THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

On Uncertainty and Data Worth in Decision Analysis for Contaminated Land

PÄR-ERIK BACK

Department of Geology
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2003

On Uncertainty and Data Worth in Decision Analysis for Contaminated Land

PÄR-ERIK BACK
Department of Geology
Chalmers University of Technology

ABSTRACT

Contaminated soil and groundwater is a problem that has received increased attention in the last decade. Decision-making about investigation strategies, protective actions, and remedial actions is based on sparse and uncertain information, primarily data of contaminant concentrations and geological information. Because of limited economical resources, cost-efficient decisions must be made. Risk-based decision analysis is a tool for evaluating the cost-efficiency of different decision alternatives.

Identification of cost-efficient site investigations can be performed by applying data worth analysis. In such an analysis, the value of additional sample data is compared to sampling cost and if the data worth is larger than the cost it will be worthwhile to carry out the sampling. Because environmental samples are uncertain, this uncertainty should be considered in the analysis. An approach for estimation of uncertainty in soil sampling is presented. It is based on the sampling theory for particulate materials developed for the mining industry. The sample uncertainty is broken down into eight basic types of uncertainty and variability. An application of the methodology is presented for the problem of soil sampling with a drill auger. The result indicate that the uncertainty in sample data can easily be in the range of 30-40 %. The sampling uncertainty is believed to be much more important that the analytical uncertainty.

A methodology for including sample uncertainty in data worth analysis is presented. It is based on a Bayesian approach to data worth. The sampling objective is to estimate the mean concentration at a site. A MathCad computer application for the calculations is supplied. An application of the data worth estimation procedure is presented for a sampling problem at a former Ferro-alloy work in Gullspång, Sweden. A conclusion is that prior estimates of contaminant concentrations may have a significant impact on the result, as well as estimates of failure cost. It is recommended to use different estimates of failure cost to study its influence. Results also indicate that when sample uncertainty is increased, the expected net value of the sampling program will decrease moderately and relatively constant.

In situations where contamination has not yet occurred, cost-efficient protective actions need to be identified to combat environmental risks. A methodology for selecting cost-efficient protective actions for water supplies along railways has been developed. The risk object is railway transport of dangerous goods. Also for this problem, estimation of failure cost is believed to be important for the result.

The need for additional development of the methodology is identified. Estimation of uncertainty in soil sampling can be improved and the described theory extended. The methodology for data worth analysis for contaminated land should be extended to take additional sampling objectives into account.

Keywords: contaminated land, data worth, decision analysis, risk, sampling, uncertainty

ACKNOWLEDGEMENTS

The Swedish Geotechnical Institute (SGI) is gratefully acknowledged for financing and supporting the project. The applied research project leading to the paper in Appendix 1 was financed by the Swedish Rail Administration.

A number of people have contributed to the present result in numerous ways. I would like to express my sincere thanks to my supervisors Docent Lars Rosén and Professor Lars O. Ericsson at the Department of Geology, Chalmers University of Technology. I also wish to express my gratitude to all my colleagues at SGI and Chalmers for their inspiration, fruitful discussions and constructive comments. I would like to particularly thank my colleague in the project, Jenny Norrman, SGI. In addition, I highly appreciate the help provided by Docent Tommy Norberg, Mathematical Statistics, Chalmers and Göteborg University. The positive attitude and helpfulness displayed by Per-Arne Olofsson and his colleagues at Gullspång municipality was an inspiration during field-work at the former Ferro-alloy work in Gullspång. A special thank is also directed to Dr. Bo Berggren, SGI, for initiating the project and for introducing me to the risk-based philosophy.

I also want to express my sincere gratitude to the reference group consisting of Tomas Gell, Swedish Rescue Services Agency, Per Olsson, the county authorities in Västra Götaland, Professor Göran Sällfors, Chalmers, Dr. Jan Sundberg, SGI, Dr. Claes Alén, SGI, Fredrika Östlund, Swedish EPA, and Dea Carlsson, Swedish EPA.

TABLE OF CONTENT

1 INTRODUCTION

1.1 Background

Contaminated soil and groundwater is a problem that has received increased attention in the last decade. The risk of unacceptable exposure to contaminants for humans and other organisms in the environment makes it necessary to protect soil and groundwater from contamination. In order to select an efficient protection strategy, the risk of contamination has to be assessed. Managing the risk for contamination may include technically advanced protective actions, often resulting in costly and sometimes over-dimensioned protections, whereas similar objects in other locations are left unprotected. Therefore, it is desirable to consider the cost-efficiency of protective measures in a more structured and objective way. Methods to evaluate the cost-efficiency of different protective actions have so far had a rather limited use but there is an increasing demand for such approaches.

If a site already has been contaminated is be necessary to investigate the degree of contamination, in order to assess the risk and select mitigating measures. Such site investigations can be arranged in different ways but due to limited economical resources it is important to select an investigation strategy that is cost-efficient. The problem is that uncertainties are often large, and the result of a cheap site investigation will not supply the required information to make the correct decisions about the site. On the contrary, a site investigation resulting in only a small amount of uncertainty may be very expensive. However, 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 land. Often, the overall uncertainty is underestimated, especially in cases where only a few field samples have been collected. Because data analysis cannot recover more information than the samples contain (Flatman and Yfantis, 1996), collecting only a few samples may result in poor site characterisation, which in turn can lead to unsuccessful and expensive remediation decisions (James and Gorelick, 1994). Therefore, it is important to estimate the uncertainty in a site investigation so that cost-efficient strategies can be identified.

In order to estimate the uncertainty in a site investigation, it is necessary to quantify the uncertainty in the collected data. 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 field sampling 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.

Usually the investigation of a contaminated site is followed by a risk assessment, where the risk for humans and the environment is assessed. If the risk is assessed to be unacceptable, some kind of remedial action is often considered. A large number of remediation techniques exist and a decision must be made which technique to use. This decision should be made based on the cost-efficiency of the different alternative actions.

As mentioned above, decision-making regarding contamination problems of soil and groundwater includes decisions about (1) protective actions, (2) investigation strategies, and (3) remedial actions. All these decisions need to take the cost-efficiency of the action into account. In order to do this, the uncertainty cannot be ignored but must be handled in a structured and transparent way. Only this will make it possible to use the limited economical resources in an efficient way for contamination problems. It is our belief that this can be achieved by applying risk-based decision analysis to such problems.

1.2 Objectives

The scope of the project is to develop and apply a framework for risk-cost-benefit (RCB) decision analysis and data worth analysis to problems regarding contaminated soil and groundwater. There are two parts of the project, one focusing on the framework for RCB decision analysis (Norrman, 2000a) and one concentrating on uncertainty and data worth. This thesis deals with the second part. In principle, tools for RCB decision analysis and data worth analysis exist but they need to be modified and adjusted to contaminated land problems.

To achieve the main goal, four different means and main activities have been identified, constituting the main part the thesis. The purpose is to present:

- state-of-the-art regarding uncertainty and data worth analysis in site-investigations,
- a methodology for estimating uncertainty in soil sampling,
- a methodology to include sampling and analytical uncertainty in data worth analysis, and
- a methodology for selecting between different alternative protective actions by applying RCB decision analysis.

Background information on uncertainty and data worth in a decision analysis framework is presented in chapter 2 and 3 respectively. The results from the work to reach the three latter goals are presented in chapters 4 and 5, and in the paper in Appendix 1. Examples of application are also presented.

1.3 Limitations

As mentioned, decision-making regarding contaminated sites may include (1) protective actions, (2) investigations strategies, and (3) remedial actions. This thesis focuses on the first two aspects of the problem, whereas remedial actions are not covered.

The focus of uncertainty in site-investigations is on geochemical properties. Geologic, hydraulic, and transport properties are considered to a smaller extent, although they are associated with the important concepts of *parameter uncertainty* and *model uncertainty* described in chapter 2. However, the data worth methodology presented in chapter 5 can

be applied to for example geological, hydrogeological and geotechnical problems with slight modifications.

Methods to estimate the uncertainty in laboratory analyses (analytical uncertainty) 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 thesis.

In the literature there is an abundance of qualitative routines for sampling but they are of minor concern in this thesis. The reason is that they rarely support any quantitative estimation of uncertainty, which is a prerequisite for the approach taken.

The problem of sampling, risk assessment, and decision-making for contaminated land is quite complex chain of activities and in order to derive at practically applicable methods, several assumptions must be made. These are especially important to consider in chapter 5, since the presented approach has limitations regarding the objective of the sampling exercise (estimation of mean concentration) as well as the spatial distribution of the contaminant (randomly collected samples with no correlation between sample points is assumed).

2 UNCERTAINTY IN GENERAL

2.1 Quantities, parameters, variables and constants

A quantity is something that can be quantified in some way. Parameter is a similar term as quantity. 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 or contaminant concentration. Often, the empirical quantities constitute the majority of quantities in models and they are often uncertain. The commonly used term "parameter uncertainty" (section 2.4) refers to uncertainty in empirical quantities.

Defined constants are by definition certain, such as the mathematical constant π or the number of carbon atoms in a certain organic contaminant. 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.

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 inap-

propriate 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 probabilistic if one or more of the input quantities, on which they depend, are probabilistic. Otherwise the outcome criteria will be deterministic.

Note that the definitions given above are not used universally. For example, the term parameter has a more specific definition in statistics. Another example can be found in Gorelick et al. (1993), who make a distinction between parameters and variables in groundwater problems. They use the term "parameter" for time-independent properties such as hydraulic conductivity, whereas the term "variable" is used for time-dependent indicators such as hydraulic head and contaminant concentration. A conclusion is that it is important to define the terms used in order to avoid misinterpretation.

2.2 Uncertainty vs. error, probability, and risk

It is important to notice the difference between the terms uncertainty and error, although they are often used as synonyms. Terms like uncertainty, reliability, confidence and risk are probability-related and refer to *á priori* conditions, i.e. the situation before an event has occurred. 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 posterior, 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 completely throughout this thesis. The main reason is that the two concepts often are used more or less synonymous by many authors. Therefore, the concept chosen by a referred author has usually 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, a way of handling the uncertainty is to try to make a reasonable estimate of the error prior to investigation. This explains way 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. Some quantitative definitions of risk are given by Kaplan and Garrick (1980). An example of how risk can be defined in a decision analysis framework is presented in the paper (Appendix 1).

2.3 Classification 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 information 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. Therefore, the uncertainties should be handled differently depending on what type of quantity they refer to.

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, and (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*, see section 2.4) and uncertainty about model structure (*model uncertainty*, see section 2.5). Sturk (1998) classifies uncertainty in geological engineering problems into three classes; (1) inherent variability, (2) modelling uncertainty, and (3) parameter uncertainty. In the following sections we will use a wider definition of parameter uncertainty and include variability in the parameter uncertainty. The argument for this is that parameter uncertainty may encompass both uncertainty due to lack of knowledge and uncertainty due to inherent variability, as described in the section below.

Another type of uncertainty is the uncertainty in the interpretation of data and other information, uncertainty in the understanding of geological, chemical and biological processes, etc. Such uncertainties are closely related to conceptual uncertainties, see section 2.5.1. One type of interpretation uncertainties occur during classification of contaminated land, see Figure 3.6 in section 3.5.2.

2.4 Parameter uncertainty

Sturk (1998) distinguishes between three types of parameter uncertainty; (1) statistical uncertainty, (2) measurement errors, and (3) gross errors. Morgan and Henrion (1990) classify uncertainty in empirical quantities (see section 2.1) 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) and it is often expressed as *precision* (Figure 2.1). Precision is defined as "*the 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, 1997b).

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*, as in Figure 2.1 (the problem with Figure 2.1 is that the true value is known, which almost never is the case in real-world problems). A positive counterpart to bias has been presented by the International Organization for Standardization (1994) by invention of the term *trueness*. Trueness is defined as "*the 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, 1997b).

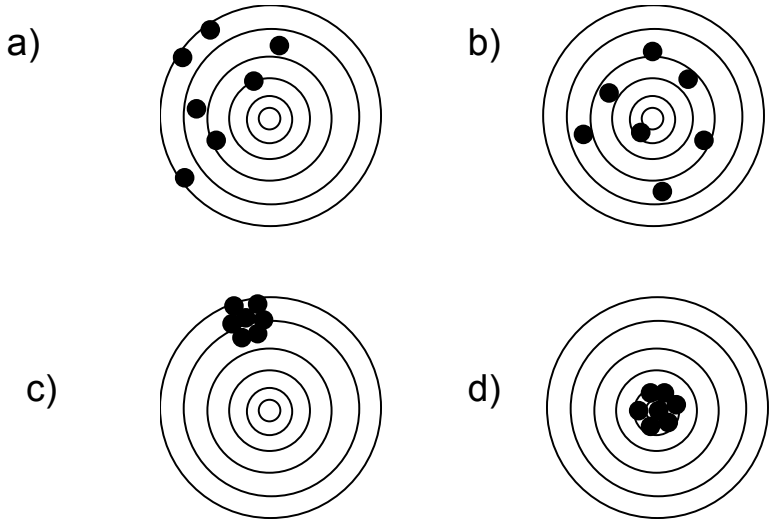


Figure 2.1 Patterns of shots at a target (after Gilbert, 1987).
(a) high bias + low precision = low accuracy
(b) low bias + low precision = low accuracy
(c) high bias + high precision = low accuracy
(d) low bias + high precision = high accuracy

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 Standardi-

zation (1994) defines accuracy as "*the 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).

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 is quite common but the way it influences the result is often not known.

Variability is the uncertainty due to a quantity that varies over space or time (spatial or temporal variability). An example of temporal variability is when the groundwater level in a monitoring well varies over time. Spatial variability is for example when the concentration of a contaminant varies over space. Similar terms are *population variability*, *geochemical variability* or *environmental variability*. It is not possible to reduce variability by taking more samples or making additional measurements, but our knowledge of the variability can be increased (Morgan and Henrion, 1990). Often, the variability of for example contaminant concentration is expressed as a variance. Variability relates to random error and statistical variation in spatial statistical problems. It is important to be aware of that spatial variability is scale dependent.

Some authors point out that it is important to distinguish clearly between uncertainty and variability (Hoffman and Hammonds, 1994; Rai et al., 1996). In this terminology, often used in human and ecological risk assessment, the word "uncertainty" is restricted to parameters with single but unknown values, whereas lack of knowledge in a parameter that is a function of space and time is called variability (McMahon et al., 2001). This is an indication of the importance to define the way uncertainty is used, in order to avoid misinterpretation. In contrast to the definition above, we will include variability when we use the word uncertainty in the following sections, but we agree that variability and other types of uncertainty should be distinguished when uncertainty is analysed.

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 (Burmaster, 1996; Morgan and Henrion, 1990). Goodman (2002) calls this type of randomness *process uncertainty*, probably including variability as described above.

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 and about the acceptable risk level (see chapter 3).

Approximation uncertainty arises because of the simplifications that are unavoidable when real-world situations are modelled. In many site-investigations the hydraulic con-

ductivity 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). Approximation uncertainty is closely related to conceptual uncertainty described in section 2.5.1.

Ramsey and Argyraki (1997) use the term *measurement uncertainty* as a way of characterising uncertainty in site-investigations of contaminated land. 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 4.3.2. Note that other authors may refer to measurement uncertainty as the analytical uncertainty only, for example Taylor (1996).

Christian et al. (1994) have categorised uncertainty in soil properties according to Figure 2.2. These uncertainties can be compared to random error, systematic error, and variability as described above.

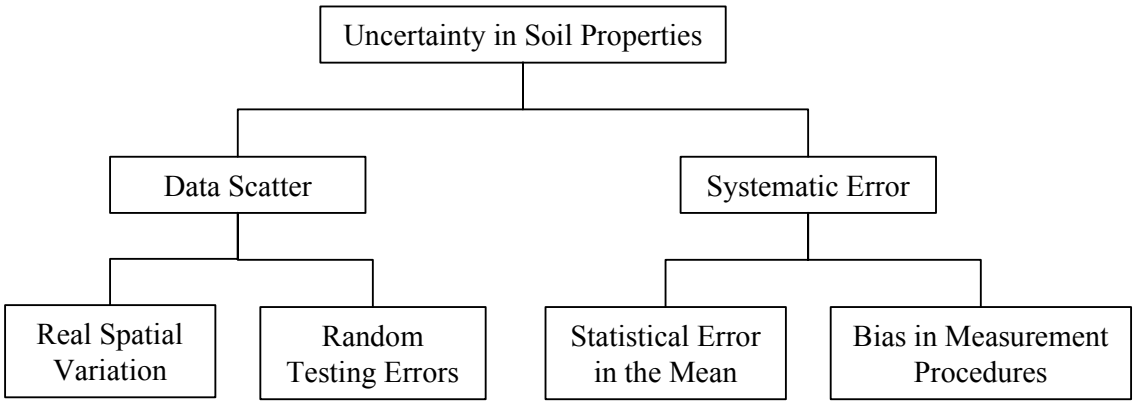


Figure 2.2 Categories of uncertainty in soil properties (after Christian et al., 1994).

For decision analysis problems of contaminated land there are several types of parameter uncertainty to consider. The most important may be uncertainty associated with the contaminant concentration. In chapter 4 and 5 the term *sample uncertainty* is used, including uncertainty in sampling and measurements (field and laboratory analyses).

2.5 Model uncertainty

2.5.1 Conceptual models

Distinction

One 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. Gustafson and Olsson (1993) define a model as an application of a theory to a specific problem (Figure 2.3). Uncertainty about model structure is believed to be more important than uncertainty about the value of a parameter. However, the distinction between model uncertainty and parameter uncertainty can be rather delicate (Morgan and Henrion, 1990). Model uncertainty is due to idealisations made in the physical formulation of a problem (Lacasse and Nadim, 1996). It can be divided into conceptual uncertainty (qualitative models) and uncertainty in mathematical (quantitative) models.

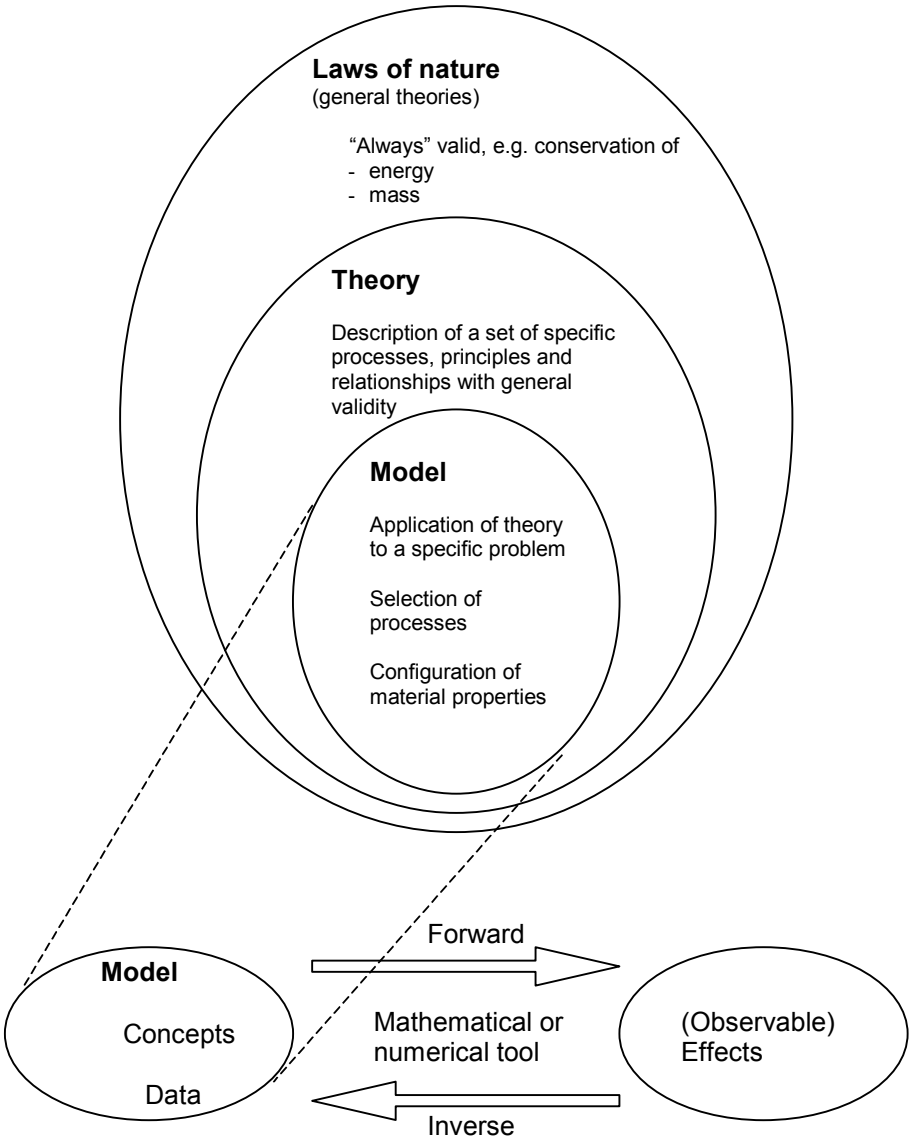


Figure 2.3 Schematic hierarchy of theories and models (after Gustafson and Olsson, 1993).

Development of a conceptual model

Conceptual uncertainty is uncertainty associated with our understanding of the system under study (McMahon et al., 2001). 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 uncertainties in the estimates of contaminant levels themselves. McMahon et al. (2001) give the following example of conceptual uncertainty in a contamination problem: *Is the decrease in contaminant concentration away from the source due to natural degradation and what processes are giving rise to and controlling degradation?* 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).

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 and U.S. EPA (1996) also gives advice on what information should be included in such a model. 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 does not mention how this should be achieved in practice.

Geological and hydrogeological conditions

During conceptualisation, it is important to consider the geological conditions, both at the actual site and in the surroundings where contaminant transport may occur. An understanding of the geological processes leading to the present situation will enable a more robust conceptual model to be developed. This will also facilitate compilation and evaluation of prior information (see sections 2.7.2 and 5.3.4). Geological and hydrogeological parameters that need to be considered are for example grain size distribution, specific surface, hydraulic conductivity, hydraulic gradients, groundwater recharge rates etc. In addition, the transport processes that result from the geological, hydrological, chemical, and biological conditions also need to be considered, e.g. advection, adsorption, diffusion, dispersion, decay etc.

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 is handled in 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.

James and Oldenburg (1997) 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 first-order second-moment (FOSM) and Monte Carlo methods. In order to transform the actual site history and conditions into a numerical simulation system, a set of four conceptual models was defined. These conceptual models were defined with the purpose to take the *conceptual model uncertainty* into account. The same set of parameter uncertainties (variances) were used for each of the four conceptual models (one base case and three variations of the base case) and uncertainty analysis was applied. 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.

Freeze et al. (1990) transformed a conceptual model into a *geological uncertainty model* and a *parameter uncertainty model* by incorporating the uncertainties. These models were 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 (see section 2.7.2). Soft information of where contamination is likely to be, was used to develop a conceptual model. This image was updated with new sample data by indicator kriging.

2.5.2 Quantitative models

In a similar way that parameter uncertainty and model uncertainty often overlap (Taylor, 1993), it is not easy to separate conceptual uncertainty from uncertainty in quantitative models. The quantitative models can be of different nature and examples include *analytical (mathematical) models*, *numerical models*, and *spatial statistical models*.

Usually, model uncertainty is aimed at the structure of 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 the input data is 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, whereas the precision can be expressed by the standard deviation of the model predictions.

Lacasse and Nadim (1996) argue that it is absolutely more rational to include model uncertainty in an analysis, rather than to ignore it. Model uncertainty is generally large but it can be reduced. 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.

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.

Sometimes, it is possible to construct a model in such a way that model uncertainty is converted to parameter uncertainty. 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:

$$E = E_S + E_R \tag{2.1}$$

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.

Geostatistical modelling includes several techniques, such as variogram analysis and a number of different kriging methods (see section 2.7.3). 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 uncertainty in their parameters, i.e. the variograms are considered deterministic. To take the uncertainty of a variogram into account, 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.

2.6 Qualitative estimation of uncertainty

In decision problems for contaminated land it is quite common to make qualitative estimations of uncertainty in parameter values and predictions. The magnitude of the uncertainty is almost exclusively expressed in words, such as “certain”, “quite uncertain”, “uncertain” etc., and the room for linguistic imprecision is therefore large. Surprisingly often, the uncertainty is not addressed at all. Systematic methods for other ways of estimating uncertainty qualitatively are extremely rare, and no such system has been identified for contaminated land problems. However, the topic of risk communication deals with how to communicate risk and uncertainty to the public, primarily in a qualitative way. Goodmann (2002) points out that the most useful description of uncertainty is a quantitative one, as described in the following sections.

Although systematic ways of describing uncertainty qualitatively are hard to come by, there exist methodologies that implicitly consider the uncertainty in a qualitative way. One such example is the Risk-Based Corrective Action (RBCA) for managing contaminant release sites in United States. RBCA consists of three different stages (called tiers) with increasing level of detail and decreasing uncertainty regarding the risk estimates. In Sweden, the Swedish EPA (Naturvårdsverket) uses a similar approach with investigation stages of increasing level of detail and reduction of uncertainty. The stages are called MIFO phase 1 and MIFO phase 2 (risk classification), simplified risk assessment, and extended risk assessment (Naturvårdsverket, 1997b; Naturvårdsverket, 1999). The uncertainties are assumed to decrease as one moves from an early stage to a more detailed one.

2.7 Quantitative estimation of uncertainty

2.7.1 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 *log-normal* 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 log-normally distributed (Ball and Hahn, 1998) or follow no predefined distribution. 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.

The standardised way of expressing uncertainty in measurement is by *standard uncertainty* or *expanded uncertainty* (Thompson and Ramsey, 1995). The former is related to standard deviation and the latter to confidence intervals. The expanded uncertainty is

obtained by multiplying the standard uncertainty by a coverage factor, normally around 2 or 3 ($\pm 2\sigma \sim 95$ % and $\pm 3\sigma \sim 99$ %).

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 whereas spatial data often is correlated.

2.7.2 Bayesian statistics

Bayes´ theorem

In contrast to classical statistics, Bayesian statistics allows all sources of information to 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). Using subjective information in addition to measurements is a big advantage in situations where only sparse data is available, which is often the case for geological problems. Therefore, Bayesian statistics are now widely used for such problems. 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):

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)} \qquad (2.2)$$

where $P(A_i \mid B)$ represents the probability of event A_i given that event B has occurred. A_i represents prior information (see below), whereas B is the new information that is used for updating the prior information.

As formulated above, Bayes´ theorem is relatively simple but when applied to continuous distribution functions the mathematics can become laborious. This is one reason risk analysts appear to split into two camps: those who use Bayes´ theorem extensively and those who do not (Vose, 1996). Another reason is different philosophical views of subjective prior information.

Prior information

Even if calculation with Bayes´ theorem is not performed, it is possible to include subjective prior information if a Bayesian approach is taken, although no *hard data* exist. Hard data are direct measurements or observations (Koltermann and Gorelick, 1996). *Soft data* on the other hand, 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.

If measurements or similar hard data is not available, soft data can be used to define a *probability density function* (PDF) of the prior information. *Prior distributions*, or prior PDFs, of subjective information can be used to characterise an individual's belief about the value of a parameter (Hammitt, 1995). They 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 much of the mathematics is 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.

A distribution based on soft data 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 (Hammitt 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 log-normal 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).

A less stringent methodology of including prior information 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) applied for hot spot detection.

Bayesian updating
In Bayesian statistics Bayes´ theorem is employed to determine *posterior* probability distributions of a variable by combining *prior* information with observed data (Vose, 1996). 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), see chapter 3 and 5. 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 (Figure 3.3). This step can be repeated several times. McLaughlin et al. (1993) calls this procedure "sequential updating".

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, 1997a; 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).

2.7.3 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). 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). Below, the principles of geostatistics are described relatively detailed compared to other approaches in this chapter. The reason is that the geostatistical approach is sometimes applied on contaminated land problems, and it is one approach that could have been used in this thesis as an alternative to the approach taken in chapter 5.

Variogram
The *variogram*, or *semivariogram*, is often used to analyse spatial or temporal correlations (Figure 2.4). 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 in each group 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 2.4 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 2.4 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 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 non-uniform contamination release, and of heterogeneity and anisotropy in the geological medium. Longer correlation distance can be expected in contaminated groundwater than in soil.

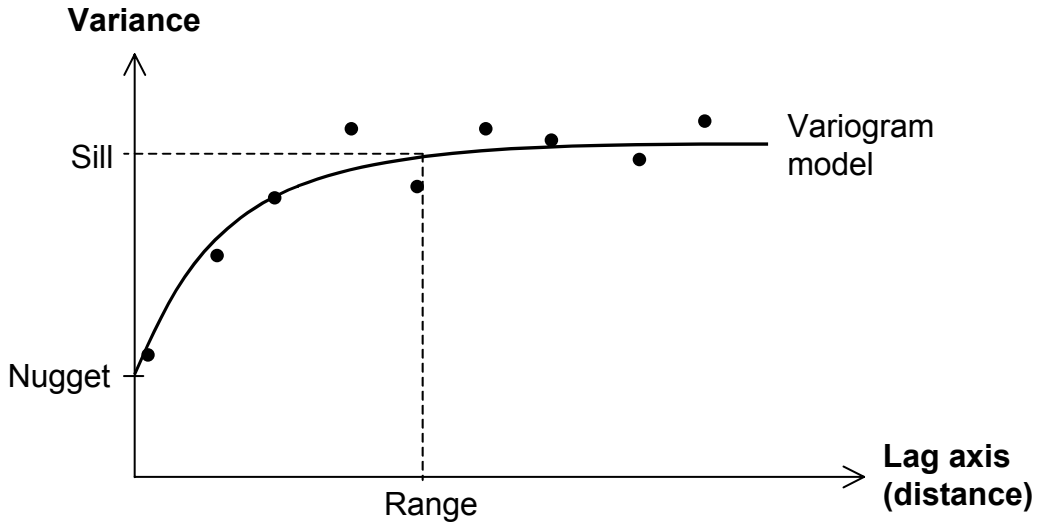


Figure 2.4 Example of a variogram with sill, range and nugget defined.

