spreadsheet for assigning probabilities according to a method by A. O'Hagen (in Swedish)
- a spreadsheet for assigning PDFs (in Swedish)
- the public domain software *SPAT* for assessing probability distributions. The program runs under DOS and was released in 1989 by P. Terlouw, Rijksuniversiteit Groningen, Netherlands.

A5 STOCHASTIC SIMULATION SOFTWARE

A5.1 Crystal ball

Crystal Ball is a commercial software from Denver, Colorado. It is a graphically oriented forecasting and risk analysis program (Decisioneering, 1996), designed for subjective selection of probability distributions. The program is an add-in to Microsoft Excel or Lotus and has a wide range of built-in probability distributions. Monte Carlo simulation and Latin Hypercube sampling can be performed. The program is available from Decisioneering (<http://www.cbpro.com>).

A5.2 @Risk

As *Crystal Ball*, *@Risk* is a commercial software. It is an add-in program to Microsoft Excel or Lotus and has similar functions to *Crystal Ball*. The software is available from Palisade (<http://www.palisade.com>).

A5.3 RiskSim

RiskSim is a low-cost software for Monte Carlo simulation. It is an add-in program to Microsoft Excel. The simulation results are presented in histogram and cumulative distribution charts. *RiskSim* is available from Decision Support Services (<http://www.treeplan.com>).

A6 DATA WORTH AND DECISION ANALYSIS SOFTWARE

A6.1 General

A list of more than 20 Bayesian network software packages (most of them public domain) is available at <http://http.cs.berkeley.edu/~murphyk/Bayes/bnsoft.html>. Links to other decision analysis software is also provided at this web site. A discussion of some of these software packages for decision analysis in real-time environment is available in Younes (1998).

A6.2 BayMar

The *BayMar* Windows software uses Markov-chain analysis for estimating probabilities of indicators for mutually exclusive states in 2D and 3D (Rosén, 1995). Probability es-

timates are updated using Bayesian statistics. Although no pre-programmed function for data worth analysis is available in the software, such analysis can still be performed with a simple sequence of functions in the program. The BayMar software is not commercially available but a demo version of the software is available in Rosén (1995).

A6.3 DATA

DATA (Decision Analysis TreeAge) is a commercial software for decision analysis. Its two main features are decision trees and influence diagrams. The trees and diagrams can be easily drawn and calculations can be performed. Other capabilities include Monte Carlo simulations and sensitivity analysis (Treeage Software, 1999). The software is available from Treeage Software (<http://www.treeage.com/products.htm>).

A6.4 DecisionPro

DecisionPro is a commercial decision analysis software with the capabilities to build decision trees and to perform Monte Carlo simulations. The software is available from Vanguard, where a demo version can be downloaded (<http://www.vanguardsw.com>).

A6.5 DPL

DPL (Decision Programming Language) is a commercial software offering two major decision-making tools; influence diagrams and decision trees. Other features include spreadsheet linking and sensitivity analysis. The software is available from Applied Decision Analysis, USA (<http://www.adainc.com>).

A6.6 Expert Choice

Expert Choice is a commercial decision support software based on the Analytic Hierarchy Process (AHP) methodology for decision making. The program allows building a model of the decision problem. The best decision is identified by via pair-wise comparisons of the relative importance of the variables. The program is available from Palisade (<http://www.palisade.com>).

A6.7 GeNIe

GeNIe is a decision analysis software for Windows, available at no cost for non-commercial research and education. A graphical interface allows the user to build Bayesian networks and influence diagrams by “drag-and-drop”. Functions include calculation of data worth (EVI and EVPI, see chapter 6). The program can be downloaded from the University of Pittsburgh (<http://www2.sis.pitt.edu/~genie>).

A6.8 HUGIN

The *HUGIN* system is a commercial software for decision analysis based on Bayesian networks. The software supports construction of graphical networks for the modelling of uncertainty and decision problems. Influence diagrams can be handled by the system. A demo-version of the software (HUGIN Lite) with limited capacity is available at no

cost from <http://hugin.dk/tryit.html>. Examples of networks for different problems are available, for example whether to make seismic soundings or drilling to determine the geological structure at a site.

A6.9 PrecisionTree

PrecisionTree is a commercial software for decision analysis. It is an add-in program to Microsoft Excel. The two main features are decision trees and influence diagrams. The software is available from Palisade (<http://www.palisade.com>).

A6.10 SmartSampling

SmartSampling performs economic risk-based decision analysis. It can estimate the probability that contamination exceeds a certain value, and use this information to choose between sampling strategies (Sullivan et al., 1997). The SmartSampling approach is described in more detail in chapter 6.

A6.11 TreePlan

TreePlan is a low-cost software for building decision trees in spreadsheets. It is an add-in program to Microsoft Excel. *TreePlan* is available from Decision Support Services (<http://www.treeplan.com>).