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 because of the *nugget* effect. 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). Cressie (1993) explains the nugget effect as a result of micro-scale variations and sampling errors. The micro-scale variation is the sum of the fundamental error and the grouping and segregation error (Myers, 1997), whereas the rest of the nugget effect is due to the random part of the materialisation errors, preparations errors and analytical errors, as described in chapter 4. The nugget effect is reduced if these errors are minimised.

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, non-linear 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. An important technique for contaminated land problems is *indicator kriging*. Each sample data is classified as either contaminated or not contaminated (1 or 0) and kriging is performed on these indicator data.

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.

Dagdelen 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 and the data set. Two major reasons are (1) that the coefficient of variation in the data set often is so large that the model assumptions are inappropriate, and (2) that the assumption of second-order stationarity often is not fulfilled because of multiple contaminant sources or different geologic environments within the study area. Cressie (1993) describes different types of stationarity, e.g. intrinsic stationarity (*the intrinsic hypothesis*) of which second-order stationarity is a special case, and ergodicity. Discussions about these can also be found in Freeze et al. (1990). The stationarity of the model is the chief assumption of kriging that is often violated (Dagdelen and Turner, 1996).

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. Such correction techniques 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 of 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).

2.8 Approaches for modelling of uncertainty

2.8.1 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 unnecessarily strong (Alén, 1998). Similar approaches are used for human health risk assessment and when generic guideline values for contaminated soil are developed.

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 thesis.

2.8.2 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. These variables are assigned as probability density functions (PDFs) describing 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 2.7.2), and it is the foundation for the approach taken in this thesis. 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.

2.8.3 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 generalisation 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 3.4.3) for subsurface soil geology interpolation. The possibility of occurrence of different geologic classes was calculated. Recently, effort has been made to combine the advantages of the possibility approach with the probabilistic approach for assessing uncertainty in risk assessment. Guyonnet et al. (2003) call this the “hybrid approach”.

3 DECISION ANALYSIS FOR CONTAMINATED LAND

3.1 Decision-making under uncertainty

The main problem at a contaminated site is how to make *decisions under uncertainty*. Because we only have limited information about the conditions at a site, the decisions to be made will always be associated with uncertainties. Therefore, such decision problems can preferably be handled in a framework based on *decision analysis*. Decision analysis is a tool for making well-founded and defensible decisions under uncertainty. The decisions involved in contaminated land problems is typically about selecting one action among a set of alternatives. Examples of different types of actions are:

1. to do nothing,
2. to perform field investigations (data collection),
3. to take protective actions,
4. to take mitigating actions (e.g. remediation), or
5. to monitor the site.

In this thesis we will concentrate on actions of type 2 and 3 (see section 3.3).

The basis for the decision-making is the available information about the site. It can be information about the site history, the geological setting, previous results from site investigations, risks for recipients on the site and in the surroundings, economical limitations, political and juridical aspects etc. Several of these factors are associated with a significant degree of uncertainty, especially regarding information about the geochemical, geological, and hydrogeological situation at the site. Therefore, the decision-making must be based on a framework taking these uncertainties into account. All relevant uncertainties in the pre-sampling, sampling and post-sampling activities have to be handled.

The decision-maker must select one alternative from a set of alternative actions, even if the uncertainty associated with the decision is large. Decision analysis makes it possible to select the most cost-efficient alternative under the present state of knowledge. When additional information is collected there is a possibility of change in decision.

The decision will also involve decision variables of different types. In the case the decision concerns site investigations, decision variables may be the following; type of investigation, medium to sample, number of sample points, location of sample points, laboratory analyses, etc. Decision-making about protective actions may instead concern protective techniques, administrative restrictions etc. On the other hand, if the problem is decision-making about mitigating actions the decision variables may be; type of remediation technique, volume of soil to remediate etc. As shown, the decision variables will be different depending on the type of action. Therefore, the decision-making problem must be structured slightly different depending on the type of action considered, but the fundamental framework can be similar.

There are several advantages of using decision analysis for decision-making under uncertainty. First of all, complex problems are structured and presented in a clear way to the involved parties, leading to more transparent decisions. Secondly, the cost-efficiency of different alternative actions can be analysed and the most cost-efficient

alternative be identified. Thirdly, if applied properly, the subjectivity in decision-making will be reduced, or at least it will become clear where the subjective opinions enter the decision-making process.

3.2 Decision analysis framework

The decision analysis framework for contaminated land problems (including groundwater) is presented in Figure 3.1. It is a risk-cost-benefit (RCB) decision analysis framework, based on traditional risk-cost-benefit analysis and the principles presented in the review paper by Freeze et al. (1990). The approach will be shortly described, roughly following the presentation given by Eklund and Rosén (2000).

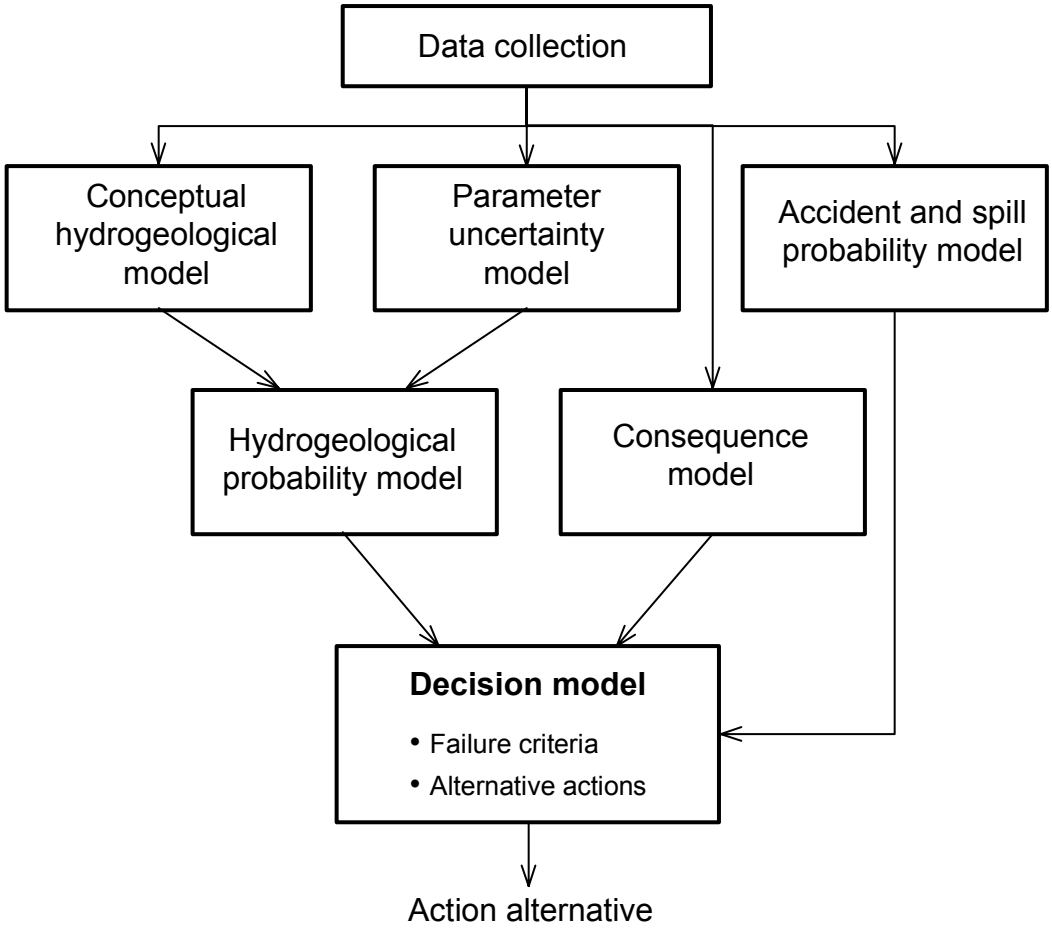


Figure 3.1 Models in the decision analysis framework (after Freeze et al., 1990).

The principle of RCB decision analysis is that the trade-offs between a given set of alternative actions are evaluated in order to select the most cost-efficient alternative. Each alternative implies a certain risk-reduction and the alternative action supplying the largest reduction in risk is the most cost-efficient one. The benefits, costs, and risks of each alternative are taken into account by an objective function, Φ_i, for each alternative i = 1,…, n. The general form of the objective function is:

$$\Phi_i = \sum_{t=0}^{T} \frac{1}{(1+r)^t} \left[B_i(t) - C_i(t) - R_i(t) \right] \quad (3.1)$$

where B_i represents the benefits of alternative i in year t, C_i is the investment and operational costs of alternative i in year t, R_i is the risks (expected cost or probabilistic cost) for alternative i in year t, r is the discount rate [decimal fraction], and T is the time horizon [years]. The objective function represents the net present value of the alternative i.

In the *decision model* in Figure 3.1, the risk is quantified as an expected annual cost (probabilistic cost). This is performed by multiplying the probability of an event with the consequence of the same event, and summing up for all possible events in the analysis. The risk-quantification is facilitated if the problem can be structured as an event tree or a decision tree. A simple example of an event tree is presented in Figure 3.4. The Risk-term R_i in equation 3.1 is quantified as:

$$R_i = \sum_{j=1}^{N} \left[P_j \cdot C_j \right] \quad (3.2)$$

where P_j is the probability of event chain j leading to a consequence, C_j is the consequence cost (expressed in monetary terms) of event chain j, and N is the total number of event chains leading to consequence costs (terminal nodes in the event tree). If only one event is of importance for the risk, this event is often defined as *failure*. Consequently, the probability is called "probability of failure", P_f, and the consequence "cost of failure", C_f (Freeze et al., 1990). Of importance for the analysis is how failure is defined, i.e. the *failure criterion*. A short discussion of different failure criteria is given in the paper in Appendix 1. For contaminated land problems, it is reasonable to define failure to occur if contaminant concentration exceeds an action level, such as a guideline value or soil screening level.

Different decision criteria can be used based on the objective function in equation 3.1, but a reasonable one is to maximise the objective function, i.e. to maximise benefits and to reduce the sum of costs and risks. Using this decision criterion, alternative actions that are more costly than the risk-reduction they provide cannot be justified. This means that none of the considered alternative actions may in fact be cost-efficient. However, it will still be possible to identify the most cost-efficient alternative among the considered ones. If a socially acceptable risk has been defined, the objective of the decision-maker is to reach that risk level to the lowest possible cost (Eklund and Rosén, 2000), as illustrated in Figure 3.2.

Now, let us as an example consider two alternative risk-reducing actions, A and B. We assume that there are no benefits from either alternative, so that only the costs and the risks for each alternative need to be considered. By applying equation 3.1 we can calculate the objective function Φ_A and Φ_B respectively. Figure 3.2 illustrates which alternative is the most cost-efficient one. Because alternative B has a higher value of the objective function (lower value of -Φ in Figure 3.2), it is the one to prefer. If we want to study how cost-efficient this alternative is, we will have to compare the cost of alternative B with the risk-reduction it provides, i.e. we must also know the present risk-level when no action has been taken. Assuming that the present risk is R_p, we can conclude

that the risk-reduction of alternative B is larger than the cost and therefore alternative B is a cost-efficient one. However, the risk is larger than the acceptable risk level set by society, as indicated in Figure 3.2, and therefore it is not an acceptable alternative by society. In reality, a problem is that the accepted risk level usually has not been defined by society.

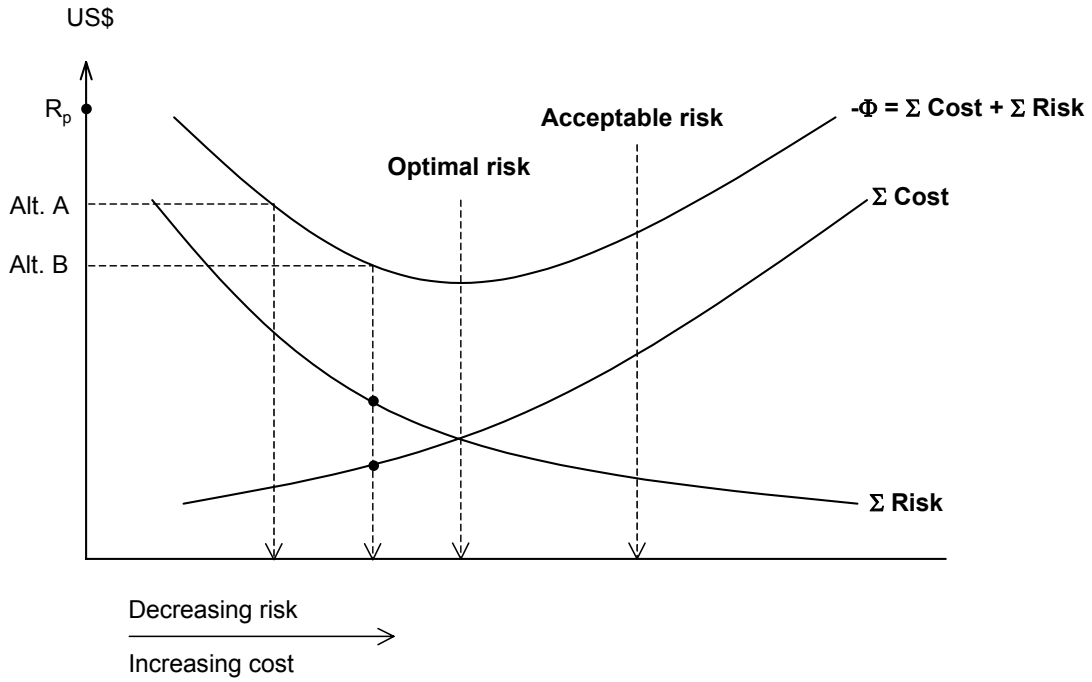


Figure 3.2 *Risk-cost minimisation as a decision criterion and the concepts of cost-efficiency, optimal risk, and acceptable risk (after Freeze et al., 1990).*

So far, only the decision model in Figure 3.1 has been described. A short discussion of the other models will also be made. The first box, *data collection*, considers collection of both hard and soft information. Of special interest is data collection by a field investigation program, as discussed in section 3.3.1.

The *conceptual hydrogeological model* is a qualitative description of the geological and hydrogeological conditions at the site (see section 2.5.1). In situations where the groundwater situation is not included in the analysis, such as in chapter 4 and 5 where attention only is pay to the contaminated soil, this model will supply information about the geological conditions at the site, e.g. soil particle distribution and other soil properties.

The *parameter uncertainty model* is a compilation of the geochemical and transport parameters and their associated uncertainty. This information is required by the *hydrogeological probability model*, which is used to estimate the probability of occurrence of different defined events. Estimation of probabilities can be performed by analytical transport models, as described in Appendix 1, or by numerical models. If only the contaminated soil and not the groundwater is addressed, only the geochemical aspects is considered in the model, i.e. the probability of exceeding some defined action level.

The *accident and spill probability model* supplies information about the likelihood of a spill event to occur. This model is only relevant for problems where the risk object is still present, i.e. contamination can still occur. The *consequence model* is used to estimate the consequences of different events. In the approach taken in this thesis, all consequences are estimated as monetary costs.

The framework for decision analysis presented in Figure 3.1 is a foundation for the approach taken in this thesis. Two types of decisions (alternative actions) will be considered; (1) how to select an investigation program, and (2) how to select protective action.

3.3 Selection between alternative actions

3.3.1 Field investigation programs

There are several different ways of collecting data from a contaminated site. The most common and direct way is to take samples, often complemented with other types of data collection such as in situ measurements, ocular inspection etc. The focus of this thesis is on sampling of soil as a means of data collection but the principles apply to other data collection strategies as well. The extent of the data collection is specified in a field investigation program. Examples of information in such a program are which medium to sample, what to measure, how many samples to take, where to take them etc.

Two types of questions concerning data collection programs can be addressed in a RCB decision analysis framework (after Freeze et al., 1992):

1. Given the available information, which data collection program from an available set of alternatives is the best one?
2. Is it worthwhile trying to improve the data set by performing additional sampling before deciding which is the best action alternative at the site?

In the first question we can define “best” as the most cost-efficient investigation program. This decision problem can be solved in the decision framework presented above, i.e. RCB decision analysis is a tool for selecting cost-efficient investigation programs.

The second question will arise at some stage in a contaminated land problem. The question can also be formulated as: “Do we have enough data or will it be cost-efficient to collect more data?” This question can be answered by applying *data worth analysis* to the problem. The principles of data worth analysis are presented in section 3.4. An application of data worth analysis for sampling of soil is described in chapter 5.

3.3.2 Protective actions

The framework in Figure 3.1 also apply to situations where no contamination has occurred yet, but where there is a risk object present, i.e. there is a possibility for accidents, spills etc. Examples of such risk objects include buried oil tanks, oil pipes, transport of dangerous goods on roads and railways etc. The decision problem in such situations is how to select cost-efficient protective actions in order to avoid contamination in the future.

The protective actions can be very different depending on the problem. If the risk object is a buried oil tank the protective actions could for example be to (1) remove the tank, (2) install a protective casing around the tank, (3) to install an oil level alarm, or (4) to inspect the tank regularly. Each protective action will imply a certain cost and reduction in risk level. Taken this into account, the most cost-efficient alternative can be identified, as explained in section 3.2 and Figure 3.2.

A methodology for decision-making about protective actions for water supplies along railways is presented in the paper in Appendix 1. It illustrates how RCB decision analysis can be applied to identify cost-efficient protective actions. The methodology is based on Figure 3.1 and event trees similar to Figure 3.4. Analytical transport models are used to estimate the various probabilities. An extended version of the methodology is presented by Back and Rosén (2001).

3.3.3 Mitigating actions

In situations where contamination has already occurred, the decision problem is how to select the most cost-efficient mitigating action. For contaminated land problems, this often means some type of remedial action, e.g. excavation of the contaminated soil, soil washing, in situ treatment, encapsulation etc. The decision framework for this problem will be similar to Figure 3.1, except that there will be no "Accident and spill probability model" because the spill has already taken place. The mitigating actions are not the focus of the thesis, although this decision problem is of great importance for contaminated land problems. The question of whether to remediate or not is only briefly discussed for the data worth problem in chapter 5.

3.4 Data worth analysis

3.4.1 Principles

In contaminated land problems, questions will inevitable arise if investment in more data is warranted or not. Such questions can be (Goodman, 2002):

- Will the cost of the data lead to a reduced cost for the decision to be made?
- Will the investment in data be more profitable if we collect different kinds of data?
- Will we achieve a greater benefit by investing the same amount of money in remedial or protective measures rather than data?

These, and other questions can be successfully answered by application of data worth analysis in a RCB decision analysis framework. The concept we will use to achieve this is *data worth*, i.e. the worth of additional information. How the concept of data worth is defined depends on the philosophy and the perspective of the decision-maker (Freeze et al., 1992). In a decision analysis framework the cost-efficient solution to a problem can for example be found by maximising the objective function in equation 3.1.

In a contaminated land project, data collection such as sampling, will continue until a *stopping rule* terminates the sampling. Different decision frameworks will have different stopping rules, and consequently the worth of data can be defined differently. A

simple definition of data worth is the difference between the cost of data collection and the expected value of the risk reduction the data provides (Freeze et al., 1992). Another definition is that additional data only has value if it can alter the choice of decision to make (Hammitt, 1995). Some of the most common or most interesting frameworks for data worth analysis are presented in section 3.5. 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.

The application of data worth analysis in chapter 5 is based on the definition of data worth by Freeze et al. (1992). More formally, they define data worth as "…*the increase in the expected value of the objective function between the prior analysis and the pre-posterior analysis due to the availability of the proposed additional measurements*". This definition of data worth can be summarised by the following equation:

$$EVSI = \Phi_{preposterior} - \Phi_{prior} \tag{3.3}$$

where *EVSI* (Expected Value of Sample Information) is the worth of additional data, Φ_{prior} is the objective function (equation 3.1) for the prior stage, and $\Phi_{preposterior}$ is the objective function for the preposterior stage (Figure 3.3).

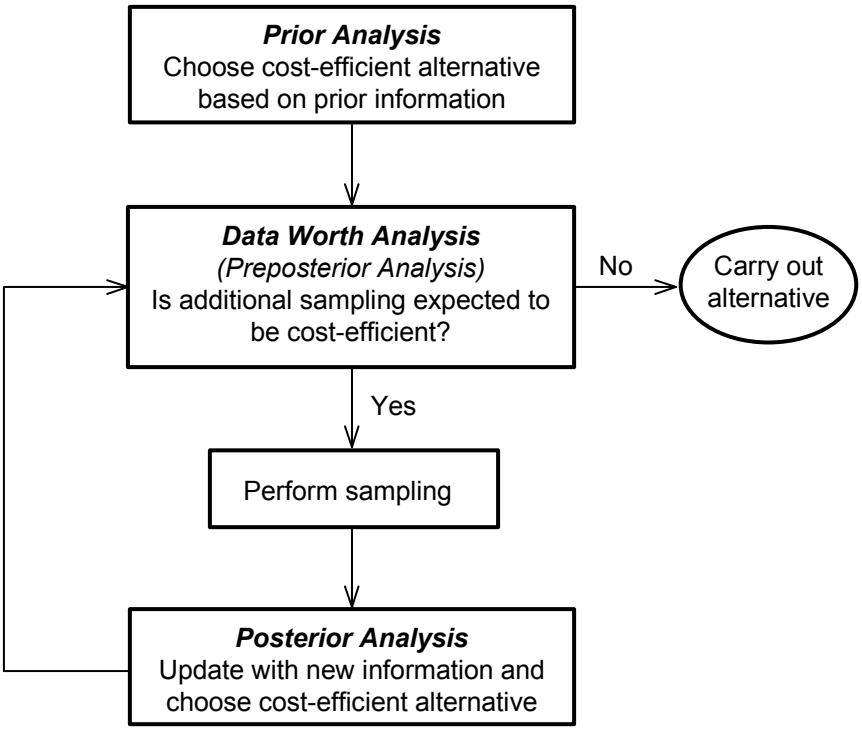


Figure 3.3 Data worth analysis in a decision framework based on Bayesian philosophy (after James and Gorelick, 1994).

Figure 3.3 illustrates the concept of data worth analysis in a decision framework. Prior, preposterior, and posterior analyses refer to the different stages of Bayesian updating (see section 2.7.2). *Prior analysis* refers to analysis of decision alternatives based on prior information only, i.e. before any sampling has taken place. *Posterior analysis* on

the other hand, is the analysis performed after sampling has occurred, i.e. when new data is available. The box "data worth analysis" in Figure 3.3 is also called *preposterior analysis*. This analysis is carried out when the data collection program has been defined in detail (number of samples, sample locations etc.) but before actually taking the samples. At this point it is known that the additional samples will supply information and thus reduce the risk-term in equation 3.1. By calculating the expected reduction in risk is it possible to estimate the worth of the new samples before they have been collected.

The framework in Figure 3.3 is based on Bayesian philosophy. It allows soft information to be combined with hard data. When sampling has been performed, the analysis can be updated with new information as it becomes available, all according to Bayes´ theorem. In this framework, the value of new data will depend a lot on the prior information supplied.

The optimal level of uncertainty in a site investigation is reached when the cost for additional investigation is equal to the expected risk-reduction associated with the new information the investigation will supply. Additional information is only cost-efficient up to that point. If the cost for additional data exceeds the risk-reduction they provide, sampling is no longer cost-efficient. 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. Data worth analysis is thus the key to cost-efficient field investigation programs.

3.4.2 Terminology

In decision analysis 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 high. 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 one 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 of 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. 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 terms related to 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 one 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 analysis 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 analysis to a water resource problem with two decision variables and three objective functions.

3.4.3 Tools

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. Table 5.2 is an example of a payoff table.

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 in section 3.5.4. Most decision analysis texts include material about decision trees and their application.

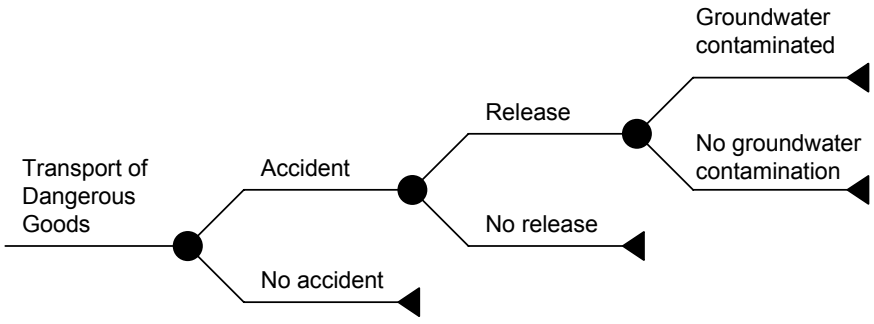


Figure 3.4 Example of an event tree for transport of dangerous goods over groundwater resources.

Decision trees are horizontal structures, which proceed with time from left to right. In a decision tree, nodes represent decisions (decision nodes), uncertain events (chance nodes), or outcome (terminal nodes) (Treeage Software, 1999). Several examples of decision trees are presented in chapter 5 and in Appendix 1. A tree without decision nodes is called an event tree. An example of an event tree is given in Figure 3.4.

Influence diagrams

A tool with growing applications in decision analysis is the influence diagram, also called relevance diagram. Figure 3.5 shows an example of an influence 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 as a tool that captures the useful features of both decision trees and influence diagrams.

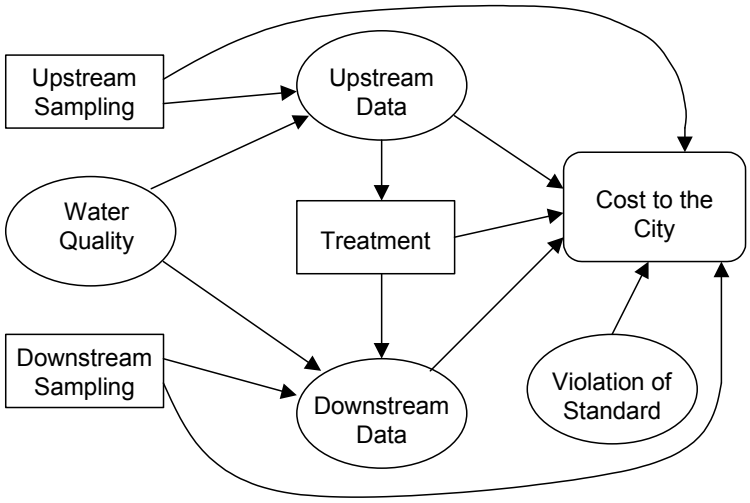


Figure 3.5 Example of an influence diagram illustrating the monitor-and-treat decision problem. Sampling is performed upstream and downstream of a discharge point in a river to study arsenic concentrations regarding sampling of river water (after Heger and White, 1997).

An influence diagram consists of chance nodes (circles or ovals), decision nodes (rectangles), value nodes (any shape except circle, oval or rectangle), 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.

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 3.5). However, no calculation of the worth of additional, uncollected data was performed.

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 2.8.3), 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 get a quantitative estimate of uncertainty from a neural network analysis. 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.

3.5 Review of approaches to data worth

3.5.1 Traditional approach: Fixed cost or fixed uncertainty

Two strategies have traditionally been used for sampling of contaminated land; (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 has 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 of primary concern. Gilbert (1987) and Borgman et al. (1996a) provide cost functions that can be applied for both strategies when the population mean should be deter-

mined. Cost functions are presented for different strategies, such as stratified random sampling, two-stage sampling, composite sampling, and double sampling.

3.5.2 Misclassification cost and loss functions

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, whereas 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, whereas the estimated concentration is below, there would be a Type II Error. Type I Errors lead to excessive remediation costs, whereas Type II Errors can bring human health risks and ecological risks. Therefore, Type II Errors are called *consumer's risk* (Gilbert, 1987). The different decision errors can be illustrated by the *decision space* in Figure 3.6.

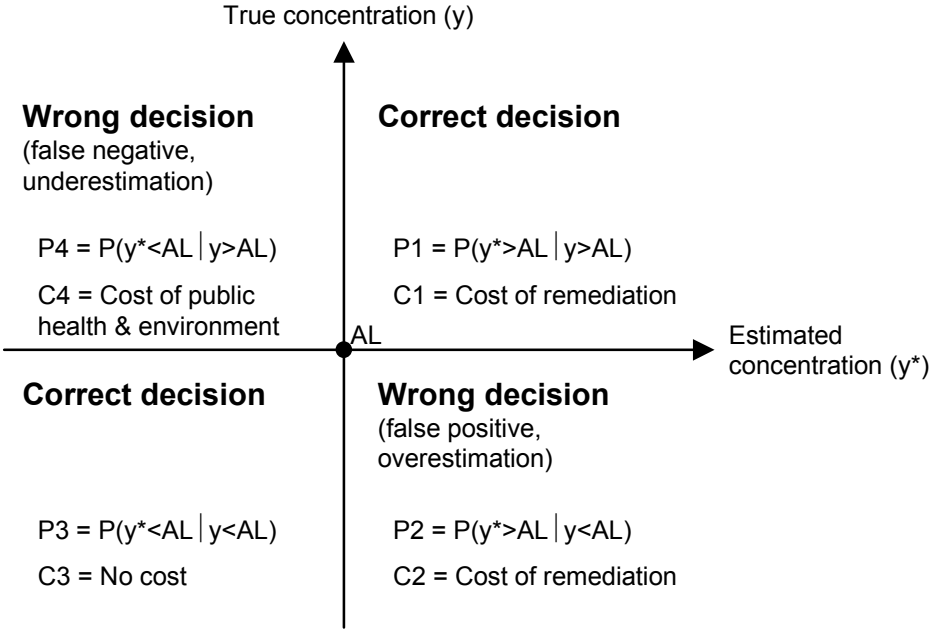


Figure 3.6 The decision space for classification of contaminated land (after Flatman and Englund, 1991). The two axis intersect at the action level AL. The probability of occurrence (P) and the cost (C) is described for each type of classification.

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 $\$_o$ and $\$_u$ 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 3.7) 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 is where the total cost is at a minimum, i.e. at the cost optimum. At this point, marginal cost equals marginal benefit. Note that the sampling cost in Figure 3.7 is linear, which of course is a simplification of reality.

The function in Figure 3.7 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.

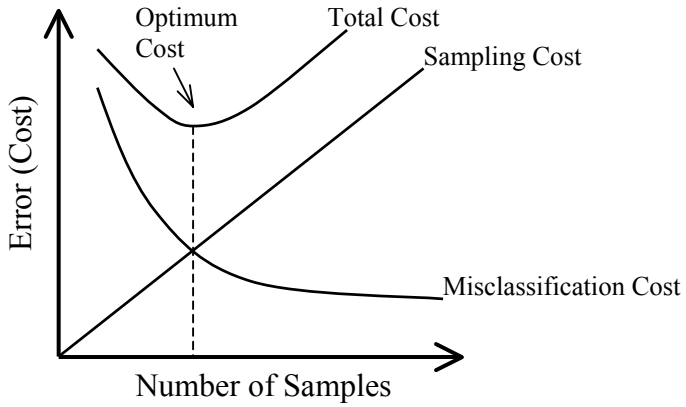


Figure 3.7 A cost optimisation graph for sampling. After J-M Rendu in Myers (1997).

Figure 3.8 presents the *V curve loss function*. 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 re-

remediation cost, whereas the right hand side of the graph is identical with the right hand side in Figure 3.8.

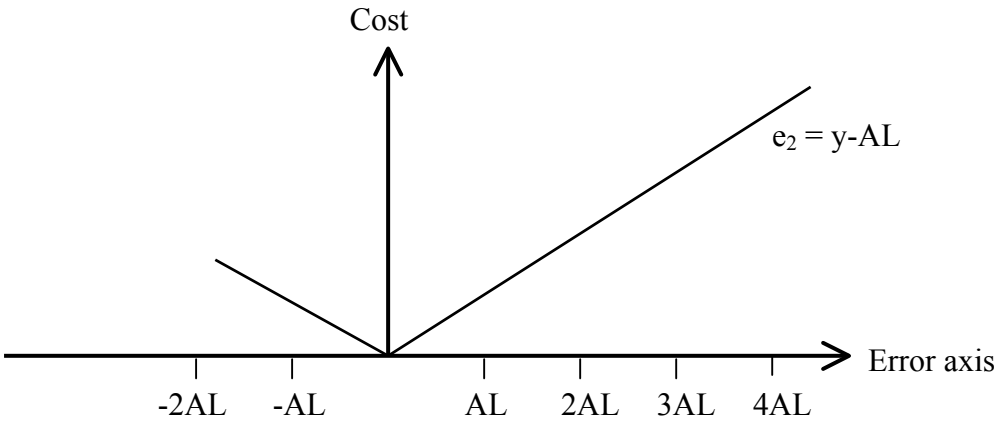


Figure 3.8 *The V curve loss function. Error (horizontal axis) is defined as the difference between the true concentration (y) and the action level (AL) for all incorrect decisions. The expected loss is a function of concentration (after Flatman and Englund, 1991).*

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.

3.5.3 Uncertainty-based stopping rules

In this section, a selection of sampling approaches with uncertainty-based stopping rules is presented, i.e. rules when additional sampling is no longer worthwhile. These stopping rules do not take sampling cost, other costs or benefits, or the economical worth of data into account. 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 % confidence level. Data worth are not considered in the approach but could probably be incorporated relatively easy. 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 data quality objectives (DQO) process. Field analysis techniques (field screening) 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 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. The sample collection ceases when there is low probability of the area outside the boundary to exceed the threshold (Stenback and Kjartanson, 2000).

3.5.4 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 additional 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 analysis 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) were 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 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:

$$TSC_i = C_i + C_r(1 - e_i)R_u \tag{3.4}$$

where C_i is the economic cost of remedial action a_i, C_r is the cost per unit risk (assumed constant), e_i is the efficiency of action a_i in reducing risk ($0 \le e_i \le 1$), and R_u is the uncontrolled risk level ($R_u \ge 0$). The concept of expected opportunity loss (EOL) is used to calculate the value of groundwater monitoring.

Freeze et al.

The framework presented by Freeze et al. (1992; 1990) is basically the framework for decision analysis we apply in this thesis. It is based on an objective function Φ similar to equation 3.1. The optimal decision from a risk perspective is achieved when the objective function is maximised. The corresponding 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., is 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.

Data worth analysis is applied as 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 program 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 programs, or (b) to be used as a stopping rule to decide when no further measurements should be made. They applied data worth analysis on a hypothetical hydrogeological problem. Some of the methods they propose to determine the probability of failure include search theory, stochastic simulation, and indicator kriging.

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 3.5.5) 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) are evaluated (see section 3.4.2). 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 used a loss function, where the loss (cost) was a function of the difference between the true contaminated area and the remediated area. Their analysis included the *expected value of including uncertainty* (EVIU) and EVPI. In a later paper Dakins et al. (1995) extended the analysis to include EVSI. They utilised Bayesian Monte Carlo methods to calculated EVSI in a preposterior 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 was 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 according to Figure 3.3. Collection of data ceases when the cost of acquisition is greater than the reduction in remediation cost the next measurement would bring.

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:

$$EVSI = Sample\ reliability \times EVPI \tag{3.5}$$

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) presented a data worth model under the acronym BUDA, Bayesian Uncertainty Development Algorithm. The purpose of the data worth model was 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. 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):

$$Total\ Cost = Characterisation\ Cost + Treatment\ Cost + Failure\ Cost \qquad (3.6)$$

In principle, this function is identical to the objective function presented by Freeze et al. (1990). The Smart Sampling approach requires explicit decision rules, such as whether or not a piece of ground is contaminated or uncontaminated. The failures are identical to the type I and type II errors described in section 3.5.2. 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). Geostatistics is applied to determine where to located new samples. The iterative procedure continues until additional samples will no longer reduce the objective function.

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.

3.5.5 Other approaches

There are some other approaches that should be mentioned in this context. One is optimisation theory. The goal of frameworks based on optimisation has often been to minimise sampling costs while estimating some quantity (James and Gorelick, 1994). Another interesting but very different approach to sampling is the *adaptive sampling* approach as described by Cox (1999). Strictly, neither of these approaches is based on decision analysis and is therefore not discussed further. A review of applications is presented in Back (2001).

4 ESTIMATION OF UNCERTAINTY IN SOIL SAMPLING

4.1 Introduction

4.1.1 Uncertainty in sample data

Traditionally, attention has been focused on reducing uncertainty in laboratory analyses, but 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 field sampling 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). A realistic assessment of the overall uncertainty in the decision-making process for contaminated land should incorporate uncertainty in both sampling and laboratory analyses. This is especially important when the concentration levels at a site are relatively close to an action level for remediation. In such situations, the decision to remediate or not might depend on whether the sampling uncertainty is taken into account or not.

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 cause, affect the remedial decision and have a negative impact on both the economy and the quality of the site cleanup.

There are different ways of approaching this problem. Traditionally, the approach has been to avoid the problem as much as possible by applying sampling standards and quality assurance/control (QA/QC) procedures to the sampling exercise. If applied successfully, such procedures may reduce errors but will rarely supply information about the uncertainty that inevitable exists, even if all precautions have been taken.

4.1.2 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 statistically sophisticated to be understood by an average scientist or engineer, or the document is too concerned with generalities such as QA/QC procedures. Another limitation 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.

Examples of the most commonly sampling objectives include:
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, and
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 literature. In practice, several other, and sometimes less stringent, sampling objectives exist. One such 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. 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. 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 question the whole population has to be analysed, but in contaminated soil problems this is impossible.

4.1.3 Geological and hydrogeological considerations

Before a sampling objective is defined it is important to consider the geological and hydrogeological conditions at the site. For example, stratification of the soil layer may lead to irrelevant results if a pre-fabricated sampling plan is followed without consideration of the geology and the geological processes that formed the site. Geological factors that influence the sampling is for example soil type, soil stratification, anisotropic and heterogeneous soil conditions, soil depth, grain size distribution etc. Contamination levels may also be closely connected to the specific surface of soil.

In estimation of sampling uncertainty, information about the grain size distribution is vital. As described in sections 4.3.5 and 4.4. It is the largest particles that will determine the size of some important sampling errors. In addition, the soil-forming processes may have lead to inhomogeneities and graded soil strata, which may affect the movement and clustering of contaminants in the soil. Therefore, geologic professional judgment must be used during conceptualisations, preparation of sampling plans, and estimation of sampling uncertainty.

Often, the sampling objective also incorporates investigation of contaminant transport. Water is often the most important transport medium and the hydrogeological conditions are therefore very important to consider. Parameters such as hydraulic conductivity, porosity, hydraulic gradient, recharge rates etc. should therefore be taken into account. A carefully developed conceptual model (see section 2.5.1) will be great importance when soil sampling plans are developed.

4.1.4 Purpose

The purpose of chapter 4 is to present a methodology for estimation of uncertainty in soil sample data. This uncertainty is important in a framework for data worth analysis because the value of new data depends on how accurate it is. Previous work is reviewed

and a methodology for estimation of uncertainty is presented. An application in the last sections illustrates the approach. It can be used as a guide for estimation of sampling uncertainty for similar sampling problems.

Although the question of sampling uncertainty applies to both sampling of groundwater and of particulate materials like soil, this chapter is aiming at soil sampling. Many aspects presented also apply to groundwater sampling but since water is a flowing medium and the particles are extremely small (suspended particles or molecules) some of the errors discussed can be neglected, whereas additional ones may be introduced.

4.2 Sampling key terms

4.2.1 Terminology

When the sampling literature is reviewed it becomes clear that there is no generic terminology. Therefore, it is wise to define the terms used in order to avoid misinterpretation. Some of the most common terms and some less common but important ones are described below. Several of them are used in the sampling theory for particulate materials described in section 4.3.5.

4.2.2 Sample

A sample is a volume of material, selected with the purpose to represent another volume (the target population, see below). One author may use the word "sample" in one way, whereas other authors may have different definitions. 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*. A *subsample* is a sample collected from another sample.

4.2.3 Increment

Another commonly used term is *increment*. It is the amount of material collected from a sample point with a device at a single time. Several increments can be mixed, forming a sample of larger mass than the single increment. If there is only one increment it is equivalent with a sample. This definition of increment is used throughout chapter 4.

4.2.4 Population

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, whereas the sampled population is the one that is examined in reality. When the target and the sampled populations are not the same, there is a potential for biased results (Chai, 1996).

4.2.5 Lot

A lot is the volume of material to be characterised. The target population can be divided into several lots and each lot sampled separately. 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 (Pitard, 1993). The problem of randomly drawing coloured marbles from a pot is an example of a zero-dimensional problem. A railroad can be approximated as a one-dimensional lot for certain sampling problems. An example of a two-dimensional lot is a contaminated site with a defined soil layer of constant thickness (Figure 4.1). Most contaminated sites can be approximated as one or several two-dimensional lots. Three-dimensional lots are the most difficult to sample and a typical example is a stockpile of excavated soil (Figure 4.1). Such problems are best handled if the stockpile can be levelled out to a constant thickness, thus transforming it into a two-dimensional lot.

The correct way of sample a two-dimensional lot is to extract a perfect cylinder of material from the lot (Pitard, 1993). The cylinder should capture the whole thickness of soil layer in order to avoid sampling errors, according to Figure 4.10 in section 4.4.4).

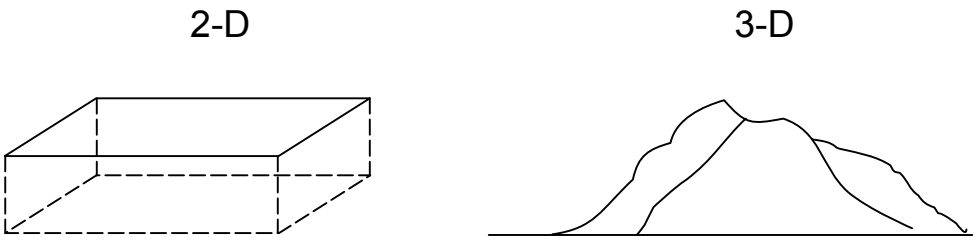


Figure 4.1 Schematic illustration of two and three-dimensional lots.

4.2.6 Sampling strategy and sampling plan

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.

4.2.7 Heterogeneity

Although used quite commonly, 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 d*istribution 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. The fundamental variability (Figure 4.8) is a result of the constitution heterogeneity.

A lot (see above) 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. It is also closely related to the *support,* described below. The distribution heterogeneity gets smaller when the support volume is increased and it is always smaller or equal to the constitution heterogeneity (Pitard, 1993). The grouping and segregation variability (Figure 4.9) is a result of the distribution heterogeneity.

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

$$h = h_1 + h_2 + h_3 \tag{4.1}$$

where:
h_1 = the *short-range heterogeneity*, typically random and discontinuous. It is influenced by the constitution heterogeneity and the distribution heterogeneity.
h_2 = the *long-range heterogeneity* (or *large scale segregation*) reflects local trends, usually non-random and non-cyclic.
h_3 = the *periodic heterogeneity* reflects cyclic phenomena. This heterogeneity is most common in flowing streams of material, such as water. For stationary material like contaminated soil, we can define h_3 to represent the *temporal heterogeneity*, i.e. variability over time.

The effect of short-range and long-range heterogeneity is illustrated in Figure 4.2.

A concept closely related to heterogeneity 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 is not included in the sampling theory for particulate materials. However, Myers (1997) integrates sampling theory and geostatistics into the concept of Geostatistical Error Management.

4.2.8 Representativeness

One objective in a typical site-investigation is to collect "representative" samples. However, the term *representative* is almost 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):

$$r^2(SE) = m^2(SE) + \sigma^2(SE) \leq r_0^2 \qquad (4.2)$$

where $m(SE)$ is the mean of the sampling selection error and $\sigma(SE)$ is the standard deviation of the same error.

4.2.9 Sample Support

When characterising a geologic or environmental area the physical size of the sample 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 4.2). The change in measured value when the volume increases is a result of heterogeneity at different scales (see *heterogeneity* above), and the REV is the volume where the large fluctuations cease (Bear, 1979).

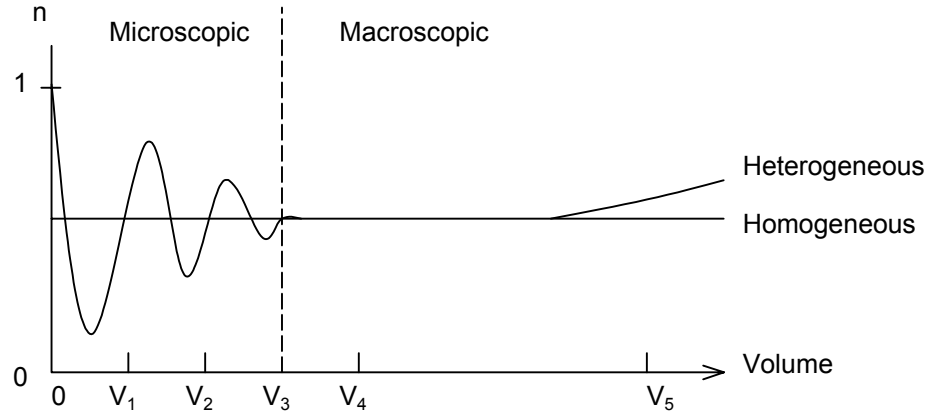


Figure 4.2 Illustration of the Representative Elementary Volume (REV) applied to porosity at the microscopic and macroscopic scale (after Hubbert in Freeze and Cherry, 1979). The REV is equal to the volume V_3.

If the support is not specified, a statement of the variability in a set 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 log-normal 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 likely be incorrect. In site-investigations of contaminated land it is common to mix results derived with different sample supports, which may lead to wrong conclusions.

Another way of describing to 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 increments in the vicinity of the designed sample point. Starks (1986) suggests taking the increments in a square grid around the sampling point, whereas Jenkins et al. (1996) used a 20 cm wide "sampling path" of circular shape. The increments can be used to produce a composite sample (see below) to save laboratory analysis 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 distribution of sample concentrations but such an approach can lead to serious estimation errors and high unexpected cost. Myers (1997) describes how the support volume 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 log-normal correction*.

4.2.10 Composite sampling

The process of compositing is a physical averaging of the material used to form the composite sample. Samples, or increments, are taken from different locations and then mixed to increase sample support (Shefsky, 1997). Under ideal conditions, the measured value for a composite sample should be equal to the arithmetic mean of the measured values from the individual increments forming the composite sample. 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.

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 heterogeneity (provided compositing is performed over large areas). 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. Naturvårdsverket (1998) presents a rule of thumb of when composite sampling can be performed, based on the heterogeneity and expressed with the coefficient of variation (CV).

4.2.11 Other key words

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 4.3.5. 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.

4.3 Previous work

4.3.1 Sampling theories

Borgman et al. (1996b) distinguish between three types of sampling theories:
1. design-based sampling (classical finite sampling theory),
2. model-based sampling, and
3. the particulate sampling theory.

The main distinction between the two first 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). Theories to determine mean concentrations by classical statistics and to detect hot spots (Gilbert, 1987) are examples of design-based sampling.

The third sampling theory shares features with both design-based and model-based sampling. The following sections will concentrate on design-based sampling and the particulate sampling theory.

4.3.2 Models of random uncertainty

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

$$s^2_{sampling} = s_1^2 + s_2^2 + \ldots s_n^2 \qquad (4.3)$$

where $s^2_{sampling}$ denotes the total random components of sampling variance, and $s^2_1 \ldots s^2_n$ are the components from sources $1 \ldots n$, 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.

Population variability
The equation above does not take population variability into account. To do this, Taylor (1996) suggests the following relationship:

$$s^2_{sample} = s^2_{sampling} + s^2_{population} \quad (4.4)$$

For large scale problems, the population variability is also called geochemical variability (Ramsey and Argyraki, 1997) or environmental variability (Clark et al., 1996). As is often the case with empirical quantities in the geologic environment, a scale problem is present. 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 4.3. 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 4.2.10).

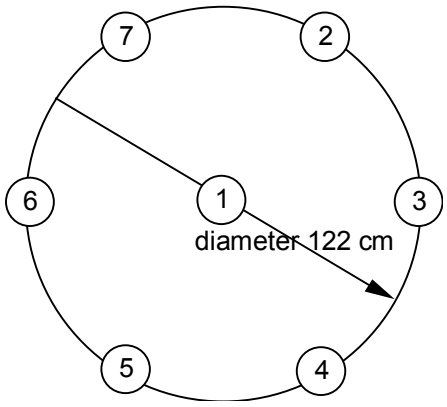


Figure 4.3 *Illustration of sampling scheme for a short-range heterogeneity study performed by Jenkins et al. (1996). Seven discrete samples (increments) were collected at each sample location.*

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 short-range variability in soil contaminated by volatile organic compounds. They found that soil properties (total organic carbon, 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 have implications for the sample support, as described in section 4.2.9.

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 4.3.3), is independent of the number of replicates (Taylor, 1996). The total uncertainty includes not only sampling and population uncertainty but also other types of uncertainty, e.g. uncertainty in laboratory analysis.

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):

$$\sigma_m^2 = \sigma_s^2 + \sigma_h^2 + \sigma_{ss}^2 + \sigma_a^2 + \sigma_b^2 \qquad (4.5)$$

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

The total variability is presented as:

$$\sigma_t^2 = \sigma_m^2 + \sigma_p^2 \qquad (4.6)$$

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):

$$s_{total}^2 = s_{geochem}^2 + s_{sampling}^2 + s_{analysis}^2 \qquad (4.7)$$

Note that geochemical variance is just another name for population variance (the only difference between eq. 4.7 and eq. 4.4 is that the analytical variance has been added in eq. 4.7). 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 4.4).

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, whereas Robust ANOVA is not (Ramsey, 1998).

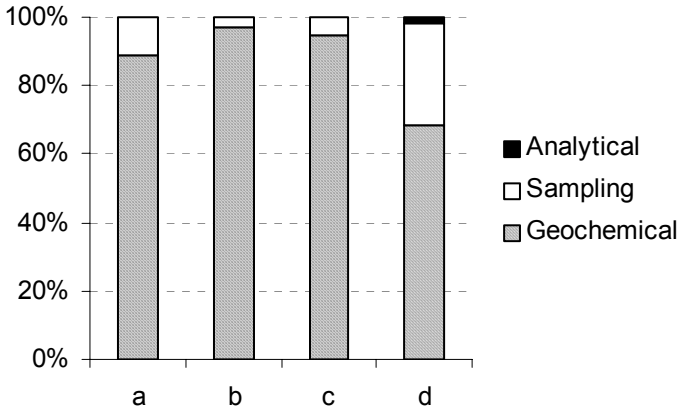


Figure 4.4 *The relative importance of analytical variance, sampling variance, and geochemical variance in a site-investigation of lead (Pb) and copper (Cu) concentrations in soil (after Ramsey et al., 1995). The analytical variance is less than 0.1 % for cases a, b, and c.*
(a) Pb data from four sampling 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.

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

$$\sigma_{\bar{x}}^2 = \frac{\sigma_c^2}{n} + \frac{\sigma_s^2}{n \cdot m} + \frac{\sigma_a^2}{n \cdot m \cdot k} \qquad (4.8)$$

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 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^2_{samp} represents small scale heterogeneity at the sampling location, whereas σ^2_s in Houseman's model represents small scale heterogeneity 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):

$$\sigma_{\bar{x}}^2 = \frac{\sigma_c^2}{n} + \frac{\sigma_a^2}{n \cdot k} \quad (4.9)$$

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

$$\sigma_{\bar{x}}^2 = \frac{\sigma_c^2}{n} \quad (4.10)$$

4.3.3 Models of systematic uncertainty

Bias in sampling

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):

$$E = \mu + b \quad (4.11)$$

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 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.

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 certain types of bias that are included in the particulate sampling theory (see section 4.3.5), e.g. bias introduced by the sampling equipment.

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 aims 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"*.

Total systematic uncertainty

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

$$B_{total} = B_1 + B_2 + \ldots B_n \qquad (4.12)$$

where B_{total} is the total systematic uncertainty, and $B_1 \ldots B_n$ are the components of uncertainty from sources $1 \ldots n$. 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.

4.3.4 Models of total uncertainty

A model of total uncertainty must include both random and systematic components of the uncertainty. Such models are not common in the environmental literature, probably because the systematic part is often ignored due to the difficulty of quantifying it. One model can be found in for example (Pitard, 1993), as described in the next section. For measurements in general, the following simple model has been suggested (Gleser, 1998):

$$m = \mu + b + e \tag{4.13}$$

where m is the measurement value, μ is the true value, b is the systematic bias and e is the random deviation from the true value. This relationship is particularly important when b and the standard deviation of e are of the same order of magnitude. In most cases b is unknown but Gleser (1998) mentions a way of establishing conservative confidence intervals based on an estimated range of possible values of b.

4.3.5 The sampling theory for particulate materials

Background

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. Other theories only cover limited 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).

The theory is applicable to the sampling of particulate materials but several of the concepts also apply to fluids (Borgman et al., 1996b). It is assumed that the contaminant is attached to particles (e.g. soil particles) or itself consists of liberated particles (e.g. lead shots). The theory was not developed with fluids in mind but it is also applicable to such problems if the contaminant exists in a solution or in suspension. In this case, the particles are extremely small (the molecules or the suspended particles).

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

Although applied to environmental sampling, it is not evident that all aspects of the theory are applicable to environmental problems. Some assumptions and limitations of the theory that can be questioned are:

- The contaminant is assumed to attach to particles, which is not always the case (e.g. non-aqueous phase liquids in the pore space in soil).
- The contaminant concentration for a particle is assumed to be correlated with its density.
- Volatile compounds are not considered in the theory.

Some of these aspects are discussed further in chapter 6.

Sampling objective

The sampling objective for the theory is estimation of mean concentrations. It can be the mean concentration of several increments or the mean concentration of a lot (see section 4.2). It is possible to use parts of the theory for other sampling objectives as well but this requires a good understanding of the theory in order to make the right assumptions.

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 Fluctuation Error* CE_1 ($=FE+GE$), the C*ontinuous Selection Error CE* ($=CE_1+CE_2+CE_3$), and the M*aterialisation 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*: $OE = TE + AE$

The relationship between the different errors is illustrated in Figure 4.5. The meaning of the basic sampling errors is not self-evident and is therefore explained below. Each sampling error consists of a random part and a systematic part. For some errors the random part is most important, whereas the systematic part (bias) may dominate for others.

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 target population have exactly the same probability of being included in the sample (Shefsky, 1997). This can also be expressed as *the equiprobability rule*. Following this rule in all aspects of the sampling will result in an unbiased sample.

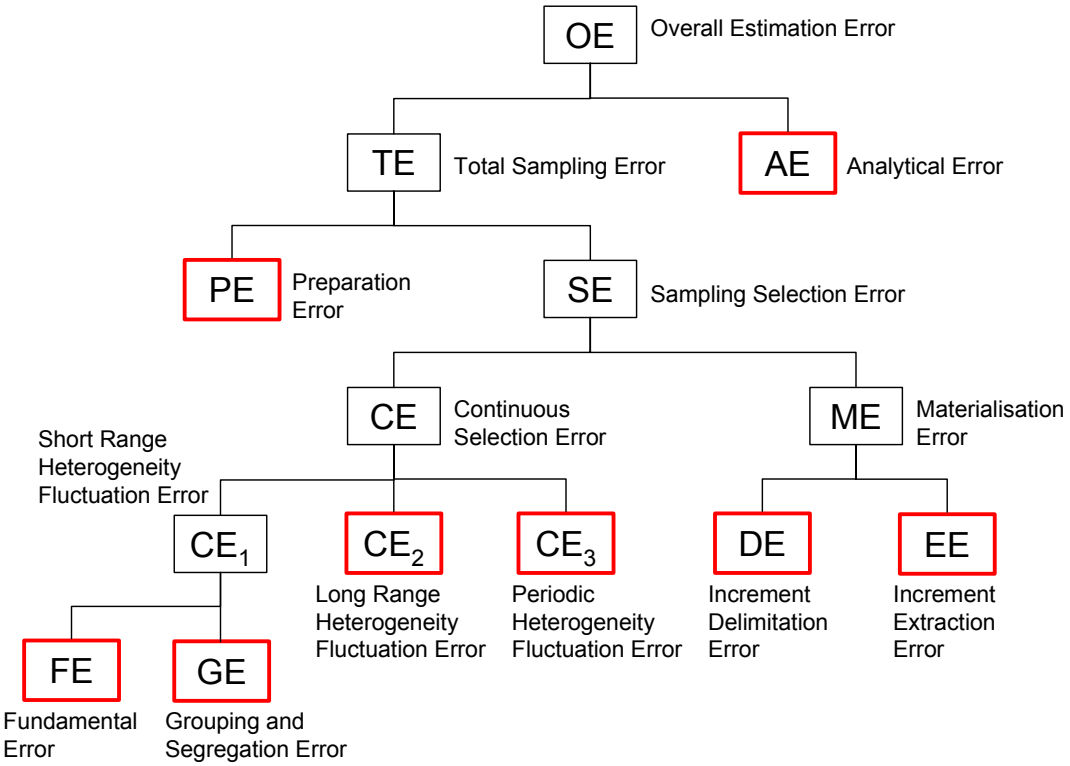


Figure 4.5 *Illustration of the relationships between different types of errors, according to Pierre Gy's sampling theory (the eight basic errors in red).*

The Fundamental Error (FE)

The fundamental error is a statistical consequence of the particulate nature of geologic samples (Shefsky, 1997), see also the Representative Elementary Volume in Figure 4.2. It is caused by the constitution heterogeneity described in section 4.2.7, i.e. by the range of particle sizes in the medium and the fact that the different particle sizes contain different amounts of the pollutant of interest. Large particles have a much greater mass than small ones and will therefore contribute more to the mean concentration in the sample. If the sample mass is small the concentration will depend very much upon how many large particles there happens to be in the sample, leading to a large fundamental error. Also, the contaminant itself can exist as large particles such as lead shots or paint chips, which may result in very large fundamental errors, as illustrated by Shefsky (1997). The principle leading to the fundamental error is illustrated in Figure 4.8. The fundamental error is the only error that can be assessed independently of the sampling method (Borgman et al., 1996b).

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 Grouping and Segregation Error (GE)
The grouping and segregation error results from the distribution heterogeneity described in section 4.2.7. The heterogeneity may be in density, particle size, shape, adhesion, cohesion, magnetism, affinity for moisture etc. so that the particles come together by groups (Flatman and Yfantis, 1996). A typical example of this error is the segregation of particles that occurs during transport of dry food, where small particles gather at the bottom and larger particles at the top. This error occurs at the scale of the sample support. Grouping and segregation of particles within the volume the sample is supposed to represent will lead to variation in concentration (Figure 4.9).

The Long-range Heterogeneity Fluctuation Error (CE_2)
This error is due to the long-range heterogeneity h_2 described in section 4.2. It is generated by local trends. The variance of this error can be quantified by the variogram of geostatistics (see section 2.7.3).

The Periodic Heterogeneity Fluctuation Error (CE_3)
This error is introduced by cyclic phenomena. It is a non-random error generated by heterogeneity h_3 (see section 4.2).

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 4.2) 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 often 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. Incorrect shape of the delimitation tool (sampling device) can lead to unequal probability of 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, especially in coarse-grained material.

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).

The Preparation Error (PE)
All the previously described errors are selective whereas 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, as listen in section 4.4.5.

The Analytical Error (AE)
This error is not part of the sampling and it does not include the subsampling or preparation of analytical samples. Such errors are instead included in the total sampling error *TE*. This may contradict how the uncertainty in chemical analyses is reported by the laboratory.

The Overall Estimation Error (OE)
The overall estimation error is quantified as the sum of all sampling and analytical errors. Usually, several stages of sampling (subsampling) and preparation are performed prior to analysis and errors introduced in all these stages must be considered. How the quantification is performed is described in section 4.4.7.

4.3.6 Other sampling uncertainty models

The DQO process (U.S. EPA, 1994) is a planning approach to develop sampling designs for data collection activities that supports decision-making. It is a 7-step procedure where uncertainties are addressed in step 6. The total study error is broken down into sampling design errors and measurement errors according to Figure 4.6. These can be compared to the different errors in the particulate sampling theory. Such a comparison is presented in Table 4.1. The terms used in the DQO process is easier to understand and could be used to present the particulate sampling theory in a more "user-friendly" way.

Table 4.1 Error types in the DQO process (U.S. EPA, 1994) *and the corresponding errors according to the particulate sampling theory* (Pitard, 1993).

DQO process	**Particulate sampling theory**
Total Study Error	Overall Estimation Error (OE)
Sampling Design Error, Inherent Variability	Continuous Selection Error (CE)
Measurement Error	Sum of Materialisation Error, Preparation Error and Analytical Error (ME + PE +AE)
Physical Sample Collection Error	Materialisation Error (ME)
Sample Handling Error	Preparation Error (PE)
Analysis Error	Analytical Error (AE)

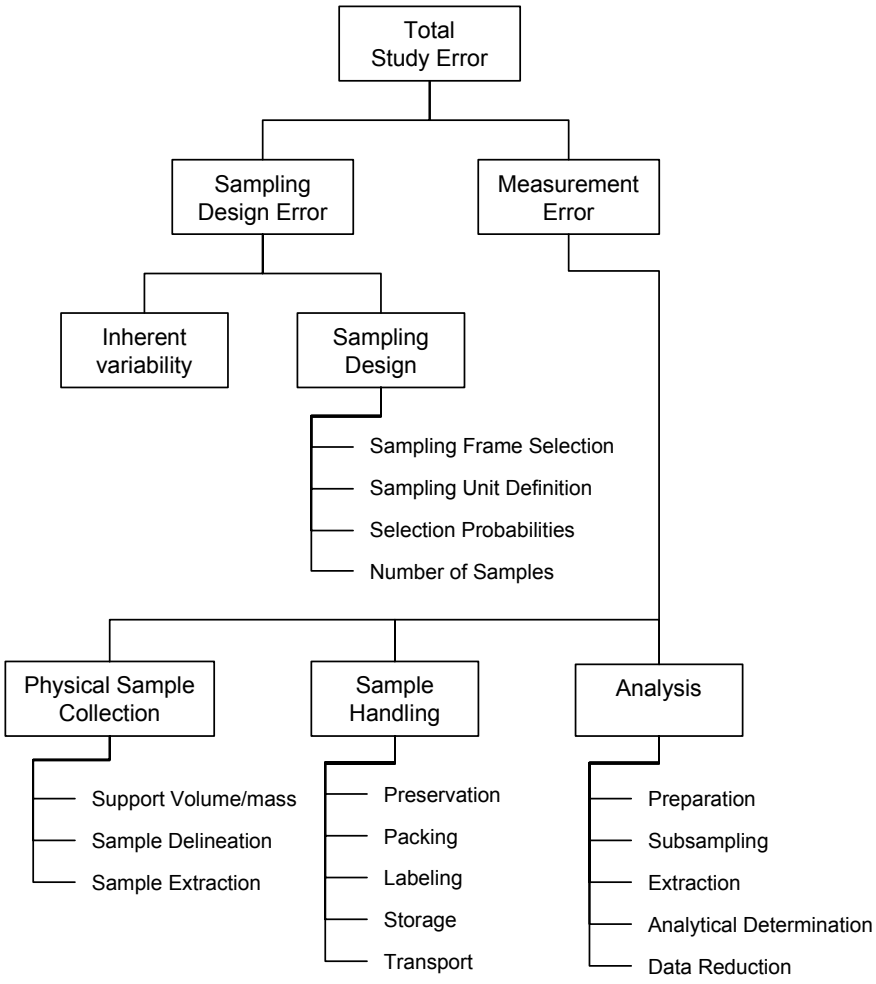


Figure 4.6. An example of how Total Study Error can be broken down by components (after U.S. EPA, 1994).

4.4 An approach for estimating sample uncertainty

4.4.1 Terminology

Surprisingly little can be found in the literature about quantitative estimation of uncertainty in sampling of contaminated soil, taking all types of uncertainty into account. Although most of the models in sections 4.3.2 – 4.3.3 are conceptually correct, they all fail to take all different types of uncertainty in soil sampling into account. The particulate sampling theory in section 4.3.5 presents a more complete picture of the sampling problem.

The approach to the problem taken in this thesis is to use the particulate sampling theory as a foundation for a slightly modified version of the theory. The modifications are made in order to:

- make the theory more applicable to sampling problems of contaminated soil,
- present the different types of uncertainties in a way that agrees with the purpose of prior estimation of sample uncertainty,
- present the theory in a way that is easier to understand.

Therefore, the names of some of the uncertainties have been slightly changed and some types have been defined differently. An illustration of the relationships between different types of sampling uncertainties is given in Figure 4.7. The names and definitions used are listed below. They can be compared to the names in Table 4.1. Generally, the term "error" has been avoided. Instead, either "variability" or "uncertainty" is used. The reason for using "uncertainty" instead of "error" is that the purpose of the methodology primarily is prior estimation of uncertainty. Errors on the other hand, occur at the moment sampling actually is performed (see section 2.2).

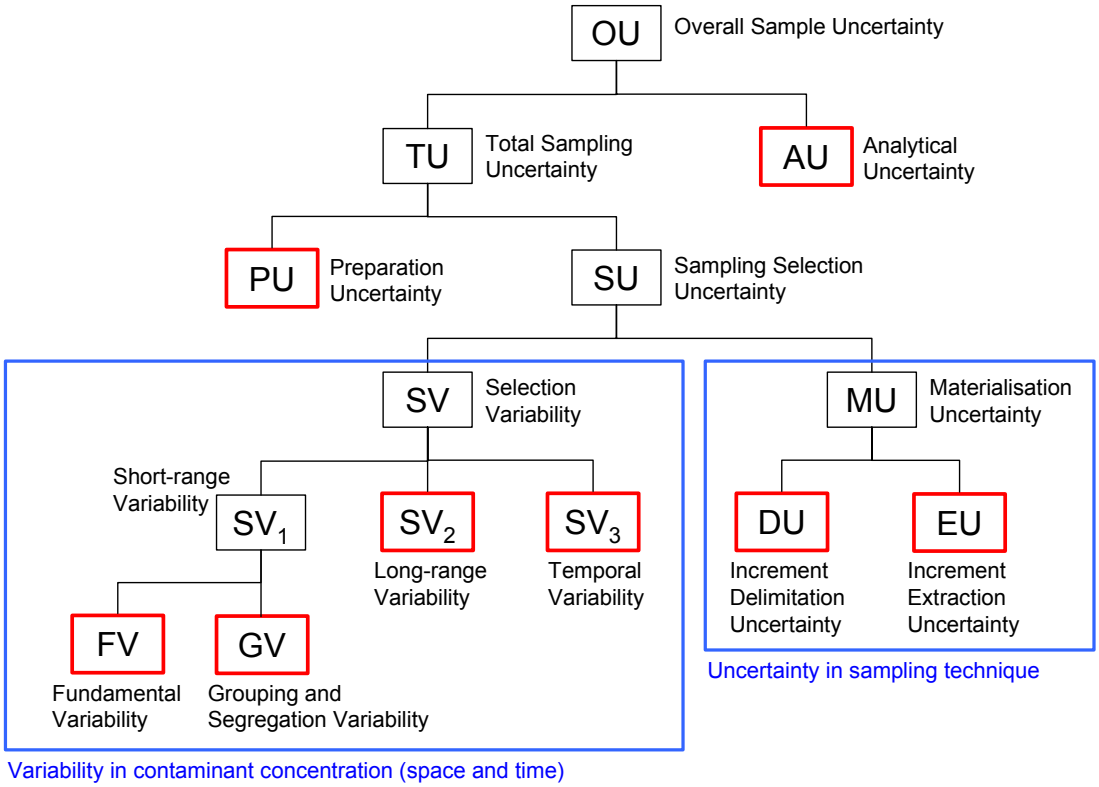


Figure 4.7 *Illustration of the different types of uncertainties in the presented approach for estimation of sample uncertainty (the eight basic types of uncertainty in bold red).*

Fundamental Variability (FV)
The fundamental variability (FV) corresponds to the fundamental error in the particulate sampling theory. Variability is a better name than error since this type of uncertainty aims at small scale variability. Fundamental variability is also the name used in a proposed ISO-standard (International Organization for Standardization) for sampling of stockpiles.

Grouping and Segregation Variability (GV)
The name has been changed for the same reason as above.

Short-range Variability (SV_1)
The name has been chosen for the same reason as above. The short-range variability is the sum of FV and GV. It is the first type of Selection Variability (SV), i.e. variability that depends on the selection of a sample.

Long-range Variability (SV_2)
The name has been chosen for the same reason as above. It is the second type of Selection Variability.

Temporal Variability (SV_3)
The sampling error CE_3 in the particulate sampling theory is of minor interest since soil is stationary and do not move much. However, temporal variability can be important for some types of problems, i.e. leaching due to infiltrating precipitation, and is therefore included instead. It is the third type of Selection Variability.

Selection Variability (SV)
The selection variability is the sum of SV_1, SV_2 and SV_3. It corresponds to the continuous selection error in the particulate sampling theory. The Selection Variability is a result of the heterogeneity at different scales at the site.

Increment Delimitation Uncertainty (DU)
This type of uncertainty is the same as in the particulate sampling theory.

Increment Extraction Uncertainty (EU)
This type of uncertainty is the same as in the particulate sampling theory but there are also important additional aspects:

1. If the sampling equipment is not clean it may contaminate the increment, i.e. contaminants that do not belong to the increment will be included.
2. The sampling operation itself may lead to a loss of the contaminant in gas phase, i.e. contaminants belonging to the increment will not be included. This type of error can be very important for soils containing VOCs.

These types of errors have a non-zero mean and therefore the systematic component is most important.

Materialisation Uncertainty (MU)
This uncertainty (or error, depending on how we choose to view the problem) is the sum of DU and EU.

Sampling Selection Uncertainty (SU)
This uncertainty is the sum of SV and MU. This indicates that the sampling uncertainty is the sum of the uncertainty introduced by the selection of the sample and by taking out the physical sample.

Preparation Uncertainty (PU)
This uncertainty is the same as in the particulate sampling theory. Note that contamination and loss errors that occur during the actual sampling are included in the Increment Extraction Uncertainty.

Total Sampling Uncertainty (TU)
The Total Sampling Uncertainty is the sum of SU and PU, i.e. equivalent to the particulate sampling theory.

Analytical Uncertainty (AU)
The Analytical Uncertainty is the same as in the particulate sampling theory.

Overall Sample Uncertainty (OU)
The Overall Sample Uncertainty corresponds to the Overall Estimation Error in the particulate sampling theory.

In total, 8 basic types of uncertainty and 6 groups of uncertainty (including variability) are now defined.

4.4.2 Sampling objective and underlying assumptions

In all questions related to sampling it is good practise to specify the sampling objective and the underlying model assumptions. The objective of the proposed methodology is to estimate the uncertainty in individual sample data. We define sample uncertainty to include uncertainty in sampling, preparation and chemical analysis. Later in the thesis (chapter 5), we will use the estimated uncertainty in sample data for determining the mean concentration in the lot.

We choose to characterise our contaminated site as a two-dimensional lot, according to section 4.2. This two-dimensional lot is our target population for sampling. Each sample (or increment) must comprise the whole thickness of our two-dimensional lot. Note that the two-dimensional approach is only applicable for the characterisation of contamination levels in the soil, not for other problems like contaminant transport etc. The thickness of the lot should be quite small and the lot should not comprise more than one type of geological unit. If the geological conditions vary significantly, the site should be divided into several two-dimensional lots.

The first sampling stage is in the field. After this stage there are often one or more additional sampling stages, in the field or in the laboratory. Errors will be introduced in all sampling stages and different assumptions may need to be made for each stage (see section 4.5).

The sampling and analytical uncertainties presented in the next sections all consist of two parts:

1. A random part, specified by the coefficient of variation (the standard deviation divided by the mean).
2. A systematic part, which is defined as the mean of the error, i.e. a positive systematic error means that the estimated concentration is systematically too high.

Each part must be estimated for each type of uncertainty in each sampling stage, in order to make a complete estimation of the uncertainty. This will be illustrated by the example of application in section 4.5.

4.4.3 Selection Variability

The selection variability is the sum of spatial and temporal variability. Spatial variability must be considered for all problems where a single sample is supposed to represent a larger volume of soil than the sample itself. For material that has been homogenised, only the short range variability needs to be considered, i.e. the Fundamental Variability and the Grouping and Segregation Variability. If a sample is supposed to represent a larger area of soil, the Long-range Variability may also need to be considered. The Temporal Variability should be considered in problems where changes in concentration is expected over time.

The Fundamental Variability

The Fundamental Variability is a statistical consequence of the fact that there is a difference in contaminant concentration between the individual particles in the sample. If all particles had the same concentration there would be no Fundamental Variability. As illustrated in Figure 4.8, the concentration in the sample depends on how many contaminated particles will be included in the sample. The larger the sample is, the more particle sizes will be represented in the sample, leading to a smaller Fundamental Variability. Thus, the size of the Fundamental Variability is directly influenced by the sample size. Large particles are more difficult to represent than small ones since they contain more mass, and therefore it is the large particles that set the limit for the Fundamental Variability.

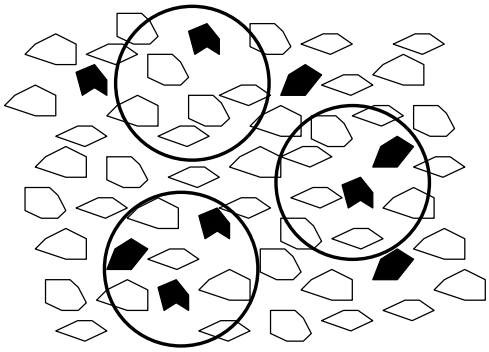


Figure 4.8 Illustration of the principle behind the Fundamental Variability (after Pitard, 1993). The number of contaminated particles in the sample has a random component depending on the exact location of the sample at the particle scale, i.e. the Fundamental Variability.

From a geological point of view, the Fundamental Variability is a result of the soil-forming processes and the individual soil particles ability to attract contaminants. Most important is the grain size distribution and the specific surface of the soil. Heterogeneity at the particle scale may also be important, i.e. the amount of large particles contributing to the Fundamental Variability.

The relative standard deviation, or the coefficient of variation, of the Fundamental Variability, CV_{FV}, can be expressed by the following relationship (after Pitard, 1993):

$$CV_{FV} = \sqrt{\left(\frac{1}{M_S} - \frac{1}{M_L}\right) \cdot c \cdot l \cdot f \cdot g \cdot d_{95}^3} \qquad (4.14)$$

where M_S is the sample mass [kg], M_L is the mass of the lot to be characterised [kg], c is the mineralogical factor [kg/m^3], l is the liberation factor [dim.less], f is the particle shape factor [dim.less], g is the particle size range factor (or granulometric factor) [dim.less], and d_{95} [m] is defined as the opening of a square mesh retaining no more than 5 % oversize particle mass. Equations and practical advice for estimating the factors c, l, f and g are presented by Minkkinen (1987), Mason (1992) , Pitard (1993), and Borgman et al. (1996b). The main difficulty for applying this equation is to estimate the liberation factor.

Equation 4.14 assumes that we have a certain amount of information about the soil we are trying to characterise, but for prior analysis in environmental sampling this may not be the case. Therefore, in many situations another approach can be used. For a sample to be representative of a soil volume, it must at least be representative of all particle size fractions (Pitard, 1993). In this case the coefficient of variation of the Fundamental Variability can be estimated by the following relation (after Pitard, 1993):

$$CV_{FV} = \sqrt{\left(\frac{1}{M_S} - \frac{1}{M_L}\right) \cdot f \cdot \rho_m \cdot \left[\left(\frac{1}{a_{Lc}} - 2\right) \cdot d_{Lc}^3 + g \cdot d^3\right]} \qquad (4.15)$$

where ρ_m is the density of the soil matrix [kg/m^3], a_{Lc} is the proportion [kg/kg] of the particle size class L_c in the lot, and d_{Lc} is the average particle size [m] in the particle size class L_c. A good approximation of the particle shape factor is $f = 0.5$ (Minkkinen, 1987; Pitard, 1993). With the critical size fraction as 5 % oversize, then by definition we have $a_{Lc} = 0.05$ and $d_{Lc} = d_{95}$. Since a_{Lc} is small, the term $g \cdot d^3$ becomes negligible (g is about 0.25) (Pitard, 1993). An approximation of equation 4.15 can now be formulated (after Pitard, 1993; Ramsey and Suggs, 2001):

$$CV_{FV} \approx 3 \cdot \sqrt{\left(\frac{1}{M_S} - \frac{1}{M_L}\right) \cdot \rho_m \cdot d_{95}^3} \qquad (4.16)$$

The term $1/M_L$ can be neglected if the sample mass is small in comparison to the lot to be sampled. However, this assumption may be too conservative for certain subsampling problems, such as sample collection from an auger (Figure 4.13) or splitting of samples. Therefore, the term $1/M_L$ is maintained in equation 4.16 in contrast to the equation presented by Ramsey and Suggs (2001) in order not to overestimate the variability.

The Fundamental Variability must be considered in primary sampling as well as in all subsampling stages. It can be deduced from equation 4.16 that it is important to consider the maximum particle size and the weight of each subsample. The maximum particle size, as defined above, is often unknown prior to sampling but it is usually possible to make a reasonable estimate based on the geological setting and inspection of the site.

Note that the factor $\sqrt{10}$ used by Ramsey and Suggs (2001) has been replaced by a factor 3 in equation 4.16 based on derivation from equation 4.15. It should also be made clear that if several increments are combined to a primary sample, M_s should be the total weight of all increments.

An additional equation for estimation of the fundamental variability is used in an ongoing standardisation work by CEN and ISO. This equation has similarities with equation 4.14 and is based on the binomial distribution:

$$CV_{FV} = \sqrt{\frac{1}{6}\pi \cdot d_{95}^3 \cdot \delta_m \cdot g \cdot \frac{1-p}{M_s \cdot p}} \qquad (4.17)$$

In this equation, p is the fraction of soil particles that contain the contaminant of interest. All of equations 4.14, 4.16 and 4.17 give comparable results but equation 4.14 requires estimation of the liberation factor l, whereas the fraction p is required in equation 4.17. These two parameters can be difficult to estimate.

When the Fundamental Variability is low it is generally normally distributed (Pitard, 1993). In this case the mean of the error is very small and can be assumed to be zero. However, large Fundamental Variability is more Poisson distributed (Pitard, 1993) and other ways to estimate the uncertainty must be used in such cases. An example of such a situation is when the contaminant of interest consists of particles of a more or less pure mineral, such as lead shots or paint chips. Whether or not one of these particles is present or not in the sample will have a significant impact on the concentration, leading to a Poisson distribution.

If the average number of contaminant particles in a sample is low, less than approximately four or five, the error distribution will have a nonzero mean and be skewed towards negative errors (Myers, 1997; Pitard, 1993). On the other hand, when the average number of contaminant particles exceeds five the Poisson distribution tends to go towards a normal distribution (Myers, 1997). For the error to follow a normal distribution more accurately, Pitard (1993) suggests that the Fundamental Variability should not exceed 16 %. To reach this, at least 40 contaminant particles should be present in the sample on average.

The Poisson distribution has the special property of having a variance equal to its mean. In the case of a contamination problem the variance and the mean is also equal to the average number of contaminant particles λ in a sample. If the contaminant particles have been spread evenly in the lot, a rough estimate of the average number of contaminant particles λ in a sample can be expressed as:

$$\lambda = N \cdot \frac{V_S}{V_{tot}} \qquad (4.18)$$

where N is the total number of contaminant particles at the site, V_S is the soil volume of the sample, and V_{lot} is the total soil volume of the lot to be characterised. The Fundamental Variability is Poisson distributed with a coefficient of variation calculated as:

$$CV_{FV} = \frac{\sigma(C_S)}{\mu(C_S)} = \frac{\sqrt{\lambda}}{\lambda} = \frac{1}{\sqrt{\lambda}} \qquad (4.19)$$

where $\sigma(C_S)$ is the standard deviation and $\mu(C_S)$ is the mean of concentration in the sample *S*.

The Grouping and Segregation Variability

The Grouping and Segregation Variability is a consequence of the distribution heterogeneity. As illustrated in Figure 4.9, different grouping and segregation patterns of particles will lead to different sampling errors. From a geological point of view, the Grouping and Segregation Variability is closely connected with micro-scale heterogeneities in the soil, e.g. irregularities that may control the infiltration pattern for organic liquids like oil. This may lead to clustering of contaminants along micro-pathways in the soil, leading to rapid concentration changes over short distances (Myers, 1997). Graded soil strata may also lead to accumulation of contaminants in certain layers. Of special importance is the organic carbon content, because many contaminants strongly adsorbs to organic matter.

The Grouping and Segregation Variability, GV, is related to the Fundamental Variability in the following way (Pitard, 1993):

$$CV_{GV}^2 = \gamma \cdot \xi \cdot CV_{FV}^2 \qquad (4.20)$$

where CV_{GE} and CV_{FE} are the coefficients of variation (relative standard deviation) for the two types of variability, γ is a grouping factor, and ξ is a segregation factor. In practice it would be extremely difficult to estimate the product in equation 4.20 and it is never done (Pitard, 1993). However, based on a large amount on experiments performed between 1960 and 1975, Gy reached the conclusion that we do not introduce an important uncertainty in the estimation of CV_{GV} if we assume that $\gamma\xi \approx 1$. It is important to note that this approximation can only be used if care has been taken to minimise the materialisation and preparation errors (Pitard, 1993). However, it can be questioned if this approximation is valid for organic substances in liquid form (APLs, aqueous phase liquids) that tend to cluster.

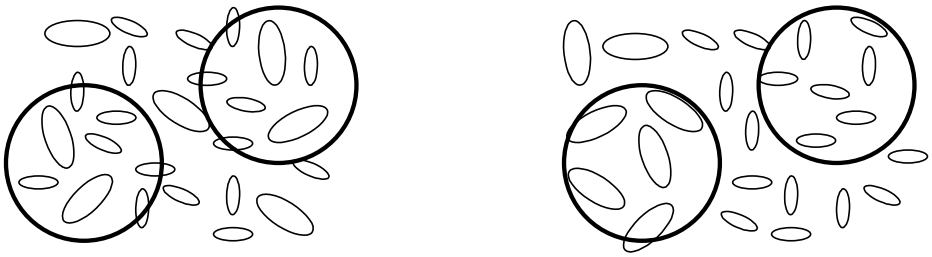


Figure 4.9 Illustration of the principle behind the Groping and Segregation Variability (after Myers, 1997). Two samples are located slightly different in two materials; low segregation (left) and high segregation (right).

In practice there are three ways of minimising the Grouping and Segregation Variability:

1. To minimise the variance of the Fundamental Variability CV_{FE}^2.
2. To minimise the grouping factor γ_s. This is performed by taking as many and as small increments as practically possible and forming the sample from these increments (Pitard, 1993). In practice, this may be the most effective way.
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 uncertainty grows larger without bounds if the size of the 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).

As described previously, a sample is often composed of several increments. The more increments that are collected, the less will the influence of the Grouping and Segregation Variability be. If the above approximation of $\gamma\xi \approx 1$ holds, the following expression can be formulated (after Myers, 1997):

$$CV_{GV} \approx \frac{CV_{FV}}{N} \tag{4.21}$$

where N is the number of increments making up the sample. According to this equation, taking 10 increments to form a sample will reduce the grouping and segregation variability one order of magnitude. Whether or not the assumption of $\gamma\xi \approx 1$ holds, the advice is to collect at least 10 random increments per sample. In this case, the following important relationship is valid (Pitard, 1993):

$$CV_{GV}^2 \leq CV_{FV}^2 \tag{4.22}$$

The Long-range Variability

The Long-range Variability is generated by spatial trends in contaminant concentration. Typically, this error is introduced when we want to use one sample to characterise a certain soil volume at a larger scale. Thus, the Long-range Variability increases when we increase the soil volume the sample is supposed to represent. The smaller soil volume the sample is supposed to represent, the smaller this error will be. If a sample is supposed to represent only the sample point there will be no Long-range Variability.

It is important to note that the variability depends on the size of the sample support. The Long-range Variability is reduced when the size of the sample increases. When the sample size approaches the size of the volume it is supposed to represent, the Long-range Variability diminishes.

The Long-range Variability is often handled by geostatistics. However, it is problematic to make prior estimates of it, especially for problems where only a one samples is used

to characterise a large area. For such problems, the Long-range Variability may well be the dominating type of uncertainty. The size of the variability is a function of soil type, type of contaminant, mode of contamination (i.e. how the contamination event occurred), time, sample volume etc. For excavated and well-mixed soil in a stockpile the Long-range Variability may be small. If a preliminary field study has been undertaken at the site, data from this can be used to assess the uncertainty. If this is not the case, the estimation must be based on experience from similar sites and professional judgement.

From a geological point of view, the Long-range Variability may be affected by changes in type or constitution of the soil over the area of interest. For example, a high silt content or organic carbon content in a certain part of the site may result in higher contaminant concentrations than in other parts of the site.

The Temporal Variability
The temporal variability is the concentration changes in the soil over time. For many contamination problems it may be neglected, e.g. for sites contaminated by non-volatile heavy metals under stable chemical conditions and no leaching, because no changes of concentration is expected. For other problems temporal variability may be of interest, primarily when organic or radioactive contaminants are of concern. Leaching due to infiltration of precipitation is another reason for considering Temporal Variability.

The Temporal Variability often has a non-zero mean. Decreasing concentration can be the result of several mechanisms, such as chemical reactions, biodegradation, loss of volatile compounds, leaching by precipitation, erosion, wind transport etc. Increasing concentrations can result from chemical reactions, on-going contamination by human activities, air deposition, wind transport etc.

Comment
Although all uncertainties described in this section are due to spatial and temporal variability, they can all be reduced by collecting more data. This may seem to contradict the definition of variability given in section 2.4, where it is stated that variability cannot be reduced by further investigations. However, the variability itself is not reduced by taking more samples but the uncertainty about the true size of the variability is.

4.4.4 Materialisation Uncertainty

The Materialisation Uncertainty is made up by two parts; (1) the Increment Delimitation Uncertainty and (2) the Increment Extraction Uncertainty. Both these Uncertainties often have a significant systematic component, i.e. these error types introduce a bias. As described in section 4.3.5, the Increment Delimitation Uncertainty results from an incorrect shape of the volume delimiting the increment or sample. It is closely related to the Increment Extraction Uncertainty, which results from an incorrect extraction of the increment or sample. 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, otherwise there will be an error. A slightly different definition must be used for contaminants occurring in liquid or gas phase in the pore space between soil particles, but the original theory for sampling of particulate materials does not consider such aspects.

For contaminated land problems, the Increment Delimitation Uncertainty is often large because equipments used are not constructed with this error in mind. For a two-dimensional sampling problem, a correctly delimited sample should consist of a cylinder extending all the way through the soil layer that the sample is supposed to characterise. If the shape deviates from a perfect cylinder there will be a difference in how the soil layer is represented in the sample, which introduces a delimitation error (Figure 4.10). Estimation of this uncertainty must be made specifically for the sample equipment used and how it is applied. Similarly, the size of the Increment Extraction Uncertainty depends very much on the equipment used. Usually, the systematic part of the uncertainty will be the most important, and it is an especially important source of sampling bias in coarse-grained materials.

Pitard (1993) states that "*when drilling through unconsolidated materials, delimitation and extraction biases are likely to take place. Vibrations and percussions are likely to alter the natural particle size distribution and vertical segregation is increasing. Pressure transmitted by the drilling machine is likely to make the coarsest fragment escape laterally unless the hole is of a large diameter*." Myers (1997) points out that most of the sampling devices available for environmental sampling are incorrect. Examples of such devices are augers, thiefs, 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).

In Sweden, the most common way to extract soil samples in the field is by drill augers. Several different types of errors may be introduced by such equipment, such as:

1. The shape of the drill auger is often not a perfect cylinder, i.e. the lower part of the auger often has a smaller diameter than the upper part. In many cases, the lowest part of the auger will extract no soil at all. This will cause the lower parts of the soil layer to be underrepresented in the soil sample.
2. Horizontal and vertical segregation of particles may result due to the rotation of the auger flights and because the delimitation "cylinder" is open, both horizontally and vertically. This may cause larger particles to be underrepresented in the increment.
3. Incorrect extraction of the delimitation "cylinder" may result due to horizontal movements of the auger during drilling or inaccuracy in the extraction in the vertical direction.
4. Soil particles that do not belong to the delimitation "cylinder" will be extracted when the drill auger is passing other soil layers. When the auger is lowered to a new sampling level, soil from the upper soil layers will be pushed down, and when the auger is lifted of, soil from upper layers will attach to the auger or soil in the auger may fall of. This may lead to so called "cross contamination" between different soil layers.
5. Contamination of the increment may occur if the drill auger has not been cleaned properly, i.e. contaminants that do not belong to the increment will be included.
6. Loss of volatile compounds may occur during drilling or when the increment is taken out. The loss occurs when the contaminated soil is disturbed and exposed to the atmosphere.

The first type of error can lead to a significant bias if the contaminant is concentrated to a certain part of the soil layer. A higher concentration in the upper part of the soil layer than in the lower will lead to a positive bias if the lower part is underrepresented, and

vice versa (Figure 4.10). An expression of the expected bias, $E[B]$, can easily be derived:

$$E[B] = E[C_s] - E[C_c] = E[C_s] - \frac{M_s \cdot E[C_s] + M_e \cdot E[C_e]}{M_s + M_e} = \frac{M_e}{M_c} \cdot (E[C_s] - E[C_e]) \quad (4.23)$$

In the equation, C_s is the mean concentration in the extracted sample, C_e is the mean concentration in the excluded soil, C_c is the mean concentration in the perfect soil cylinder, M_s is the mass of the soil sample, M_e is the mass of the excluded soil, and M_c is the total mass of the soil cylinder. If the contaminant is randomly distributed over the depth of the soil layer there will be no bias, only an increase in variance due to the reduction in sample support volume.

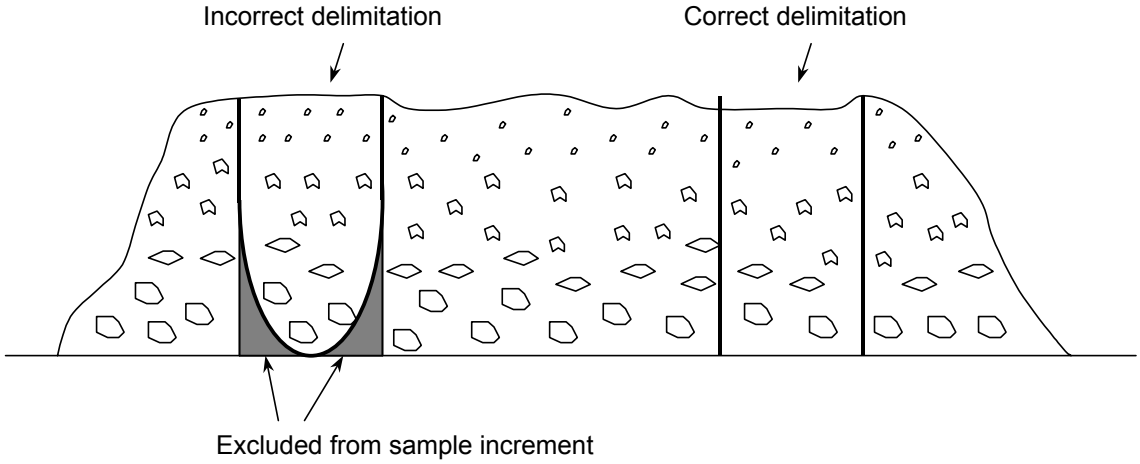


Figure 4.10. Illustration of the Increment Delimitation Error (after Pitard, 1993). Less soil from the lower part of the lot is included in the sample increment and consequently a bias is introduced.

The second type of error depends on the soil particle distribution and how the drilling operation is performed. This error is probably negligible for relatively homogeneous and fine soils. On the other hand, the error may be significant for unsorted soils and soils containing large particles. If the soil contains large particles they will be underrepresented in the sample, resulting in a positive bias because the contaminant of interest often adheres to the smaller particles.

The third type of error may be of importance in situations where the small-scale heterogeneity is large and inaccurate equipment or drilling procedures are used.

The fourth error type is a major drawback of soil sampling with drill augers. Estimation of this error is difficult, although attempts have been made (Karlström, 2001). Good sampling procedures can minimise this error. Soil that attaches to the auger when it is lifted up can often be removed before the sample is collected from the auger. On the contrary, it is impossible to separate upper soil that is pushed down into the hole. If the soil is dry the problem can be soil falling of the auger flights.

The fifth error can be eliminated by following good cleaning and drilling hygiene procedures.

Error number six only occurs for volatile compounds and is difficult to estimate, but it is expected to be larger for small increments than for larger ones. This is because small increments are more exposed to the atmosphere (relative to their mass) than larger ones, making it easier for volatile compounds to escape. However, more important may be the equipment and the technique used to take out the increment from the soil. The more the sampling device disturbs the soil, the larger error can be expected.

Other types of sampling equipment may produce other errors than those described above. Estimation of the Materialisation Uncertainty must be based on the specific sampling problem; the sampling equipment, soil type, type of contaminant, procedures etc. Often, a reliable quantification of the uncertainty is not possible but an interval of expected bias or some type of qualitative estimation should be made. At least, it is always possible to discuss each type of Materialisation Uncertainty individually and assess if the bias is positive or negative.

The geology of the site most be considered in order to control the Materialisation Uncertainty. Delimitation of an increment without consideration of the geology may lead to problems, or samples that are not representative of the site. During extraction of the increment, coarse material may not be included at all or be underrepresented.

4.4.5 Preparation Uncertainty

The primary sample may be submitted to several stages of subsampling (secondary, tertiary sampling etc.). Between each stage of the selecting process there are non-selective steps such as handling, comminution, screening, mixing, drying, packing, transport etc. We will refer to all errors that can be introduced in these steps as Preparation Uncertainty. There are different types of errors that may be introduced (Pitard, 1993):

- *Contamination errors*, due to improper procedures, contaminated equipment, dust problems etc.
- *Loss errors*, e.g. loss of fragments (mainly fines) or volatilisation of VOCs due to improper storage and handling procedures.
- *Alteration of chemical composition* is related to loss errors and includes for example biological decomposition of chemical substances in the sample.
- *Unintentional mistakes* include human error such as dropping samples, mixing labels, mixing fractions from different samples etc.
- *Sabotage and fraud*. This type of error may be rare but the possibility for it still exists.

A statistical analysis of the Preparation Uncertainty is usually not possible. Often, it is assumed that the errors are so small, or the probability or their occurrence so low, that they can be neglected. However, this may not always be the case, especially not for VOCs where it is known that same part of the contaminant is lost during the preparation stages.

It can be assumed that the Preparation Uncertainty has a nonzero mean for most contaminants. For VOCs the error mean will most likely be negative, but for several other compounds, such as metals, a positive mean can be expected due to contamination errors.

4.4.6 Analytical Uncertainty

In this thesis we will refer to the Analytical Uncertainty as the uncertainty associated with the measurements performed at the laboratory, not including errors in subsampling in the laboratory. This distinction is made because analytical uncertainty is not part of the sampling uncertainty. This is also consistent with a definition of analytical error given in a proposed ISO-standard for sampling of stockpiles. Methods to estimate the Analytical Uncertainty in detail is beyond the scope of this thesis but can be found in the extensive chemical analysis literature, e.g. Ellison et al. (2000). Some advice and "rules of thumb" are given below.

The size of the Analytical Uncertainty is available from the laboratory through its analytical quality control procedures. However, it is important to check which types of uncertainty are included in the specified analytical uncertainty. It is not self-evident that the uncertainty as specified on the laboratory protocols should be used. For example, it is important to check whether uncertainty associated with subsampling in the laboratory is included or not. If it is included, it must be remembered that this sampling uncertainty depends on the grain-size distribution and cannot be generalised to a fixed value for all types of samples (see section 4.3.3).

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):

$$s_a^2 = \sum_{i=1}^{n} s_i^2 \qquad (4.24)$$

where n is the number of procedures in the analysis method. Generally, the analytical procedures are 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):

$$CV(\%) = 2^{(1-0.5\log C)} \approx 2 \cdot C^{(-0.1505)} \qquad (4.25)$$

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 problematic if it is based on data obtained at the limits of capability of the methodology.

4.4.7 Overall Sample Uncertainty

Posterior estimation (i.e. after the sampling has been carried out) of the random part of the Overall Sample Uncertainty (OU) is simple if several samples has been taken from the area a sample is supposed to represent. Then, the random part of OU is equal to the total variability in sample concentrations reported by the analytic laboratory.

For prior estimation, the Overall Sample Uncertainty is equal to the sum of all estimated uncertainties described in previous sections. All types of sampling and analytical uncertainties in all stages of sampling and analysis most be considered. Each of the uncertainties described above has two components, one random and one systematic component. When estimating OU each component has to be summed up separately. The random uncertainty is estimated by summing up the variances, whereas the systematic error is the sum of the expected mean errors. The assumption is that all uncertainties are independent. In this case, the random part of the Overall Sample Uncertainty can be estimated by:

$$CV_{OU}^2 = CV_{AU}^2 + \sum_{n=1}^{N} CV_{TU,n}^2 \quad (4.26)$$

where

$$CV_{TU}^2 = CV_{FV}^2 + CV_{GV}^2 + CV_{SV2}^2 + CV_{SV3}^2 + CV_{DU}^2 + CV_{EU}^2 + CV_{PU}^2 \quad (4.27)$$

CV represents the coefficient of variation (relative standard deviation) for each described error type, *n* is the number of the sampling stage, and *N* is the number of subsequent sampling stages. The systematic part of the Overall Sample Uncertainty is estimated by:

$$m_{OU} = m_{AU} + \sum_{n=1}^{N} m_{TU,n} \quad (4.28)$$

where

$$m_{TU} = m_{FV} + m_{GV} + m_{SV2} + m_{SV3} + m_{DU} + m_{EU} + m_{PU} \quad (4.29)$$

and where *m* represents the expected mean deviation from the true value for each type of uncertainty.

There are some important aspects of the expressions above: (1) there are many different types of uncertainty to consider, (2) errors can enter all stages of sampling, preparation and analysis, and (3) the uncertainties are additive. Because sampling is performed in stages, most of the uncertainty can enter the problem many times and the stage with the largest uncertainty will form the lower bound on the Overall Sample Uncertainty (Borgman et al., 1996b).

4.5 Application

4.5.1 The problem

The methodology of estimating sample uncertainty can be applied to at least two types of problems:

1. Prior estimation of sample uncertainty, i.e. before sampling has been performed.

2. Posterior identification and quantification of reducible sources of uncertainty (error types), i.e. after the sampling exercise.

An application of the methodology will be described for the problem of prior estimation of sample uncertainty. The application is a real case but some aspects of the problem have been slightly modified in order to keep the example short and not too complicated. The application of the methodology illustrates how it is possible to analyse a sampling problem by breaking it down to a number of small and well-defined problems that can be studied separately.

A site-investigation of contaminated soil is planned at a former Ferro-alloy work in the municipality of Gullspång, Sweden. The contaminant of concern is chromium. One part of the industrial site will be characterised by 12 randomly located soil samples in order to estimate the mean concentration (each sample consists of a single increment). The field sampling is performed by auger drilling, with a drill auger of 1.0 m in length. When the auger is lifted up, a sample is collected from the auger flights by hand tool. The sample is collected over the depth 0 – 1.0 m. This sample is sent to a laboratory for chemical analysis.

From this short description of the problem, three different sampling stages can be identified:

1. Field sampling by drill auger (Figure 4.12).
2. Sampling from the auger (Figure 4.13).
3. Subsampling at the laboratory (Figure 4.14).

All these sampling stages must be analysed separately because they constitute different sampling problems. This chain of sampling problems is illustrated in Figure 4.11. In the following three sections each sampling stage will be analysed in detail. Many of the uncertainties can only be estimated based on professional judgment because they have not been studied much for contaminated land problems and not much information is available. All estimations made are summarised in Table 4.2.

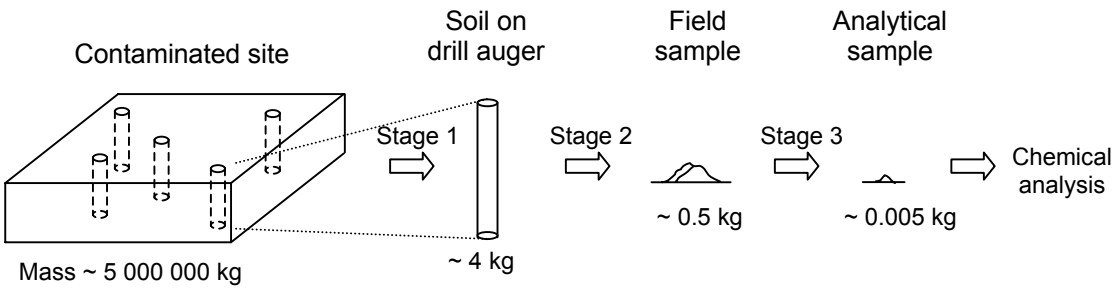


Figure 4.11 Illustration of the three sampling stages involved in drill auger sampling of contaminated soil at a former Ferro-alloy work in Gullspång, Sweden.

4.5.2 Stage 1: Field sampling by drill auger

Field sampling by drill auger is illustrated in Figure 4.12. The purpose of this sampling stage is to characterise the defined lot, i.e. the part of the contaminated site under study (target population).

A compilation of all estimated uncertainties is presented in Table 4.2. The random parts are estimated quantitatively as coefficients of variation but it is believed that it is too difficult to make reasonable quantitative estimates of the systematic parts. Therefore, the systematic parts are described qualitatively.

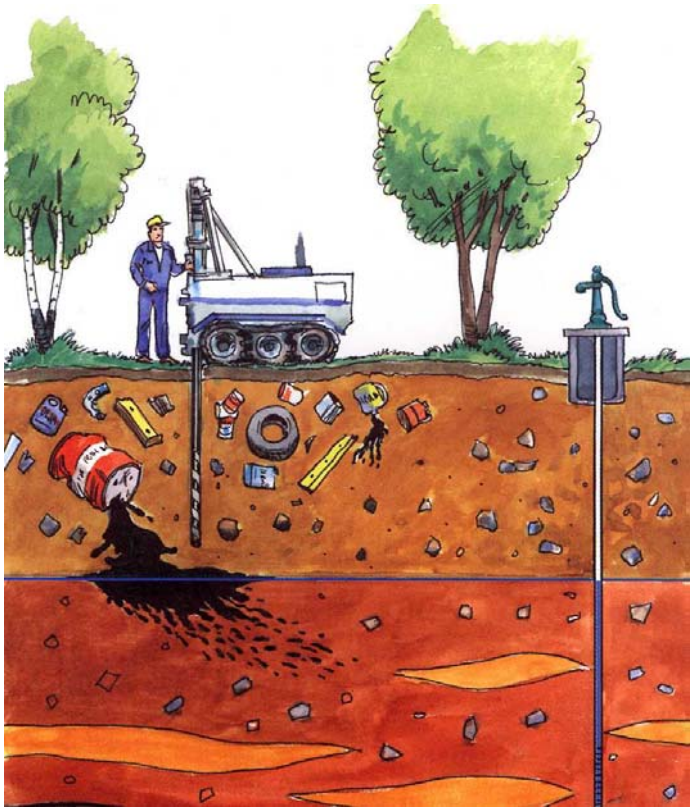


Figure 4.12 Field sampling by drill auger at a contaminated site.

Selection Variability, SV

For this particular sampling problem, each randomly located sample will only represent the sample point, not an area around the sample point. The spatial variability is handled by calculation of the mean concentration from 12 samples. Since geostatistics is not used, this means that other aspects of the spatial variability are ignored. The consequence of this is that short-range and long-range variability can be ignored in the first sampling stage, i.e. the Selection Variability is zero.

We do not expect any temporal changes in contaminant concentration to influence our sampling result and therefore the Temporal Variability is set to zero.

Materialisation Uncertainty, MU

For a two-dimensional sampling problem, the increment taken from the lot should have the shape of a perfect cylinder in order to eliminate the Increment Delimitation Uncertainty. The auger used for drilling is not expected to have a perfect cylindrical shape, rather a slightly conical shape. This will result in more soil being collected from the upper parts of the soil layer. Because slightly higher concentrations are expected in the upper part of the soil layer, the sample will be biased. This means that the concentration we expect in the sample will be higher than if a perfect cylinder had been used to delimit the sample, i.e. a positive systematic error is expected.

In addition, a positive systematic error is expected when the sample is taken out. There are at least three reasons why we can expect a positive systematic error. First, the rotation and vibration of the auger will probably have a segregating effect on the soil, leading to fewer large particles in the sample than in the undisturbed soil. Because the contaminant is expected to adhere more to the fines, this will lead to a higher sample concentration. Secondly, we can expect the upper part of the soil to contaminate the lower part of the auger sample, either when the auger is pushed down or pulled up. A higher concentration in the upper part of the soil will lead to a positive systematic error. Thirdly, a hand tool is used to remove the outermost soil layer from the auger. The purpose is to remove soil that is attached to the auger from upper soil layers. However, this procedure may result in large particles being underrepresented.

There will also be a random component of the Increment Extraction Uncertainty due to imperfections in the sampling device, vibrations, randomness in the exact drilling depth, randomness in the removal of the outer soil layer at the auger etc. (see section 4.4.4). There are no known studies available on to what extent these factors may introduce random errors. Therefore, an uncertainty estimate must be based on subjective information, such as practical experience from field work with drill augers, combined with reasoning about the different sources of materialisation uncertainty in section 4.4.4. With this approach it is believed that a subjective estimate of 10 % is realistic for this uncertainty.

Preparation Uncertainty, PU

There will be no preparation of the sample in sampling stage 1.

4.5.3 Stage 2: Sampling from the auger

Sampling from the auger is illustrated in Figure 4.13. It is a sampling stage that often is overlooked, although it may be important. For this sampling stage, the lot is equal to the soil collected from the ground by the auger. The purpose of the sampling is to produce a sample that represents the soil at the auger flights. Usually, only a portion of the soil at the auger is collected in order to keep the sample volume down. The sampling is performed by taking portions of soil (increments) from the auger by hand or hand tool, and place it in the sample container. Sampling from a drill auger is an example of a 3-dimensional sampling problem (Figure 4.1), and it is therefore very difficult to produce correct and unbiased samples. All estimations of uncertainty are presented in Table 4.2.

Selection Variability, SV

The Selection Variability is the sum of the Short-range Variability, the Long-range variability, and the Temporal Variability. The Short-range Variability is equivalent to the variability at the sample support scale due to differing contaminant concentration between individual soil particles and between groups of particles. This variability is the sum of the Fundamental Variability and the Grouping and Segregation Variability.

The Fundamental Variability is estimated according to equation 4.14, 4.16 and 4.17 (they all give similar estimates). The sample mass is approximately 0.5 kg, the mass of the lot about 4 kg, and the particle size d_{95} is estimated to 2 mm. This gives a Fundamental Variability of about 0.02 (2 %).

The Grouping and Segregation Variability is difficult to estimate due to two reasons:

1. It is not known to what degree contaminated particles may have clustered in the field.
2. It is not known to what degree the auger may have caused homogenisation or segregation of particles.

We can expect the Grouping and Segregation Variability to be larger than the Fundamental Variability because the soil at the auger is not properly homogenised. Actually, a sample made from only one increment, collected from a segregated material, may be affected by a small Fundamental Variability but an overwhelming Grouping and Segregation variability (Pitard, 1993). It is possible to reduce the variability by homogenisation prior to sampling. Another way of reducing the uncertainty is by taking a large number of randomly located small increments from the auger but this is rarely done in practice. Based on this reasoning we expect the uncertainty to be about 10 %. As far as known, there are no studies available to back it up.

No Long-range Variability exists for this sampling stage. Similarly, no Temporal Variability is expected.

Figure 4.13. Soil sampling from a drill auger.

Materialisation Uncertainty, MU

Because of the shape of the auger there is no easy way to delimit the sample correctly in order to eliminating the Increment Delimitation Uncertainty. The best way would probably be to remove all soil from the auger and take out the sample by using a splitting device. In reality, the sample is often collected from the auger as a set of increments collected from the auger flights (Figure 4.13) in a quite subjective way. The volume of the increments taken from different parts of the auger will most likely differ, and each

increment will have a different concentration because of variability in the vertical direction, leading to a random error. This error can be substantial, depending on how much care is taken by the sampling team.

As far as known, no investigations have been made on how large this uncertainty generally is. However, studies have been made on vertical variability in soil, e.g. Schumacher and Minnich (2000) who found that variability in the vertical direction can be significant in soil. Due to lack of information, an uncertainty estimate must be based on reasoning and experience from practical field work with drill augers. Such experience indicates that increment delimitation is highly subjective and varies substantial. Using this subjective approach, we will assume a coefficient of variation of 0.20 for this error, i.e. 20 % variation in soil concentration depending on how the sampling from the auger is performed.

In addition, a systematic error is expected if the increments are taken from the outer parts of the auger, because this soil may be slightly segregated. If the coarser particles are more concentrated to the outer parts there will be a negative systematic error (lower contaminant concentration in coarse particles than in the fines). However, we have earlier stated that the outer part of the soil is removed by a hand tool and therefore we will disregard this systematic component.

We do not expect the extraction itself to produce errors, only the delimitation of the increments. Therefore, the Increment Extraction Uncertainty is estimated to be zero.

Preparation Uncertainty, PU

The preparation of the sample consists of packing, storage, and transportation to the laboratory. We do not expect errors to be introduced in this chain.

4.5.4 Stage 3: Subsampling at the laboratory

Subsampling at the laboratory is illustrated in Figure 4.14. For this sampling stage the lot is equal to the sample sent to the laboratory. The purpose of the sampling is to produce an analytical subsample that represents the lot. All estimations of uncertainty are presented in Table 4.2.

Selection Variability, SV

The Fundamental Variability is estimated according to equation 4.14, 4.16 or 4.17 (they all give quite similar estimates). The sample mass is 0.005 kg, the mass of the lot is 0.5 kg and the particle size d_{95} is estimated to 2 mm. This gives a Fundamental Variability of about 20 %.

The Grouping and Segregation Variability depends on how well the original sample is homogenised and how the subsample is taken out. We will assume that this is performed in a professional way so that we can use a “rule of thumb” for this error. Pitard (1993) states that we do not make an unreasonable assumption if we assume this error to be in the same order as the Fundamental Variability. However, the laboratory is expected to do a professional job and we therefore estimate GV to be around 10 %, i.e. smaller than the Fundamental Variability.

No Long-range Variability exists for this sampling stage and no Temporal Variability is expected.

Figure 4.14. Preparing for subsampling at the laboratory.

Materialisation Uncertainty, MU

We will assume that all increments, or a single subsample, are delimited and taken out correctly at the laboratory, i.e. we will neglect Materialisation Uncertainty.

Preparation Uncertainty, PU

The preparation of the subsample consists of unpacking the sample delivered to the laboratory, storage, and homogenisation etc. at the laboratory. We do not expect significant preparation errors to be introduced during this procedure.

4.5.5 Analytical Uncertainty

We will assume the Analytical Uncertainty at the laboratory to be $CV_{AU} = 10$ %. This also agrees quite well with the Horwitz equation, discussed in section 4.4.6, for contaminant concentrations in the order of 100 mg/kg.

4.5.6 Overall Sample Uncertainty

The random part of the Overall Sample Uncertainty is estimated according to equations 4.26 and 4.27 in section 4.4.7. We estimate the uncertainty for two different values of

the Long-range Variability in sampling stage 1, as described in section 4.5.2. From Table 4.2 the Total Sampling Uncertainty can be estimated as:

$$\sum_{n=1}^{3} CV_{TU,n}^2 = CV_{EU,1}^2 + CV_{FV,2}^2 + CV_{GV,2}^2 + CV_{DU,2}^2 + CV_{FV,3}^2 + CV_{GV,3}^2 =$$
$$= 0.1^2 + 0.02^2 + 0.1^2 + 0.2^2 + 0.2^2 + 0.1^2 = 0.33^2$$

where *n* is the number of the sampling stage. The Overall Sample Uncertainty is the sum of the Total Sampling Uncertainty and the Analytical Uncertainty:

$$CV_{OU} = \sqrt{CV_{AU}^2 + \sum_{n=1}^{3} CV_{TU,n}^2} = \sqrt{0.1^2 + \left(0.33^2\right)} = 0.35$$

All results are presented in Table 4.2.

Table 4.2 *Estimated á priori uncertainties for three sampling stages at a contaminated site in Gullspång, Sweden. Random parts are expressed quantitatively by coefficients of variation (CVs). Systematic parts (error mean, m) are expressed qualitatively according to the following scale:*

- - = strong negative systematic error
- = negative systematic error
0 = no or negligible systematic error
+ = positive systematic error
++ = strong positive systematic error

	Sampling stage 1		**Sampling stage 2**		**Sampling stage 3**	
	Random, *CV*	System-atic, *m*	Random, *CV*	System-atic, *m*	Random, *CV*	System-atic, *m*
Selection variability, SV						
Fundamental Variability, FV	0	0	0.02	0	**0.2**	0
Grouping & Segr. Variab., GV	0	0	**0.1**	0	**0.1**	0
Long-range Variability, SV_2	0	0	0	0	0	0
Temporal Variability, SV_3	0	0	0	0	0	0
Materialisation Uncert., MU						
Increment Delim. Uncert., DU	0	+	**0.2**	0	0	0
Increment Extr. Uncert., EU	**0.1**	+	0	0	0	0
Preparation Uncertainty, PU	0	0	0	0	0	0
Σ Total sampling Uncert., TU	*0.10*	++	*0.22*	*0*	*0.22*	*0*

	Random, *CV*	Systematic, *m*
Total Sampling Uncertainty for all sampl. stages, Σ TU	0.33	+ +
Analytical Uncertainty, AU	**0.1**	0
Overall Sample Uncertainty, OU	**0.35**	+ +

From Table 4.2 can be concluded that the uncertainty in the proposed sampling exercise is in the order of 30-40 % (coefficient of variation). In chapter 5, a coefficient of variation of 0.4 is assumed for the individual sample data based on the example presented above. A positive bias is also expected, i.e. the measured concentration will probably be higher than the real concentration for the defined sampling problem. How large the bias will be is difficult to estimate at present state of knowledge.

The random part of the uncertainty is estimated by summing squares of the relative standard deviations. Therefore, only the largest uncertainties will actually influence the result. Addition of squared CVs implies that only errors larger than about one third of the largest of all errors will affect the Overall Sample Uncertainty, since other errors will be too small to have any influence (Morgan and Henrion, 1990). In the described application it is therefore only necessary to consider the random errors in bold in Table 4.2. Other uncertainties will be too small to affect the Overall Sample Uncertainty. As Table 4.2 illustrates, the influence from the Analytical Uncertainty is very small. Reduction of the Analytical Uncertainty will have no significant impact on the Overall Sample Uncertainty.

5 INCLUDING UNCERTAINTY IN DATA WORTH ANALYSIS

5.1 Introduction

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 ecosystems makes it necessary to investigate the degree of contamination at a site. The results from these site investigations typically contain a large amount of uncertainty due to lack of information. The heterogeneity of contaminant distribution over the site is often large and in many countries relatively few samples are collected to characterise the geochemical situation at the site. In addition to the spatial variability in contaminant concentration, uncertainty is also introduced during sampling and analysis procedures as well as during interpretation of the data. Thus, decisions that are made are based on relatively uncertain information.

Because of stiff competition on the market the consultant with the cheapest site-investigation often gets the job. A consequence is that it will be difficult to distinguish 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 problems or environmental effects. The opposite situation may also exist, i.e. huge amounts of data are collected at high cost based on demands from environmental authorities (James and Freeze, 1993; LeGrand and Rosén, 2000). Both these situations are results of bad data collection strategies from a cost perspective.

Three strategies have traditionally been used to determine the size of the sampling effort; (1) to minimise the sampling cost for a specified level of accuracy (usually variance), (2) to minimise uncertainty for a given sampling budget, or (3) to respond to demands on sampling made by the legal authorities. The concept of data worth analysis, or value-of-information analysis, constitutes a fourth alternative approach that can be used in a risk-based decision analysis framework. In this approach the optimal level of uncertainty in a site-investigation is reached when the cost for additional information (e.g. from additional sampling) is equal to the expected benefits associated with the new information. If the cost for additional information exceeds the benefit, sampling is no longer cost-efficient.

The purpose of this chapter is to include sample uncertainty in the data worth analysis within a RCB decision analysis framework, and at a complexity level that can be applied to practical contaminated land problems with limited amount of data. The uncertainty in individual samples is estimated with the approach described in chapter 4. The methodology is applied to a sampling problem at a former Ferro-alloy work in Gullspång, where metal contamination is expected (see also section 4.5.1).

5.2 Previous work

During the 1970s a number of applications of data worth analysis were reported for various water-related problems (Davis and Dvoranchik, 1971; Gates and Kisiel, 1974; Maddock, 1973) and additional work for hydrogeological problems was reported during the 1980s (Ben-Zvi et al., 1988; Grosser and Goodman, 1985; Reichard and Evans, 1989). In the last decade, data worth analysis has been used for a number of groundwa-

ter contamination problems (Abbaspour et al., 1996; Freeze et al., 1992; James and Freeze, 1993; James and Gorelick, 1994; James et al., 1996a; Russell and Rabideau, 2000).

For problems of contaminated particulate materials such as soil or waste material the spatial variability is usually larger than for groundwater contamination problems, especially at small scales. Applications of data worth analysis to such problems are more limited. Dakins et al. (1994; 1995) presented a decision framework for remediation of PCB-contaminated sediments. Rautman et al. (1994) used a risk-based decision analysis approach to compare the reliability of different characterisation techniques for uranium contaminated soil. James et al. (1996b) presented a simple risk-based decision analysis framework for remediation of contaminated soil, whereas Kaplan (1998) described software with the purpose of locating sample points based on their data worth. A detailed review of previous work on data worth analysis is presented in section 3.5.

The complexity of several of the previously presented approaches has restricted the application of data worth analysis to real-world problems. Although complex, none of the frameworks above explicitly accounts for all sampling and analytical uncertainties, although these may have a significant impact on the remedial decision. However, Freeze et al. (1992) discussed how sampling uncertainty could easily be included in their framework using stochastic process theory.

5.3 An approach for including uncertainty in data worth analysis

5.3.1 Underlying assumptions

Random sampling

The purpose of this chapter is to estimate the worth of data from a proposed a sampling program. The objective of the sampling is to estimate the true mean concentration at the site or part of the site of interest. The spatial distribution of the contaminant is not known and we do not know if the concentrations at the sample scale follow a particular statistical distribution. All samples will be located randomly over the entire area, in order to derive at an unbiased estimate of the mean concentration. It is important that all parts of the area have an equal probability of being selected as a sample point, otherwise our estimate will be biased.

Uncorrelated samples

It is assumed that all samples are uncorrelated. For contaminated soil, correlation distance is often short so the assumption is that the distances between sample points are longer than the correlation distance, leading to uncorrelated samples. This assumption may not hold completely when many samples are collected from a relatively small area. More common though, is that relatively few samples are taken, which makes the assumption reasonable.

Sample uncertainty

The random part of the sample uncertainty is quantified as a coefficient of variation, CV_{OU} (relative standard deviation of the Overall Sample Uncertainty). The systematic uncertainty is denoted m_{OU}, all according to chapter 4.

Approach to data worth

The methodology for data worth analysis will be described rather thorough from a geological point of view, and with practical application in mind. A solution to our problem, with the emphasis on the statistical aspects, has been presented in shorter mathematical notation by Lindley (1997).

A relevant question is why we have chosen the above approach instead of using traditional geostatistics. There are two reasons for this. The first one is that the approach taken in this thesis is to apply a methodology that is capable of handle the great number of contaminated land problems for which geostatistics is no good option. The limitations of the geostatistical approach have restricted its use to large projects with many samples. At least 25-30 sample points is required to produce a reliable variogram model, and even with this number of samples the experimental variogram is often difficult to interpret. Many real-world contaminated land problems encompass only a handful of samples. Since the correlation length in soil is quite short, collection of only a few samples will lead to large distances between sample points and consequently, uncorrelated samples. A more fundamental problem with geostatistics is the assumption of stationarity in the kriging approach (the intrinsic hypothesis, see section 2.7.3). This assumption is often not valid for sites contaminated by a point source. Such situations are quite common for contaminated land problems, and therefore kriging can be questioned in many cases.

The second reason is the wish to describe data worth analysis in a way that is understandable also for persons with little knowledge in geostatistics. Data worth analysis is not easy to understand in all its details and including geostatistics would make the presentation even more complicated.

5.3.2 Procedure

The principles of data worth analysis have been described briefly in section 3.4.1 and Figure 3.3. In the following sections, these principles are applied on our problem defined above. Figure 5.1 presents the procedure of carrying out the data worth analysis for the defined problem.

The procedure consists of five steps. The first step is to define the sampling program or the set of sampling programs that are evaluated. Secondly, the prior information about the mean concentration at the site must be transformed to a probability density function (PDF). Thirdly, four different types of probabilities are estimated, based on the proposed sampling program and the prior information. Fourthly, all involved benefits and costs must be estimated. The last step is then to estimate the data worth based on probability estimates and cost estimates. The presentation in the following sections will roughly follow these steps in Figure 5.1. Finally, the methodology is applied to a real case and conclusions are drawn.

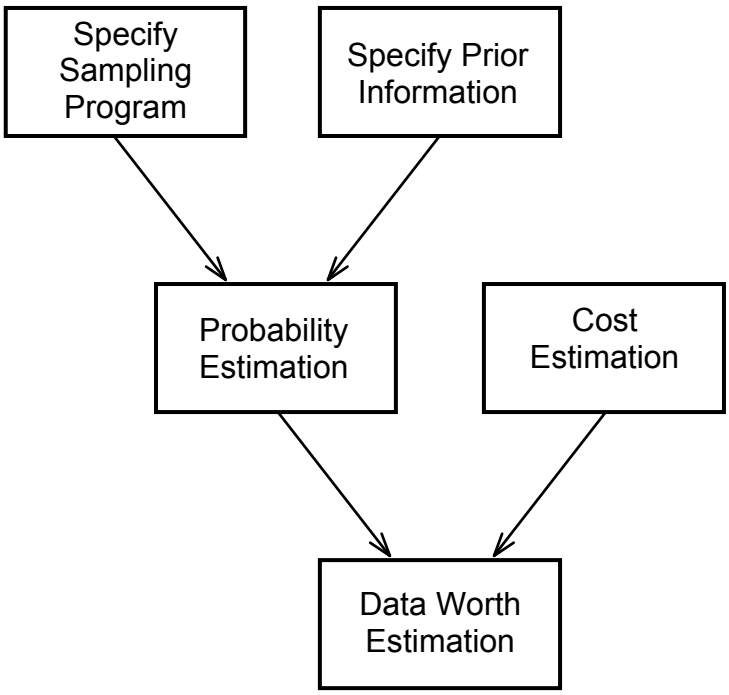


Figure 5.1 Procedure for estimation of data worth of a proposed sampling program.

5.3.3 Step 1: Sampling program

The sampling objective is to determine the mean concentration at a site or part of a site. For this problem, two parameters need to be specified:

- the number of samples (n) in the sampling program, and
- the uncertainty associated with an individual sample.

The uncertainty in sample data is equivalent with the Overall Sample Uncertainty, defined in chapter 4. It is expressed as a coefficient of variation (relative standard deviation), and following the notation in chapter 4 it is denoted CV_{OU}.

In addition, the sampling program must specify random sampling, i.e. all parts of the site must have the same probability of being sampled.

5.3.4 Step 2: Prior information

Prior information is information that is available before the sampling has been carried out. For our problem, prior information about the mean concentration μ at the site can be expressed by means of PDFs. It is important to distinguish the PDF of the expected mean concentration from the PDF of the expected sample concentrations. The PDF of sample concentrations is of no concern in our approach, only the PDF of the mean concentration. This PDF should reflect our belief of the likelihood of different mean concentrations at the site.

In a Bayesian approach like ours, prior estimation of the mean concentration can be based on soft information if no hard data is available. There are several possible sources of soft information, e.g. experience from similar sites, professional judgement, literature

etc. The prior PDF can be specified in a number of ways. In our approach, the following must be specified as prior information:

1. Type of PDF
2. Reasonable minimum value, *a*, of the mean concentration
3. Most likely value (mode), *m*, of the mean concentration
4. Reasonable maximum value, *b*, of the mean concentration

Another possibility is to specify the prior probabilities P(*C*+) and P(*C*-) and base the shape of the selected PDF on these probabilities (see section 5.3.5). However, this approach has not been used in this thesis.

From the vast number of available PDFs we will consider only four:

- uniform (rectangular)
- triangular
- normalised normal (normalised gaussian)
- log-normal

These PDFs are illustrated in Figure 5.2 and are described below. Each PDF is defined solely by the parameters *a*, *m*, and *b*, which makes it easy to define the PDFs for practical problems. The probability P(μ<*b*) is denoted P_s and is illustrated in Figure 5.2, and described for each PDF.

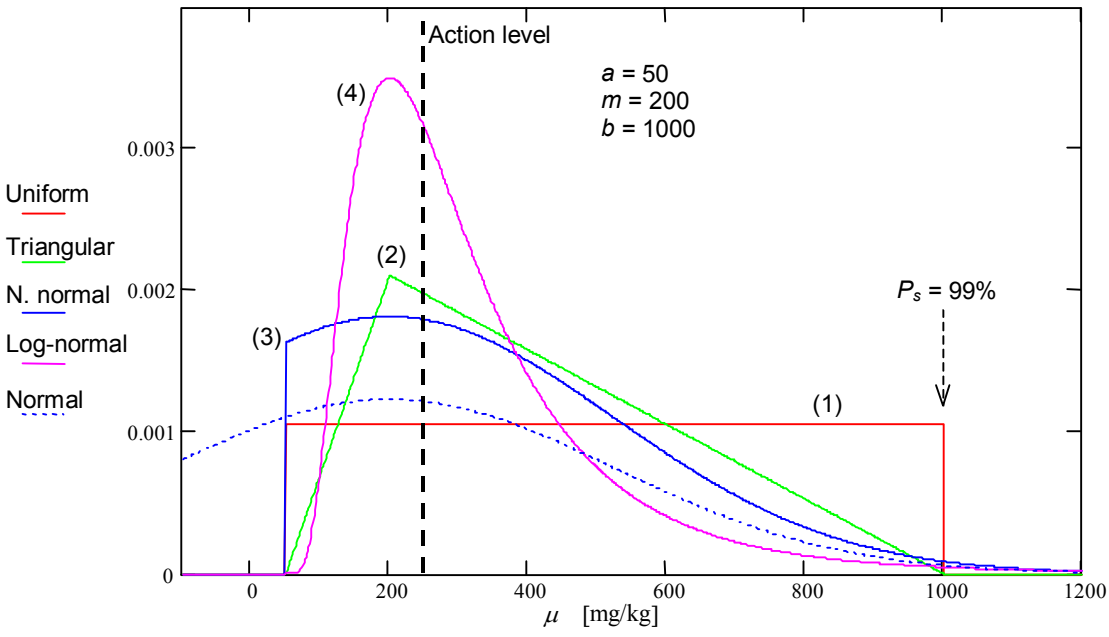


Figure 5.2 *Four different probability density functions defined by the minimum value a=50 mg/kg, the most likely value m=200 mg/kg, and the maximum value b=1000 mg/kg: (1) uniform distribution, (2) triangular distribution, (3) normalised normal distribution, and (4) log-normal distribution. A normal distribution is illustrated with a dotted line. The action level is set to 250 mg/kg.*

The uniform PDF

The *uniform* distribution is defined as:

$$f_{uni}(\mu) = \begin{cases} \dfrac{1}{b-a} \\ 0 \;\; if \;\; a < \mu < b \end{cases} \tag{5.1}$$

where μ is the independent variable, in our case the mean concentration we want to express by a PDF. The probability P_s is equal to 1 for the uniform PDF.

The triangular PDF

The *triangular* distribution is defined as:

$$f_{tri}(\mu) = \begin{cases} \dfrac{2(\mu - a)}{(b-a)\cdot(m-a)} \;\; if \;\; a < \mu < m \\ \dfrac{2(b-\mu)}{(b-a)\cdot(b-m)} \;\; if \;\; m < \mu < b \\ 0 \;\; otherwise \end{cases} \tag{5.2}$$

The probability P_s is equal to 1 for the triangular PDF.

The normalised normal PDF

The *normal* distribution has the drawback of providing values below the practical minimum value a, including values below zero (Figure 5.2). This can be avoided if the distribution is normalised in the following way:

$$f_{nn}(\mu) = \begin{cases} \dfrac{f_n(\mu)}{1 - P_n(\mu < a)} \\ 0 \;\; if \;\; \mu < a \end{cases} \tag{5.3}$$

where $f_{nn}(\mu)$ is the normalised normal distribution, P_n represents a probability based on the normal distribution below, and $f_n(\mu)$ is the normal distribution:

$$f_n(\mu) = \frac{1}{\sqrt{2\pi}\cdot\sigma_n} \cdot EXP\left[-\frac{(\mu - m_n)^2}{2\sigma_n^2}\right] \tag{5.4}$$

Equation 5.3 implies that the whole probability mass (=1) of the normalised normal distribution lies above the minimum value a. Figure 5.2 illustrates how the normalisation "lifts" the distribution to compensate for values of $\mu < a$.

In equation 5.4, σ_n is the standard deviation of the normal distribution and m_n is the mean of the distribution. The task is to define the distribution from parameters *a*, *b*, and *m*. For a normal distribution, m_n is equal to *m*. We define σ_n so that *b* is the upper limit of a certain percentile of the distribution in equation 5.3, typically the 95 % or 99 % percentile, i.e. 95 % or 99 % of the values of the distribution fall below *b*. This can be expressed as:

$$P_s = P_{nn}(\mu < b) \tag{5.5}$$

where P_s is the specified percentile at *b* [decimal fraction], and P_{nn} represents a probability based on the normalised normal distribution in equation 5.3. Equation 5.5 can be rewritten in the following way:

$$P_s = \frac{P_n(a < \mu < b)}{1 - P_n(\mu < a)} \tag{5.6}$$

where P_n represents a probability based on the normal distribution in equation 5.4. The proper value of σ_n is found by solving equation 5.6. This can be achieved with the help of equation solver software.

The log-normal PDF

The log-normal distribution for our case can be expressed as:

$$f_{\log}(\mu) = \frac{1}{(\mu - a)\cdot\sqrt{2\pi}\cdot\sigma_l}\cdot EXP\left[-\frac{(\ln(\mu - a) - m_l)^2}{2\sigma_l^2}\right] \tag{5.7}$$

where m_l is the log mean and σ_l is the log standard deviation. First, the proper value of σ_l is determined. We select σ_l so that *b* is the upper limit of a certain percentile of the distribution in equation 5.7, typically the 95 % or 99 % percentile. The proper value of σ_l is found when:

$$P_s = P_l(\mu < b) \tag{5.8}$$

where P_s is the specified percentile [decimal fraction], and P_l represents a probability based on the log-normal distribution in equation 5.7. Equation 5.8 can be solved for σ_l with the help of equation solver software.

Secondly, we want to specify m_l. However, based on our prior information it is much easier to specify the mode *m* of the distribution than the log mean. The relationship between the log mean and the mode is:

$$m_l = \ln\left[(m - a)\cdot e^{\sigma_l^2}\right] \tag{5.9}$$

Thus, the log-normal distribution is defined solely by parameters *a*, *m*, and *b*.

To apply the presented methodology of defining prior PDFs, the following steps should be taken: (1) estimate the lowest, most likely and the highest reasonable values of the mean concentration, (2) study the behaviour of different PDFs defined by the three parameters *a*, *b*, and *m*, and (3) select the PDF that best captures the prior knowledge.

5.3.5 Step 3: Probability estimation

There are four types of probabilities to consider in the analysis. Freeze et al. (1992) denote them:

1. P[state], i.e. the probability of the two alternative states *contaminated* site or *not contaminated* site.
2. P[sample/state], i.e. the conditional probability of the sampling result, given the state.
3. P[sample], i.e. the probability of the two alternative sampling results; *detection* or *no detection* of contamination.
4. P[state/sample], i.e. the conditional probability of the state, given the sampling result.

The probabilities are estimated in the same order as listed above. Starting with the prior analysis, we estimate the prior probability P[state] based on prior information. Our particular interest is to update the prior probability to a preposterior probability P[state/sample]. This is achieved with Bayes´ theorem but first we need to estimate the probabilities P[sample/state] and P[sample]. The latter two types of probabilities can be estimated when the sample uncertainty has been included and weighted with the prior information. All probability estimations are described below, roughly following the procedure for data worth analysis presented by Freeze et al. (1992), but with different underlying assumptions and including sample uncertainty.

Prior probabilities

The prior probabilities of the true state, P[state] are defined as:

$P(C^+) =$ the probability of contaminated site
$P(C^-) =$ the probability of uncontaminated site

The site is regarded as contaminated if the true mean concentration exceeds an action level *AL*, such as a guideline value or a soil screening level. The prior probabilities are estimated from the prior PDF as the area above and below the action level. Here, we ignore the fact that action levels themselves can be regarded as uncertain.

Including sample uncertainty

Assumptions for the problem are described in section 5.3.1. In addition, let μ be the true but unknown mean concentration in soil, which we try to estimate from a planned sampling program of samples $i = 1..n$. Each expected data value x_i is afflicted with uncertainty due to sampling errors and analytical errors etc., as described in chapter 4. All these uncertainties are combined into a single coefficient of variation, CV_{OU}, for each sample value x_i, i.e. the random component of the Overall Sample Uncertainty. The sys-

tematic component is expressed as m_{OU} for each x_i. It is assumed that all samples are associated with the same uncertainty. The expected mean concentration x, derived from planned samples, is assumed to be normally distributed around μ. The assumption of normally distributed errors around the mean value is commonly applied and is believed to be valid also for this problem. The basis for this assumption is the central limit theorem that states that the sum of a large number of independent errors is normally distributed. The standard deviation σ_x is estimated from the uncertainty in individual sample data, CV_{OU}, and the number of samples, n:

$$\sigma_x = \frac{CV_{OU} \cdot \mu}{\sqrt{n}} \tag{5.10}$$

The expected bias is the same for all samples. In this case, the bias in mean concentration, m_x, can be formulated as:

$$m_x = m_{OU} \tag{5.11}$$

If m_{OU} is quantified it is possible to correct for the bias so that x instead is normally distributed around $\mu + m_x$. We can summarise the distribution of x in the following way:

$$x \sim N\left(\mu + m_x; \sigma_x^2\right) \tag{5.12}$$

Figure 5.3 illustrates the distribution of x for four of the infinite values of μ. The increasing width of the distribution for higher values of μ is a result of using a constant coefficient of variation (CV) to characterise the random part of the uncertainty in sample data. This is reasonable because larger absolute errors can be expected for higher concentrations than for lower.

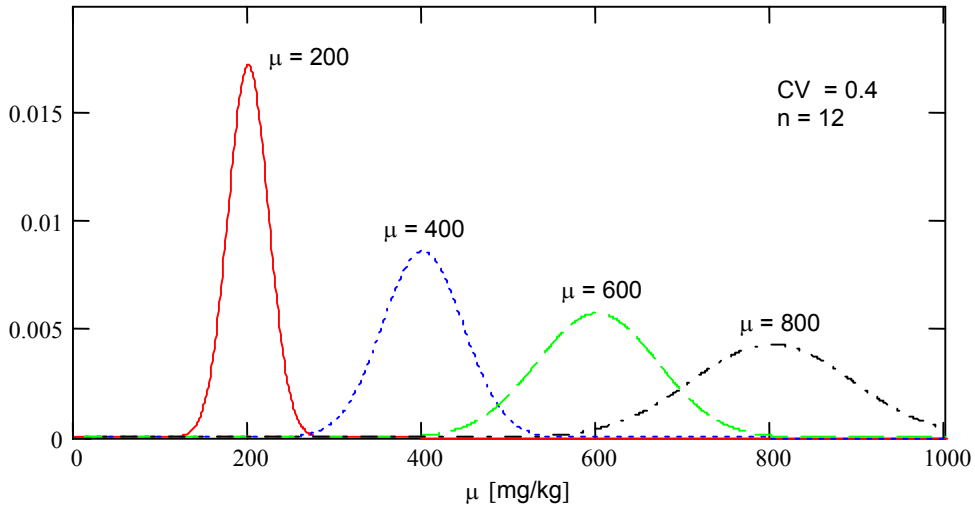


Figure 5.3 *The distribution of mean concentration x from a sampling program with uncertainty in sample data, as a function of the true mean concentration μ. The distribution of x is illustrated for four of the infinite number of values of μ, assuming random error but no systematic error.*

We can now write an expression for the probability of x exceeding an action level AL as a function of the true mean concentration:

$$p_1(\mu) = P(x > AL|\mu) = P(D^+|\mu) = P_x(x > AL) \tag{5.13}$$

where P_x is a probability based on the normal distribution of x according to equation 5.12, and D^+ symbolises that contamination is detected, i.e. the site is classified as contaminated because $x > AL$. Similarly, the opposite situation is formulated as:

$$p_2(\mu) = P(x < AL|\mu) = P(D^-|\mu) = P(x < AL) \tag{5.14}$$

The functions 5.13 – 5.14 include tail probabilities for $x<0$, and even small probabilities of x exceeding the highest concentration that possibly can be encountered. This may introduce an error because the tail of the distribution may reach significantly below zero for certain problems, which is unrealistic. Therefore, functions 5.13 – 5.14 are normalised for the low tail probability below zero. The normalised versions of the functions are:

$$p_{x>AL}(\mu) = \frac{P_x(x > AL)}{P_x(x > 0)} \tag{5.15}$$

$$p_{x<AL}(\mu) = \frac{P_x(0 < x < AL)}{P_x(x > 0)} \tag{5.16}$$

In functions 5.15 – 5.16, no normalisation is performed for the high tail exceeding the highest possible concentration, and for two reasons: (1) For most problems, μ will be much lower than the maximal concentration, i.e. the high tail probability will be insignificant. (2) If normalisation is performed for the high tail, the functions will behave strange when μ approaches the maximal concentration. The distribution would be stretched upwards more and more when μ increases, causing $P_x(x>AL)$ to actually decrease when μ increases. This phenomenon will occur because the normalisation is a function of μ (note that the low tail probability $P_x(x<0)$ actually is constant, i.e. independent of μ).

Weighing sample uncertainty with prior information

In previous sections we have defined PDFs to express prior information and formulated expressions of uncertainty in mean concentration. The next step is to combine these parts for estimation of probabilities of false and correct classification of the site as "contaminated" or "not contaminated", similar to the decision space illustrated in Figure 3.6. Freeze et al. (1992) used the notation P[sample/state] for the following four probabilities:

$P(D^+|C^+)$ = the probability of correctly classifying the site as contaminated

$P(D^-|C^+)$ = the probability of incorrectly classifying the site as uncontaminated

$P(D^+|C^-) =$ the probability of incorrectly classifying the site as contaminated

$P(D^-|C^-) =$ the probability of correctly classifying the site as uncontaminated

where C^+ and C^- represent the true state of the site, i.e. contaminated and not contaminated, respectively. These four probabilities represent the expected outcome of the proposed sampling program.

From Figure 5.3 it is clear that false classification as contaminated or uncontaminated may occur, because of the uncertainty in individual samples. However, Figure 5.3 does not give information of how likely different values of μ are. This information is supplied by the prior PDF from section 5.3.4. The different probabilities P[sample/state] are estimated by weighing all possible distributions of x around μ (Figure 5.3) with the prior PDF of the true mean concentration μ (Figure 5.2). This is performed by integrating upwards or downwards from the action level, with respect to μ:

$$P(D^+|C^+) = \int_{AL}^{\infty} p_{x>AL}(\mu) \cdot \frac{f_{prior}(\mu)}{P(C^+)}\, d\mu \tag{5.17}$$

$$P(D^-|C^+) = \int_{AL}^{\infty} p_{x<AL}(\mu) \cdot \frac{f_{prior}(\mu)}{P(C^+)}\, d\mu \tag{5.18}$$

$$P(D^+|C^-) = \int_{0}^{AL} p_{x>AL}(\mu) \cdot \frac{f_{prior}(\mu)}{P(C^-)}\, d\mu \tag{5.19}$$

$$P(D^-|C^-) = \int_{0}^{AL} p_{x<AL}(\mu) \cdot \frac{f_{prior}(\mu)}{P(C^-)}\, d\mu \tag{5.20}$$

where $f_{prior}(\mu)$ is the selected prior distribution of the mean concentration. An assumption is that the high tail of the distribution, above the highest possible concentration, is negligible.

The probabilities of detecting or not detecting contamination, P[sample], can now be estimated by simple probability theory:

$$P(D^+) = P(C^-) \cdot P(D^+|C^-) + P(C^+) \cdot P(D^+|C^+) \tag{5.21}$$

$$P(D^-) = P(C^-) \cdot P(D^-|C^-) + P(C^+) \cdot P(D^-|C^+) \tag{5.22}$$

By "detection" we actually mean that the mean of sample data (x) exceeds the action level (AL).

Preposterior probabilities

The last step in probability estimation is to update the prior probabilities to preposterior probabilities P[state/sample] with the expected outcome of the proposed sampling pro-

gram. The updating is performed with Bayes´ theorem (see section 2.7.2). This is a simple arithmetic exercise because all involved probabilities are known:

$$P\left(C^{+}\middle|D^{+}\right)=\frac{P\left(D^{+}\middle|C^{+}\right)\cdot P(C^{+})}{P\left(D^{+}\middle|C^{+}\right)\cdot P(C^{+})+P\left(D^{+}\middle|C^{-}\right)\cdot P(C^{-})} \tag{5.23}$$

$$P\left(C^{-}\middle|D^{+}\right)=\frac{P\left(D^{+}\middle|C^{-}\right)\cdot P(C^{-})}{P\left(D^{+}\middle|C^{+}\right)\cdot P(C^{+})+P\left(D^{+}\middle|C^{-}\right)\cdot P(C^{-})} \tag{5.24}$$

$$P\left(C^{+}\middle|D^{-}\right)=\frac{P\left(D^{-}\middle|C^{+}\right)\cdot P(C^{+})}{P\left(D^{-}\middle|C^{+}\right)\cdot P(C^{+})+P\left(D^{-}\middle|C^{-}\right)\cdot P(C^{-})} \tag{5.25}$$

$$P\left(C^{-}\middle|D^{-}\right)=\frac{P\left(D^{-}\middle|C^{-}\right)\cdot P(C^{-})}{P\left(D^{-}\middle|C^{+}\right)\cdot P(C^{+})+P\left(D^{-}\middle|C^{-}\right)\cdot P(C^{-})} \tag{5.26}$$

Now, all probabilities involved in the analysis have been estimated.

5.3.6 Step 4: Cost estimation

Four types of cost- or benefit-related values should be specified:

1. Benefits provided by the selected alternative action.
2. Costs induced by the selected alternative action (investment costs, operational costs etc. for mitigating or protective actions).
3. Failure costs, i.e. costs brought about by making the wrong decision.
4. Cost of the proposed sampling program.

The three first values are used in calculations of the objective function. The sampling cost is necessary to estimate the expected net value of the sampling program. The principles for cost estimation are beyond the scope of this thesis, but some of these costs are briefly discussed in Appendix 1.

5.3.7 Step 5: Data worth estimation

The estimation of data worth is performed in three steps:

1. calculation of the prior objective function (prior analysis)
2. calculation of the preposterior objective function (preposterior analysis)
3. calculation of the worth of sample data

Each of these steps is described below.

Prior analysis

The first step in the data worth estimation is to perform a prior analysis according to Figure 3.3. The decision tree for the prior analysis is presented in Figure 5.4, based on Freeze et al. (1992).

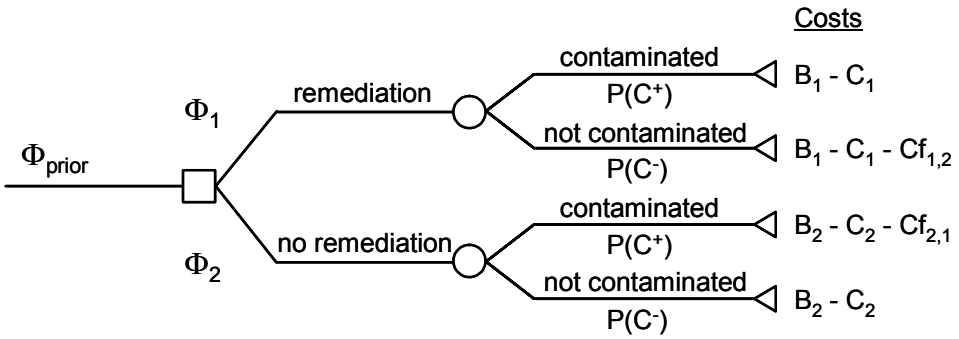


Figure 5.4 *Example of a decision tree for prior analysis of a contaminated land problem with two decision alternatives, remediation or no remediation, and two states, contaminated or not contaminated (after Freeze et al., 1992).*

The objective function in equation 3.1 is applied for the prior stage but formulated differently. If we apply it to the decision tree in Figure 5.4, and neglect time considerations and interest rates, the prior objective function for two decision alternatives can be formulated as:

$$\Phi_{prior} = \max\left[\Phi_1 \;;\Phi_2\right] \tag{5.27}$$

where Φ_i is the objective function for decision alternative i. Equation 5.27 implies that the prior objective function is equal to the expected (probabilistic) cost for the decision alternative leading to the lowest expected cost. The objective function Φ_i is estimated as:

$$\Phi_i = B_i - C_i - \sum_{j=1}^{n}\left[P_{i,j} \cdot Cf_{i,j}\right] \tag{5.28}$$

where B_i is the benefit from decision alternative i, C_i is the cost for decision alternative i, $P_{i,j}$ is the probability of state j for decision alternative i, $Cf_{i,j}$ is the failure cost at state j for decision alternative i, and n is the number of states for one decision alternative (terminal nodes in the decision tree). For the problem in Figure 5.4 we have $n = 2$. The probability $P_{i,j}$ is identical to P[state] estimated in section 5.3.5.

Preposterior analysis

The preposterior analysis is performed when the sampling program has been specified, but before the actual sampling has been carried out. A decision tree for the preposterior analysis is presented in Figure 5.5. The preposterior objective function for the problem in Figure 3.4 can be formulated as:

$$\Phi_{preposterior} = P\left(D^{+}\right) \cdot \Phi_{D+} + P\left(D^{-}\right) \cdot \Phi_{D-} \tag{5.29}$$

where Φ_{D+} and Φ_{D-} represent the objective functions for the upper and lower half of the decision tree in Figure 5.5, corresponding to "contamination detected" or "contamination not detected". These objective functions are calculated in the same manner as Φ_{prior} in equation 5.27, but taking into account the two states of detection. The state of detection will affect the probabilities but not the costs. The objective function Φ_D is estimated as:

$$\Phi_D = \max\left[\Phi_{1,d} \;; \Phi_{2,d}\right] \tag{5.30}$$

where d represents the state of detection. The objective function $\Phi_{i,d}$ is estimated as:

$$\Phi_{i,d} = B_i - C_i - \sum_{j=1}^{n}\left[P_{i,j,d} \cdot Cf_{i,j}\right] \tag{5.31}$$

This expression is identical to equation 5.28 except that during the preposterior analysis $P_{i,j,d}$ represents the probability P[state/sample] estimated in section 5.3.5.

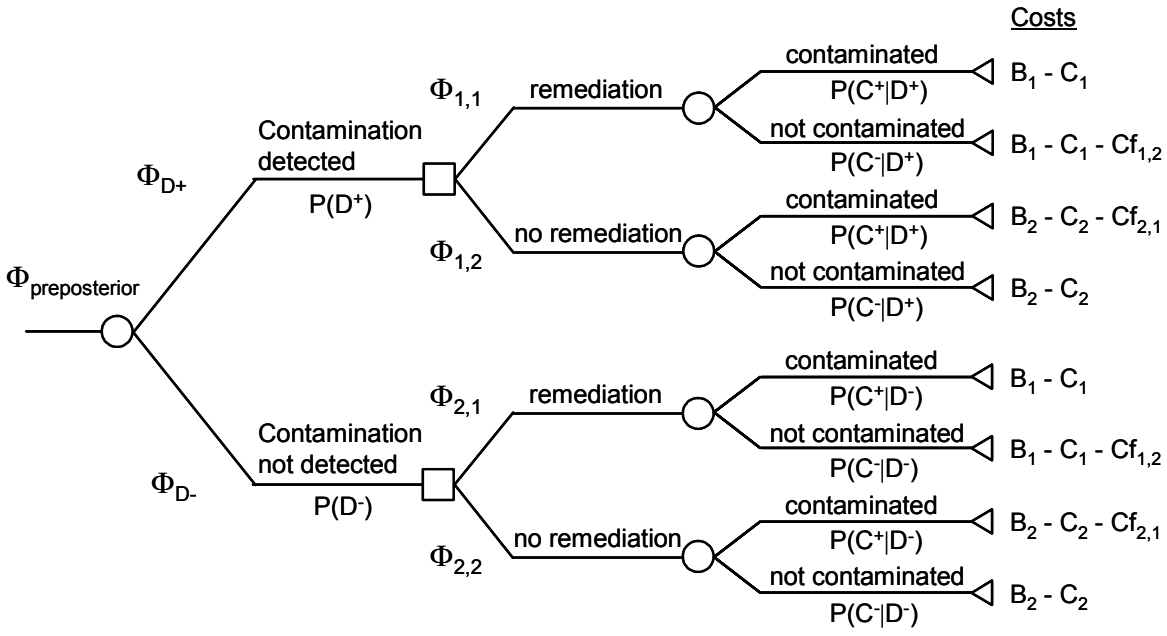


Figure 5.5 Example of a decision tree for preposterior analysis of a contaminated land problem with two decision alternatives, remediation or no remediation, and two states, contaminated or not contaminated (after Freeze et al., 1992).

Data worth

We will denote the estimated data worth as Expected Value of Sample Information (*EVSI*) according to section 3.4.2. To avoid misinterpretation it should be noted that EVSI is referred to as the value of the whole sampling program, not the value of individual samples. Freeze et al. (1992) estimate the data worth as:

$$EVSI = \Phi_{preposterior} - \Phi_{prior} \tag{5.32}$$

EVSI is estimated for a sampling program that is not perfect, i.e. there will still be some uncertainty regarding the true state when the sampling has been performed. It is possible to estimate the data worth for a perfect sampling program eliminating all uncertainty. This upper bound on the worth of a proposed sampling program is called Expected Value of Perfect Information (*EVPI*), see section 3.4.2. It is estimated from the prior decision tree in Figure 5.4 as:

$$EVPI = \Phi_{\max} - \Phi_{prior} \tag{5.33}$$

Φ_{max} can be estimated as:

$$\Phi_{\max} = P\left(C^{+}\right) \cdot \max\left[B_{i,1} - C_{i,1} - Cf_{i,1}\right] + P\left(C^{-}\right) \cdot \max\left[B_{i,2} - C_{i,2} - Cf_{i,2}\right] \tag{5.34}$$

where B_i represents benefit, C_i cost, and Cf_i failure cost, for the true states j=1 (C^+) and j=2 (C^-) for decision alternative i.

According to James et al. (1996b) and equation 3.5 we can formulate the reliability of the sampling program, i.e. the value of the sampling program in relation to perfect information:

$$reliability = \frac{EVSI}{EVPI} \tag{5.35}$$

So far, only the worth of the proposed sampling program has been addressed. To be able to study the cost-efficiency of the sampling program we must also consider the cost of the sampling program. This cost is expressed as a cost function $C_p(n)$. A simple cost function is the following:

$$C_p(n) = n \cdot C_s \tag{5.36}$$

where n is the number of samples in the sampling program, and C_s is the cost per sample. This simple function does not take into account that the cost per sample often is lower for large numbers of samples than for few. The cost-efficiency is expressed as the Expected Net Value (*ENV*) of the sampling program:

$$ENV = EVSI - C_p(n) \tag{5.37}$$

The expected net value can be used to identify cost-efficient sampling programs and to compare different programs. The higher *ENV*, the more cost-efficient is the sampling program. Only sampling programs with positive *ENV* will be cost-efficient, whereas negative *ENV* implies that the reduction in risk due to the sampling cannot balance the sampling cost.

5.4 Application

5.4.1 The problem

The described methodology for data worth analysis was applied to a real-world case. It is a continuation of the application described in section 4.5. Some simplifications have been made in order to keep the example relatively simple.

A site-investigation of contaminated soil is planned at a former Ferro-alloy work in the municipality of Gullspång, Sweden. The industry began in 1909 and the main production has been Ferro-silica, Ferro-tungsten, and Ferro-molybdenum. The contaminant of concern in the analysis is chromium. Two decision alternatives are considered; *remediation* and *no remediation* of the site. The remediation technique is assumed to be excavation of contaminated soil. Two true states of the site are evaluated; *contaminated* and *not contaminated*. The problem is structured according to the decision trees presented in Figure 5.4 and Figure 5.5.

The objective of the sampling program is to estimate the mean concentration. The estimated mean concentration will be compared to an action level, in our case a chromium concentration of 250 mg/kg, which is the generic guideline value for land with less sensitive use in Sweden (Naturvårdsverket, 1997b). Remediation is suggested if the mean concentration exceeds the action level. Our particular interest is to determine if the sampling program is cost-efficient, and if there are other sampling programs that are more cost-efficient than the proposed one.

The presentation will follow the five step procedure presented in section 5.3.2. In short, the procedure can be described as:

Step 1: Specify the sampling program. The number of samples and the uncertainty in sample data must be specified.

Step 2: Use prior information to define a prior PDF of the mean concentration at the site. Estimate minimum, most likely, and maximum value and select PDF type.

Step 3: Estimate probabilities in the following order: P[state], P[sample/state], P[sample], and P[state/sample].

Step 4: Estimate the following costs: (1) benefit and cost for each decision alternative, (2) cost of failure, and (3) sampling cost.

Step 5: Estimate the worth of the sampling program as Expected Net Value (and other types of estimates if necessary).

5.4.2 Step 1: Sampling program

The area of interest is defined as a part of the industrial site, 100×30 m^2. It will be characterised by a sampling program consisting of 12 randomly located soil samples at the level 0-1.0 m below ground. Each sample data will be associated with uncertainty, as described in chapter 4. In our case, the uncertainty is expected to be about 0.4 (coefficient of variation of the Overall Sample Uncertainty), based on the estimation of sample uncertainty in section 4.5. No systematic error is taken into account.

5.4.3 Step 2: Prior information

Our prior information about the mean chromium concentration should be expressed as a PDF. We have no information about soil concentrations at similar Ferro-alloy works. Our prior PDF is instead based on background levels of chromium in Swedish population centres (Naturvårdsverket, 1997a). Minimum (*a*), most likely (*m*), and maximum (*b*) mean concentration at the site is estimated. The estimates are based on the assumption that the chromium concentration in the soil is significantly higher at the site than in population centres. The following estimates are made:

Minimum mean concentration, *a*: 50 mg/kg
Most likely mean concentration, *m*: 200 mg/kg
Maximum mean concentration, *b*: 1000 mg/kg

As mentioned, the action level is 250 mg/kg. Calculations are performed for all four PDFs described in section 5.3.4. Several different arguments and advice can be given for the selection of PDF-type: (1) If the prior information is so weak that all values between the minimum and the maximum is believed to equally probable, the uniform PDF is the best choice. (2) On the other hand, a concentration near the minimum or the maximum is often considered less probable, and therefore a triangular distribution would better capture the prior information. (3) However, there are also arguments that the normal distribution best capture the uncertainty in mean concentration. If a large number of analysts would estimate the mean concentration, the result is expected to be normally distribution around the true mean according to the central limit theorem (systematic error ignored). However, the normal distribution must be normalised to avoid negative values, which may result in a rather strange shape (Figure 5.2). (4) This is avoided if the log-normal PDF is used. The log-normal PDF also implies that the peak of the distribution is displaced towards low values and thus assuming high values to be less probable.

The purpose of using four different PDFs is to study the effect of different prior distributions. All estimated PDFs are illustrated in Figure 5.2. The maximum value *b* is defined as the 99 %-percentile for the normalised normal distribution and the log-normal distribution. Occasionally, only the calculations based on the uniform distribution is presented. It is believed that it best represents the weak prior information about mean concentration at the industrial site in Gullspång.

5.4.4 Step 3: Probability estimation

The prior probabilities P[state] are estimated as the area above the action level for each PDF (Figure 5.2). All estimated probabilities are listed in Table 5.1. The prior probability of a contaminated site is estimated to 57 % - 79 %, depending on the choice of prior distribution.

Table 5.1 Estimated probabilities based on four different prior distributions during data worth analysis of a proposed sampling program at a former Ferro-alloy work in Gullspång, Sweden.

Probability	Uniform PDF	Triangular PDF	Normalised normal PDF	Log-normal PDF
Prior: $P(C^+)$	0.789	0.740	0.647	0.566
Prior: $P(C^-)$	0.211	0.260	0.353	0.434
$P(D^+\|C^+)$	0.982	0.965	0.963	0.932
$P(D^-\|C^+)$	0.018	0.035	0.037	0.068
$P(D^+\|C^-)$	0.051	0.078	0.052	0.077
$P(D^-\|C^-)$	0.949	0.922	0.948	0.923
$P(D^+)$	0.786	0.735	0.641	0.560
$P(D^-)$	0.214	0.265	0.359	0.440
$P(C^+\|D^+)$	0.986	0.972	0.972	0.941
$P(C^-\|D^+)$	0.014	0.028	0.028	0.059
$P(C^+\|D^-)$	0.067	0.098	0.067	0.088
$P(C^-\|D^-)$	0.933	0.902	0.933	0.912

Estimation of probabilities P[sample/state] is performed according to equations 5.10 – 5.20. In calculation of Table 5.1 a value $CV_{OU} = 0.4$ has been used, but we will later study the effects of different sample uncertainties (section 5.4.6). Some interpretations of the probabilities in Table 5.1 will be given as examples. The probability $P(D^+|C^+)$ can be interpreted as: *The probability of correctly detecting contamination is 93 % - 98 %, depending on the choice of prior distribution*. The most important probability of failure (associated with the most serious consequence) is $P(D^-|C^+)$, which can be interpreted as: *The probability of not detecting contamination when the site is in fact contaminated is 2 % - 7 %, depending on the choice of prior distribution*.

The probabilities P[sample] are estimated by equations 5.21 – 5.22. The result in Table 5.1 indicates that the probability of classifying the site as contaminated based on sample information is 56 % - 79 %, depending on the choice of prior distribution.

The last set of probabilities, P[state/sample], is estimated by Bayesian updating according to equations 5.23 – 5.26. The first of the four probabilities, $P(C^+|D^+)$, can be inter-

preted as: *The probability of the site being contaminated when contamination has been detected is 94 % - 99 %, depending on the choice of prior distribution*.

5.4.5 Step 4: Cost estimation

The data worth analysis requires estimates of benefits and costs for the different decision alternatives, and failure costs for making wrong decisions. However, the purpose of the thesis is not to go into details on cost-estimation. Therefore, only rough estimates of costs are made. These estimates would probably change somewhat if the costs were analysed in detail. All estimated benefits and costs are presented in Table 5.2 in the unit 1000 SEK (1000 Swedish Krona; 1 € ~ 9 SEK). It is assumed that all benefits and costs have been transformed to net present value.

Table 5.2 *Payoff table with estimated benefits and costs. Estimates are made for data worth analysis of a proposed sampling program at a former Ferro-alloy work in Gullspång, Sweden.*

Decision alternative	**True state of the site**	**Benefit, *B*** (k SEK)	**Cost, *C*** (k SEK)	**Failure cost, *Cf*** (k SEK)	**Total *B-C-Cf*** (k SEK)
Remediation	Contaminated	-	1 000	-	-1 000
	Not contaminated	-	1 000	-	-1 000
No remediation	Contaminated	-	-	2 000	-2 000
	Not contaminated	-	-	-	0

A benefit is assumed to arise when the site is contaminated and remediation is performed. The benefit constitutes the increase in the value of real estate property due to the remediation. At the particular site of interest, such increase in value is expected to be small because of low demand on the local real estate market. Benefits are therefore neglected.

The investment cost in remediation is the cost for excavation, transportation, and disposal of the contaminated soil. This cost arises when the decision is taken to remediate, whether or not the site in reality is contaminated. It is of course a simplification to say that remediation costs are the same even if the site is not contaminated, but we assume that monitoring during the excavation is unable of determining the true state of the site, and therefore remediation will proceed to full extent. The remediation cost is assumed to 1 000 000 SEK. The cost of the other decision alternative, not to remediate, is zero.

One type of failure is to decide to remediate when the site is not contaminated, as discussed above. This type of failure does not involve any failure cost, only an unneeded investment cost of 1 000 000 SEK. A second and more important type of failure occurs when the decision is made not to remediate when the site is in fact contaminated. This will lead to undesirable human and environmental effects, which are defined as a failure cost. Estimation of this cost is difficult and we will not go into detail on this matter. A low estimate of the failure cost is to set it equal to the double remediation cost, 2 000 000 SEK. This is a quite low estimate, but the effect of different failure cost estimates is studied in more detail in section 5.4.6.

In addition, the cost of the sampling program should be estimated. The total cost for the sampling program is assumed to follow the simple cost function in equation 5.36. The cost per sample, C_s, is estimated to 4 000 SEK including sampling, laboratory analyses, and a simple evaluation of the sampling results (estimation of mean concentration).

5.4.6 Step 5: Data worth estimation

Prior analysis

The prior objective function value is calculated by equations 5.27 – 5.28. The result of the calculation is presented in the prior decision tree in Figure 5.6. Because there is no benefit, only costs, the objective function will be negative. The objective function Φ_{prior} is estimated to -1 000 000 SEK, using the uniform prior PDF.

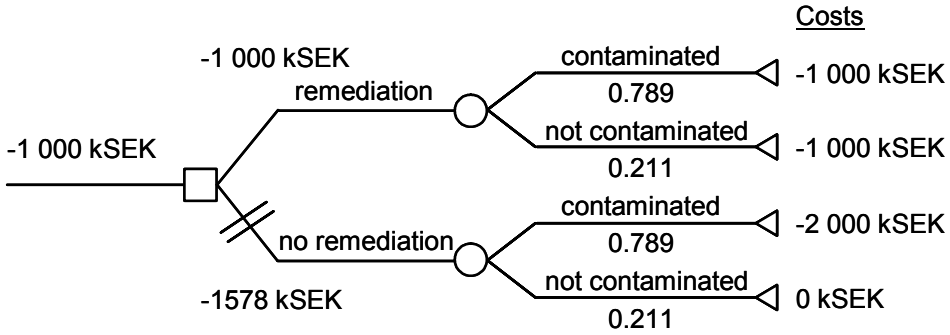


Figure 5.6 Decision tree for the prior analysis with probabilities, costs, and objective function values, based on a uniform prior distribution.

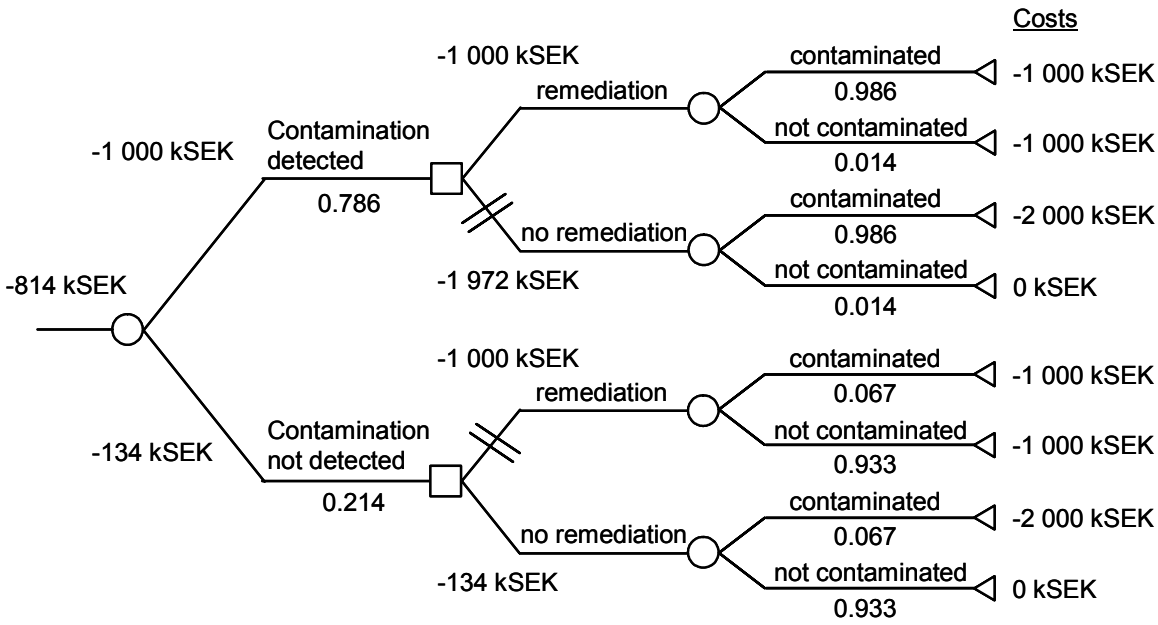


Figure 5.7 Decision tree for the preposterior analysis with probabilities, costs, and objective function value, based on a uniform prior distribution.

Preposterior analysis

The preposterior objective function value is calculated by equations 5.29 – 5.31. The result of the calculation is presented in the preposterior decision tree in Figure 5.7. The objective function $\Phi_{preposterior}$ is estimated to -814 000 SEK, using a uniform prior PDF.

Data worth

The value of the sampling program is estimated by equation 5.32:

$$EVSI = \Phi_{preposterior} - \Phi_{prior} = \left(-814\ 000 + 1\ 000\ 000\right)\ SEK = 186\ 000\ \ SEK$$

This data worth can be compared to the expected value of perfect information, which is the upper bound of data worth. The expected value of perfect information is estimated by equations 5.33 – 5.34:

$$EVPI = \Phi_{\max} - \Phi_{prior} = \left(-789\ 000 + 1\ 000\ 000\right)\ SEK = 211\ 000\ \ SEK$$

The reliability (James et al., 1996b) of the sampling program is estimated to:

$$reliability = \frac{EVSI}{EVPI} = \frac{186\ 000}{211\ 000} = 88\ \%$$

The cost-efficiency of the sampling program, expressed by the expected net value according to equation 5.37, is:

$$ENV = EVSI - C_p(n) = \left(186\ 000 - 12 \cdot 4000\right)\ SEK = 138\ 000\ \ SEK$$

This result indicates that the sampling program is cost-efficient and that it is worthwhile to carry out the sampling. The result is based on a uniform prior PDF. Table 5.3 summarises the results for all four prior distributions.

Table 5.3 Result of data worth analysis of a proposed sampling program at a former Ferro-alloy work in Gullspång, Sweden. Four different prior distributions of mean concentration are evaluated.

Result	**Uniform PDF**	**Triangular PDF**	**Normalised normal PDF**	**Log-normal PDF**
EVPI (k SEK)	211	260	353	434
EVSI (k SEK)	186	214	311	362
ENV (k SEK)	138	166	263	314
Reliability	88 %	82 %	88 %	83 %

From Table 5.3 it is quite clear that the way of expressing prior information has an impact on the result. For example, the largest expected net value is more than twice the lowest value, and the reliability of the sampling program ranges from 82 % to 88 %.

The results in Table 5.3 are based on one single sampling program. However, there may be other sampling programs that are more cost-efficient than the proposed one. Two factors in the sampling program that can affect the cost-efficiency will be studied, i.e. the sample uncertainty and the number of samples.

Sample uncertainty and the number of samples

The same calculation of ENV as above has been performed with a range of values on sample uncertainty CV_{OU} and on the number of samples n. The range for CV_{OU} is from 0.2 to 2.0, and the number of samples ranges from 2 to 60. The results are presented in Figures 5.8 – 5.11 for each of the four types of prior distributions.

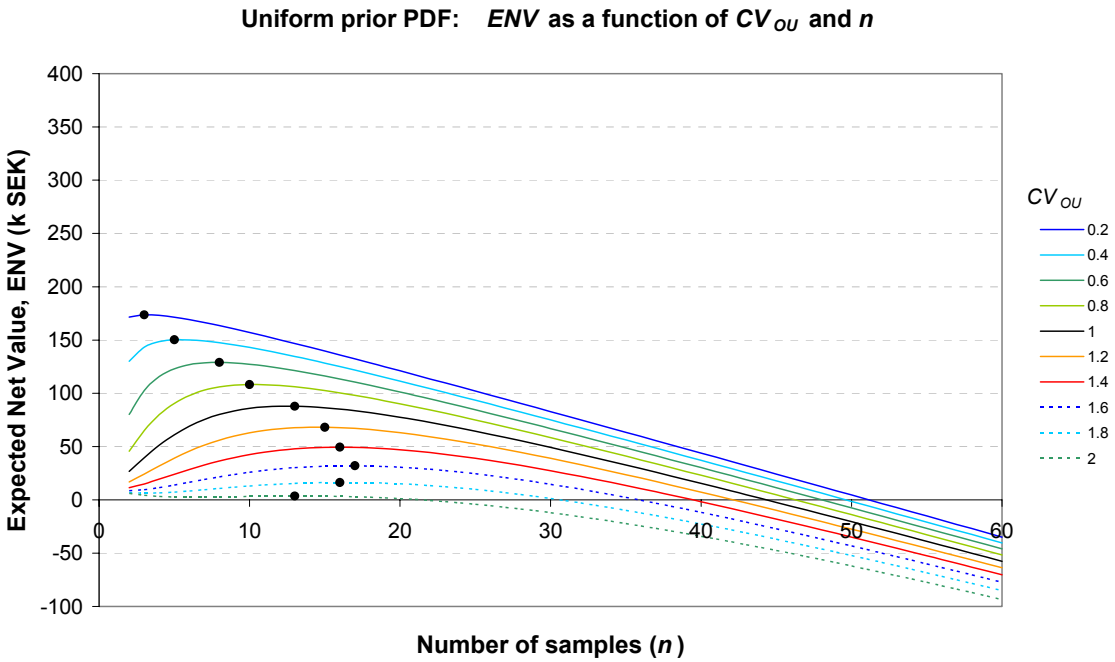


Figure 5.8 *Expected net value as a function of sample uncertainty and number of samples. The sample uncertainty is expressed with a coefficient of variation. The prior information about the mean concentration is expressed with a uniform PDF.*

Figures 5.8 – 5.11 show that the expected net value of the sampling program is positive for most sampling programs, except some with very large sample uncertainty and those with large number of samples. The shapes of the curves are quite similar for all prior distributions but the absolute values of ENV differ significantly. The log-normal distribution results in the highest ENV, whereas the uniform PDF produces to lowest. The reason is that the prior probabilities P(C+) and P(C-) of the log-normal distribution are quite close to 0.5, i.e. almost as large for *not contaminated* as for *contaminated* (Table 5.1). This results in a large value of the data worth. In the case of a uniform PDF, the prior probability P(C+) is much larger than P(C-) and therefore new sample information has a lower value.

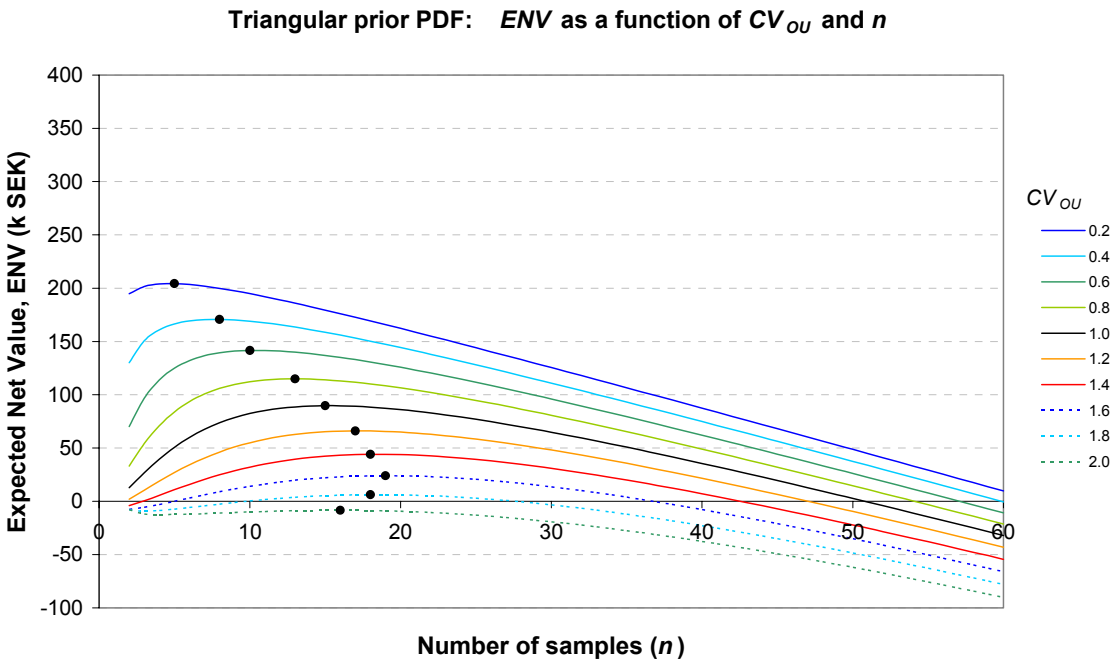


Figure 5.9 *Expected net value as a function of sample uncertainty and number of samples. The sample uncertainty is expressed with a coefficient of variation. The prior information about the mean concentration is expressed with a triangular PDF.*

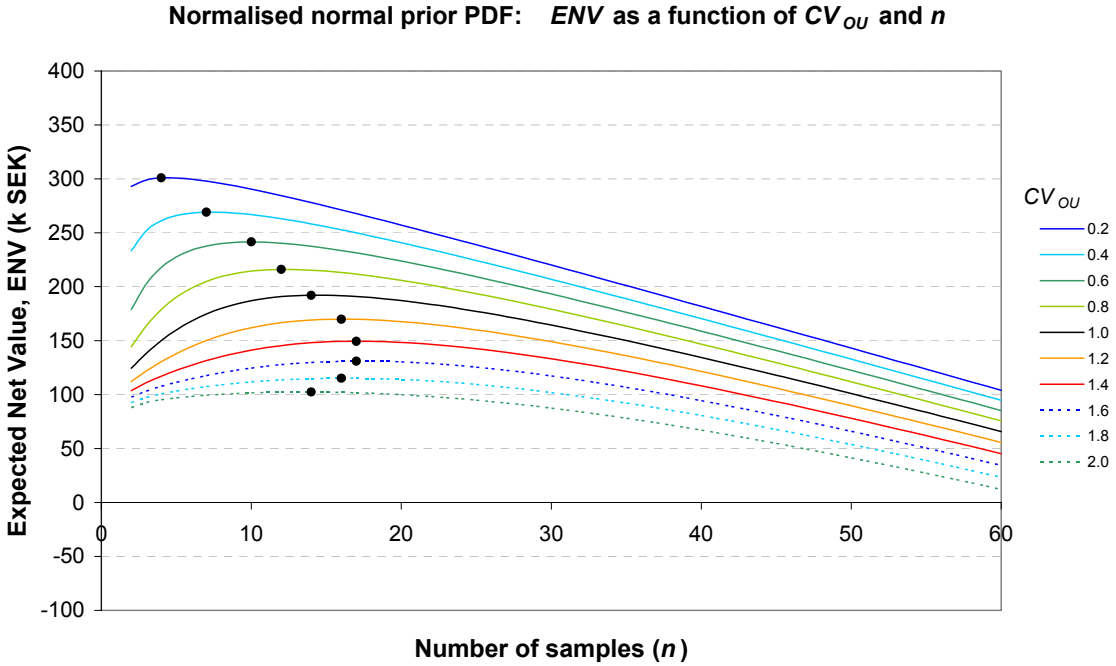


Figure 5.10 *Expected net value as a function of sample uncertainty and number of samples. The sample uncertainty is expressed with a coefficient of variation. The prior information about the mean concentration is expressed with a normalised normal PDF.*

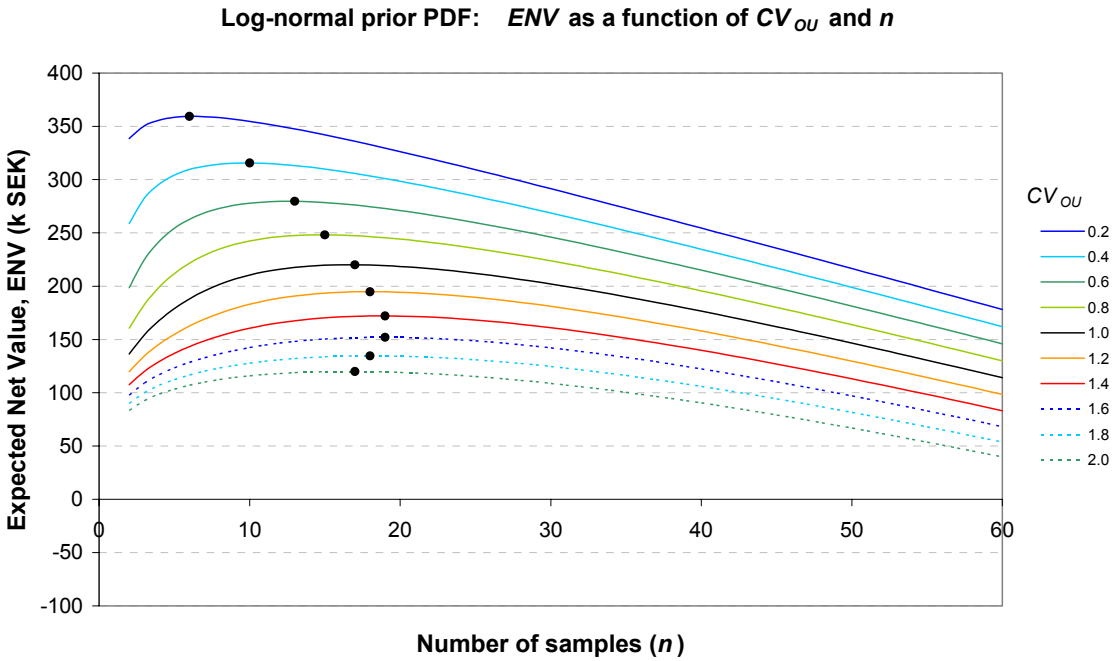


Figure 5.11 *Expected net value as a function of sample uncertainty and number of samples. The sample uncertainty is expressed with a coefficient of variation. The prior information about the mean concentration is expressed with a log-normal PDF.*

Independent of the selected prior distribution, we can say that the sampling program of 12 samples with a CV_{OU} of 0.4 is cost-efficient, since ENV is lager than zero. As a matter of fact, it appears that with this sample uncertainty, we are quite close to the optimal sampling program, i.e. the peak of the curve (the optimal number of samples is indicated with black dots in Figures 5.8 – 5.11). Starting with 12 samples and keeping the same CV_{OU} of 0.4, we can see that the optimum number of samples is in the range of 5 to 10, depending on prior PDF. If more samples than the optimal number are collected, the sampling will still be cost-efficient of to a point where ENV is reduced to zero.

The optimal number of samples increases when sample uncertainty increases, but only up to a certain level of uncertainty ($CV_{OU} \sim 1.5$). At larger uncertainties the optimal number of samples is again lower. The calculations have been performed with a fixed sampling cost but in reality, laboratory analysis may be exchanged with cheaper field screening techniques with larger uncertainty. This may lead to a different situation.

A reason why it is cost-efficient to collect so many samples is that in our problem, the most likely value (200 mg/kg) is quite close to the action level (250 mg/kg), and that the sample cost is low compared to the failure cost. Note that the cost-efficiency of sampling plans with many samples assumes that all samples are collected and evaluated together. If the sampling is performed in stages the problem will be different. If some of the samples are collected and evaluated in a first and separate stage, it is not certain that that it would be cost-efficient to continue to collect the rest of the samples later on. Whether or not a second sampling stage should be carried out depends on the outcome of the first set of samples. A second analysis of data worth should be undertaken before the second sampling stage is initiated. This is in accordance with Figure 3.3.

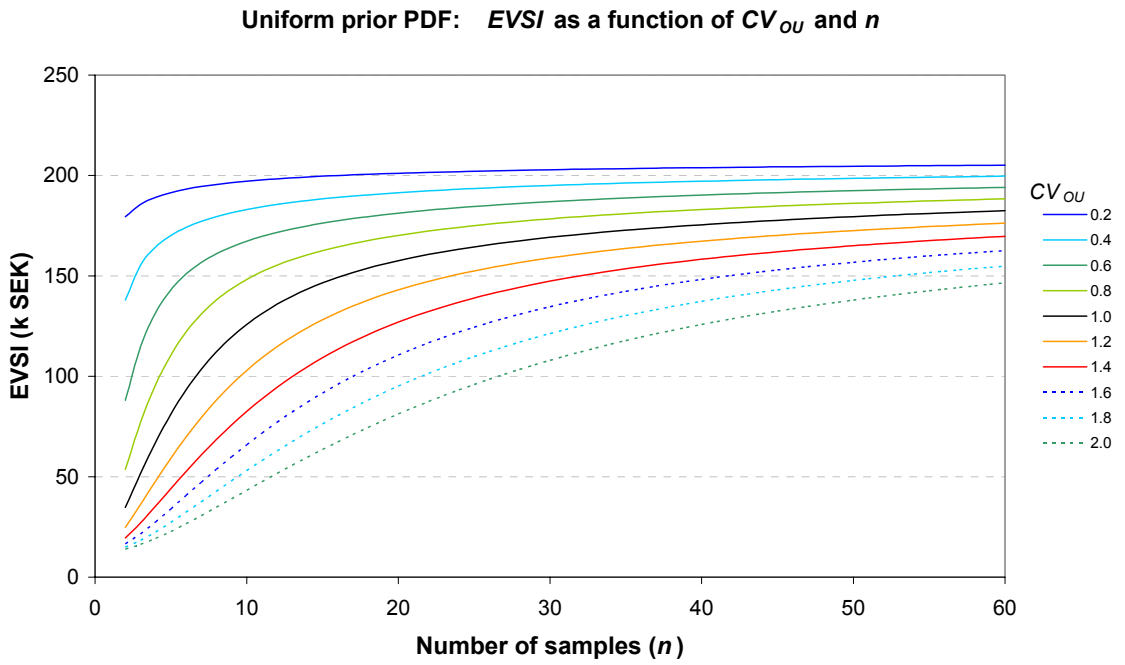


Figure 5.12 *Expected data worth (EVSI) as a function of sample uncertainty and number of samples. The sample uncertainty is expressed with a coefficient of variation. The prior information about the mean concentration is expressed with uniform PDF.*

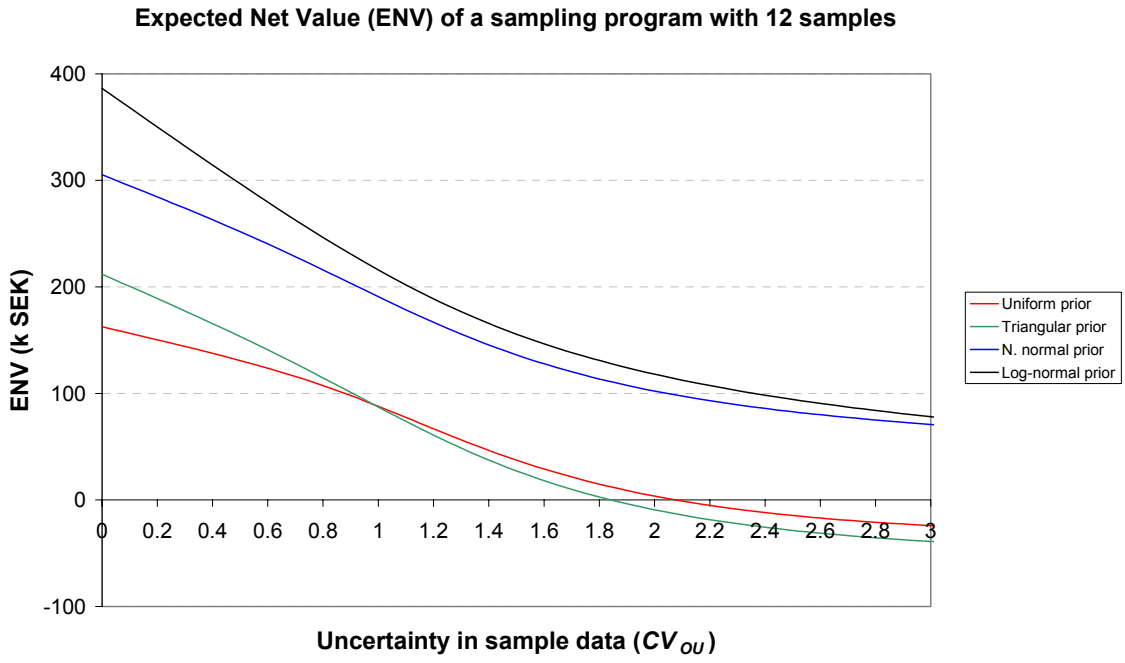


Figure 5.13 *Expected net value of a sampling program of 12 samples as a function of sample uncertainty (coefficient of variation). The prior information about the mean concentration is expressed with four different PDFs.*

If we do not consider sampling cost, the data worth can instead be expressed by EVSI. In Figure 5.12, EVSI is plotted for a uniform prior PDF. The difference in Figure 5.12 from the plot of ENV in Figure 5.8 is that EVSI always increases when the number of samples is increased. The worth of the sampling program exhibits a sharp increase for the first 5 to 20 samples (depending on sample uncertainty). The increase in EVSI is more modest for sampling program with more than 5 to 20 samples.

In Figure 5.13 the expected net value is plotted against the uncertainty in sample data. With very large sample uncertainty, around 1.9, there will be no net value of the sampling program if the prior information has been represented by a uniform or triangular PDF. The reason is that the uncertainty is simply too large, so large that the proposed sampling program cannot solve this sampling problem in a cost-efficient way.

The effect of prior information

As described previously, the choice of prior PDF may have a significant influence on the absolute value of ENV, according to Figures 5.8 – 5.11. Not only will the type of PDF affect the result but also the assumptions of minimum value, most likely value, and maximum value of the mean concentration. An extreme case is illustrated in Figure 5.14 for three different values of the maximum concentration b. The upper curve is identical to the curve for $CV_{OU} = 0.4$ in Figure 5.11. As illustrated, a small difference in estimate of b may lead to significantly different expected net value. The difference in ENV between the curves for $b = 500$ mg/kg and $b = 400$ mg/kg is more than 100 000 SEK. This indicates that the expected net value can be quite sensitive to prior estimates in certain situations. In the case in Figure 5.14, the reason for the sensitivity is a high peak of the log-normal prior distribution close to the action level.

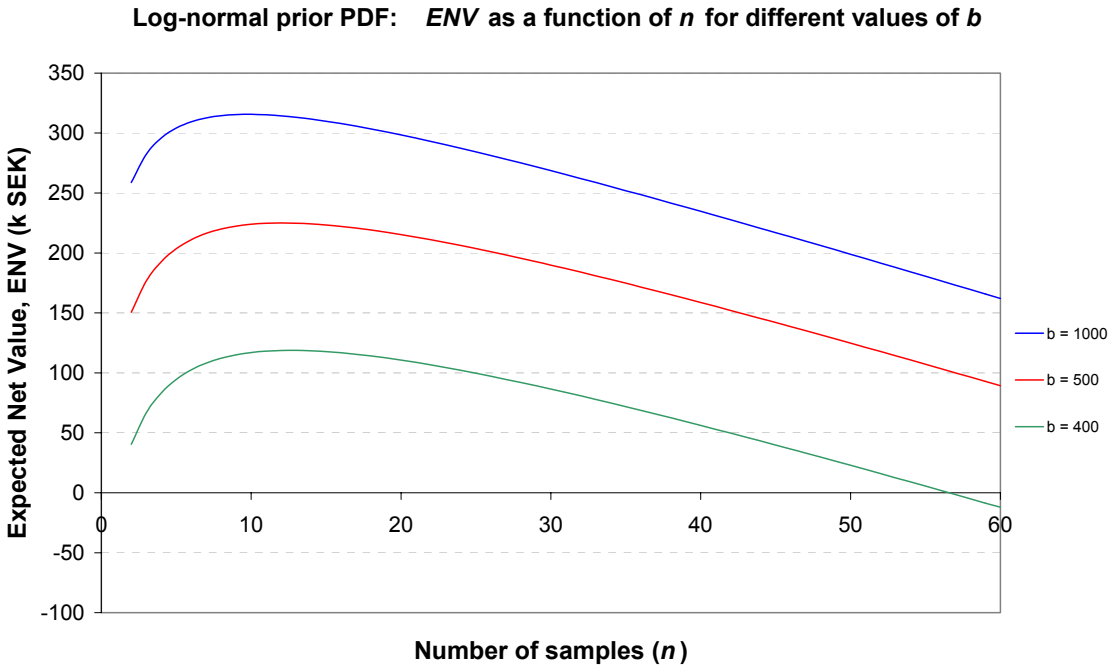


Figure 5.14 *The expected net value of a sampling program for three different prior assumptions of the maximum mean concentration for a log-normal prior PDF. The sampling program consists of 12 samples with a sample uncertainty of $CV_{OU} = 1.0$.*

The effect of cost estimates

The cost estimates in Table 5.2 are uncertain and therefore their effect on the data worth should be analysed. Here, the effect of different estimates of the failure cost has been analysed because this costs is believed to be the most uncertain and difficult to estimate. If the decision is made not to remediate when the site in reality is contaminated this will lead to a cost of failure, see Figure 3.6. This cost is the environmental cost of leaving the area contaminated, as a present and future treat to humans and the environment.

In the calculations, the cost of remediation has been maintained according to Table 5.2. The effects of different estimates of failure cost are presented for three different sample uncertainties. The results are illustrated in Figures 5.15 – 5.17. At low failure costs there will be no data worth and the reason is that it is more cost-efficient to make the decision to remediate directly, without performing any sampling. The data worth will also be zero if the failure cost is very high. This is a result of the high probabilistic risk. Other values of failure cost will produce a positive value of data worth (EVSI) but it is quite sensitive to the estimated failure cost.

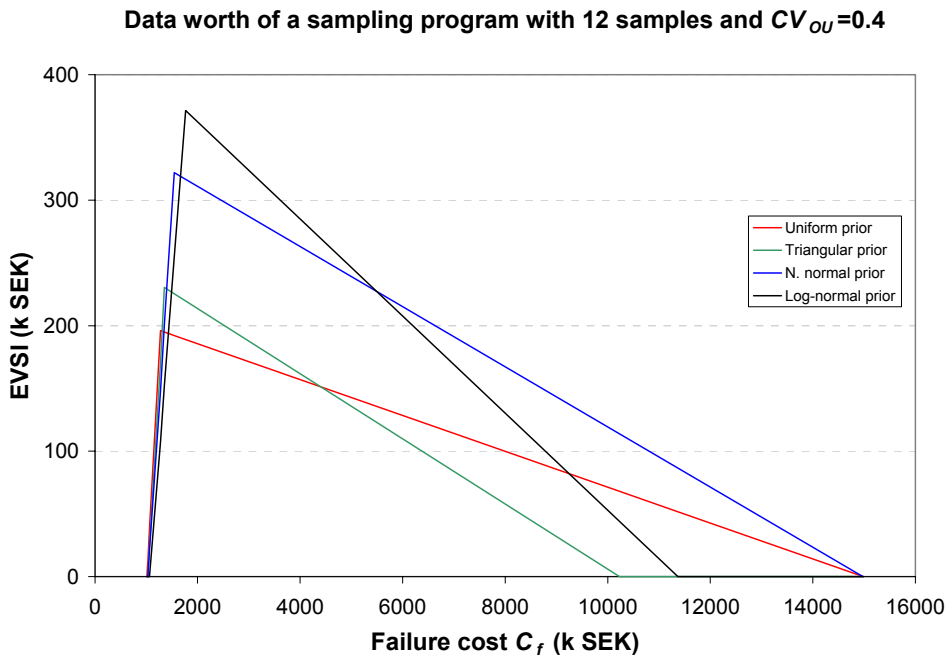


Figure 5.15 The data worth (EVSI) of the sampling program as a function of failure cost. The uncertainty in sample data is moderate, CV_{OU} = 0.4.

Figure 5.15 – 5.17 represent three different uncertainties in sample data. A comparison clearly indicates that the curve of EVSI is steeper for high uncertainty in sample data than for small. This implies that the data worth analysis gets more and more sensitive to the estimated failure cost as sample uncertainty increases.

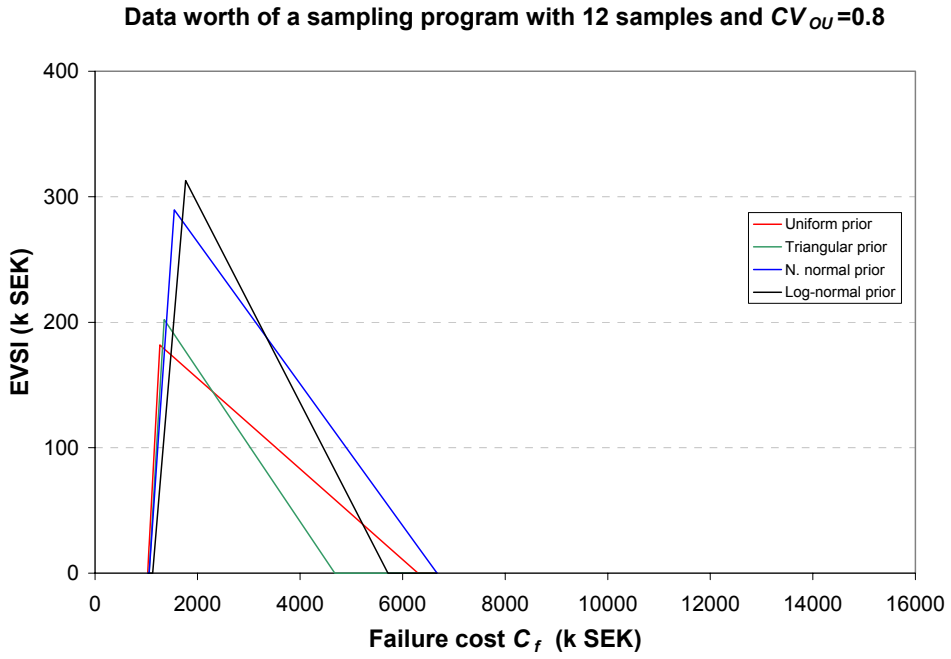


Figure 5.16 The data worth (EVSI) of the sampling program as a function of failure cost. The uncertainty in sample data is high, CV_{OU} = 0.8.

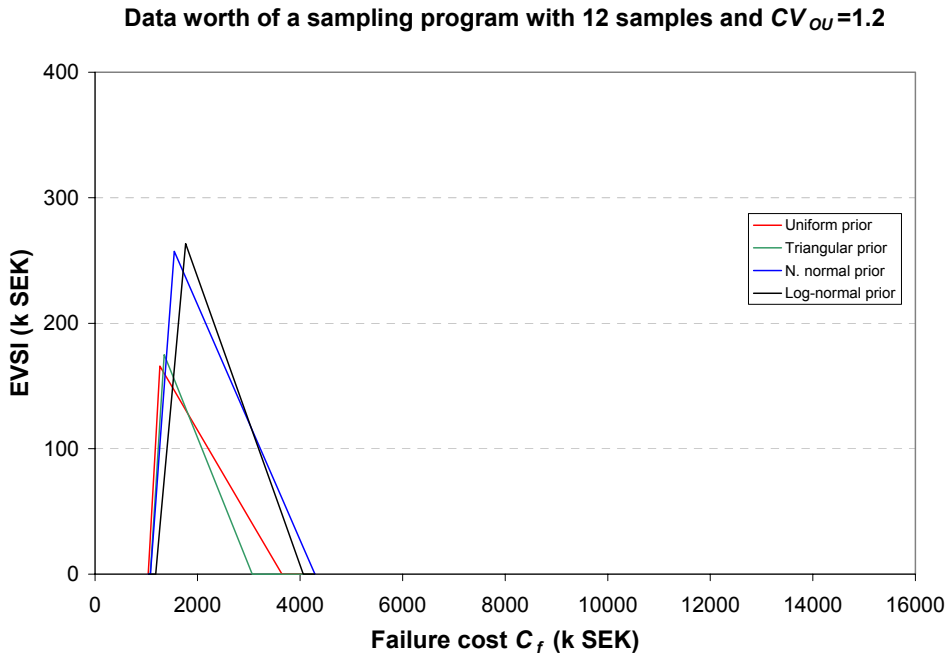


Figure 5.17 The data worth (EVSI) of the sampling program as a function of failure cost. The uncertainty in sample data is very high, CV_{OU} = 1.2.

6 CONCLUSION AND DISCUSSION

6.1 Uncertainty in sampling

The methodology presented in chapter 4 aims at estimation of uncertainty in soil sample data. There are two practical applications of the methodology; (1) estimation of uncertainty in sample data prior to sampling, or (2) analysis of an existing set of sample data to draw conclusions about the involved uncertainties and how they can be reduced. The approach taken in this thesis is the first one, i.e. prior estimation of sample uncertainty. However, it is believed that the methodology will be useful for most soil sampling problems at contaminated sites, even after sampling has been performed. The methodology gives a better understanding of the sampling problem and the involved uncertainties, issues that today often is not given enough attention. This understanding is necessary in many situations, for example when the sampling technique is selected, when choosing between laboratory analyses and field screening methods, and during evaluation of sample data.

The presented application of the methodology in section 4.5 illustrates the importance of certain sampling uncertainties that often is overlooked. The application indicates that sampling uncertainty is more important than analytical uncertainty, at least for this and similar problems. This is in accordance with comments found in the literature, e.g. by van Ee et al. (1990). In the application, a combination of reasoning and calculation is used during estimation of uncertainties. At present state of knowledge, subjective reasoning is unavoidable during the estimation because there is not much information available of the uncertainty of different sampling techniques for contaminated soil. Today, only qualitative statements of the reliability of different sampling techniques are available. Detailed studies of different techniques are desirable and would make possible better estimation of sampling uncertainty. This is especially important for the materialisation uncertainty (delimitation and extraction of the sample), which is an often overlooked type of uncertainty and one that can vary significantly between different sampling techniques.

One limitation of the methodology is that the theoretical foundation for some types of uncertainty, especially the fundamental variability, can be questioned for some problems, mainly because the theory was developed for mining problems and not for environmental problems. For example, the situation is very different if the contaminant is composed of a more or less mobile aqueous phase liquid (APL) than if it consists of solid particles like lead shots. Equations for estimation of the fundamental variability are based on models, and like all models they contain simplifications and assumptions. Some of these will be briefly discussed.

First of all, the contaminant is assumed to be attached to particles, or consist of liberated particles itself (like lead shots or paint chips). This assumption may hold for metals and some other contaminants but for common organic compounds, or mixtures of compounds, the situation may be different, especially at high concentrations. Oil or other APLs for example, will occupy the pore space between the particles as blobs and ganglia, which can form a very different situation.

Secondly, one assumption made in the estimation of fundamental variability is that the contaminant concentration of a particle is correlated with the density of the particle (Pitard, 1993), which may be a reasonable assumption for mining problems. However, for contaminated soil problems it is believed that the specific surface is more important. It is not known if the above assumption introduces significant errors when the theory is applied to environmental problems.

Although the conceptual description of the different sampling uncertainties is correct in the presented theory, a conclusion is that the theoretical foundation for quantitative estimation needs to be thoroughly analysed, especially for the fundamental variability, in order to validate the theory for certain environmental problems. The basic assumptions should be reconsidered, taking different types of contaminants and contamination scenarios into account. The result should be used to formulate the fundamental variability in a proper way for different types of problems.

The principles in chapter 4 apply to soil but it is believed that the methodology could also be used to estimate uncertainty in sediment or groundwater sampling. Some modifications of the theory would be needed, especially for groundwater sampling, because water is a moving medium in contrast to soil.

The terminology of sample uncertainty in the thesis has been taken from the sampling theory of particulate materials, with some changes made. The appropriateness of certain terms could be questioned because they are quite long and complicated. A well-structured and simplified version of the theory, together with well thought-out names could be an important contribution to practitioners in the field of contaminated soil. The presented methodology can also quite easily be implemented in spreadsheet software.

One important lesson to be learned from chapter 4 is that all sampling problems are scale problems. Not only does the sample uncertainty depend on the scale we choose to consider, but also the concentration itself is defined by the scale. The various uncertainties and errors will be more or less important depending on how the problem is defined. One extreme-case is if the sample is defined to only be representative of itself. In this case there is no sampling uncertainty at all (but we are not allowed use the sample data to draw any conclusions about the surrounding soil). The other extreme-case is when a sample is supposed to be representative of the whole site. In addition, measurements of concentration always imply mean concentrations, it is more a question of the scale at which we want to know the mean. It is a mean concentration in a certain volume we refer to when we use the concept “concentration”. If we do not want to estimate the mean concentration for a whole site we often want to know the mean for a sub-region of the site, or at least for a certain volume of soil. Therefore, studying sample uncertainty without consideration of the scale is believed to lead to serious mistakes.

6.2 Data worth analysis

A methodology for data worth analysis including sample uncertainty is presented in chapter 5. The described approach for estimation of the value of a sampling program is a complete and applicable methodology for the particular sampling objective, with a computer program presented in Appendix 2. The methodology effectively answers the commonly asked question of how many samples one should collect. This simple ques-

tion is found to be embarrassingly difficult to answer even for statisticians (Lindley, 1997). By applying an approach of cost-efficiency, this question can be answered.

Based on the presented application in section 5.4 and additional test runs with the computer program, some general recommendations can be given for the design of sampling programs and for analysis of data worth, based on the objective of estimating mean concentration and no correlation between sample points:

- From a perspective of cost-efficiency, the optimal number of samples is usually in the order of 5-20 for reasonable values of sample uncertainty and the different costs. The optimal number of samples is lower for high sampling cost than for low.

- The data worth analysis is sensitive to failure cost and it is recommended to use different estimates of failure cost to study the influence. The analysis is extremely sensitive in situations where the failure cost is nearly the same as the remediation cost. If the failure cost is lower than the remediation cost it will not be cost-efficient to perform any sampling at all. In this case the site could be remediated based on prior information only.

- The sensitivity of failure cost increases with increased sample uncertainty. This implies that when an uncertain sampling technique is used, more effort should be put on estimation of failure cost.

- The prior PDF should be selected and defined carefully because it may have a significant influence on the data worth, at least on the absolute value. PDFs with scattered probability mass like the uniform PDF result in lower net expected value than PDFs with more concentrated probability mass, at least when the PDFs are defined according to the procedure in this thesis (see discussion below). A recommendation is to consider the prior probabilities carefully when the prior PDF is defined, since the prior probabilities seem to have a significant influence on the data worth (see below).

- The data worth does not appear to be extremely sensitive to sample uncertainty. With increasing sample uncertainty, data worth decreases relatively constant. The implication of this is that sample uncertainty may not be so important to consider if the question about a sampling program to be answered is a simple “yes” or ”no” question: *Is the sampling program cost efficient or not cost efficient?* On the other hand, if the question is how cost efficient a sampling program is, sample uncertainty should be included in the analysis.

The approach in this thesis does not take spatial correlation into account. One implication of this is that if correlation exists between samples, this is not considered in the analysis. This means that some information that is available in sample data is not used, i.e. the full potential of sample data is not utilised. Therefore, data worth estimates in the thesis will be lower than the potential value of the sampling program. Taking spatial correlation into account would therefore produce larger data worth estimates. One recommendation for future work is to apply data worth estimation to spatial problems and to take sample uncertainty into account, as indicated by Freeze et al. (1992).

The methodology considers only one of several possible sampling objectives, i.e. the objective of estimating the mean concentration. Decision-making based on mean concentration is but one example of a situation where RCB decision analysis can be used as a tool. Site investigations for contaminated land problems are also designed for other objectives, as discussed in section 4.1. Typical examples include identification of hot spots and delimitation of the spatial distribution of a contaminant. Therefore, data worth estimation methods also need to be developed for such sampling objectives, taking sample uncertainty into account.

As mentioned, it is quite common that sampling plans for real-world problems are developed with several sampling objectives in mind. This makes data worth estimation more problematic because the methodology must take multiple objectives into account. An important development would be to apply data worth estimation to real-world multiple objective problems. This would allow data worth to be estimated for a wide range of situations where today only subjective decisions can be made concerning the cost-efficiency of site investigations.

The presented approach to data worth analysis has limitations regarding some other aspects. Only one chemical substance is considered in the analysis, whereas in reality several contaminants are usually studied at the same time within a single site investigation. Another aspect is that one remedial action at a contaminated site often is carried out with several contaminants in mind, e.g. excavation of contaminated soil. In other situations, different contaminants may require different remedial actions. More work is needed to take these aspects of multiple contaminants into account in a structured way.

Another aspect that needs to be considered further is estimation of failure cost. Chapter 5 indicates that data worth can be quite sensitive to failure cost, which is in accordance with other findings in the literature, e.g. Russell and Rabideau (2000). Therefore, a recommendation is to study the sensitivity using different estimates of failure cost during data worth analysis.

In addition, the issue of expressing prior information needs to be further analysed. Prior information is in the thesis represented by PDFs. The prior PDFs are based on subjective estimates of minimum, most likely, and maximum values of the mean concentration. This means that prior information about *concentration* is expressed. As a consequence, different PDFs will produce different prior probabilities depending on the shape of the distribution. This leads to a paradox: Uniform distributions, that are used to represent less certain prior information compared to e.g. log-normal distributions, result in lower estimates of data worth. One would expect the opposite to be true since data is of more worth in uncertain situations. The reason is that the methodology of defining prior PDFs in the thesis, actually leads to higher prior probabilities for the uniform distribution than the log-normal. For example, when using a uniform distribution the prior probability of contaminated site, P($C+$), may be about 0.8, whereas a log-normal distribution based on the same parameters may produce a prior probability of nearly 0.5.

A different approach would be to specify the prior probabilities P($C+$) and P($C-$) and base the shape of the PDF on these *probabilities* instead of estimates of concentration. It is possible that such an approach would be less sensitive to the choice of prior PDF than the approach taken in the thesis. In addition, this approach might be more in accordance with expert knowledge about contaminated land. For example, it is reasonable to believe

that it is easier to make good estimates of prior probabilities than of the maximum value of μ (denoted b in the thesis).

Sample uncertainty has successfully been included in data worth analysis in the approach taken here. This implies a more realistic analysis compared to ignoring uncertainty. Including sample uncertainty requires more complicated calculations but these can be automated by computer software. An interesting task would be to estimate the expected value of including uncertainty (EVIU) in the analysis, as described by Morgan and Henrion (1990). Including uncertainty implies a cost due to additional work and the question is how much better of one will be by including it.

6.3 Protective actions

A methodology for selecting between alternative protective actions for water supplies along railways is presented in the paper in Appendix 1. It is based on RCB decision analysis, as described in chapter 3. So far, experience from practical applications along railways is sparse but a similar framework developed for roads has been applied at about 30 sites (Rosén, 2002). Therefore, some conclusions about the methodology can nevertheless be drawn.

The framework contains different types of uncertainty that may influence the risk-estimates in various ways. One type of uncertainty that is taken into account is parameter uncertainty, which is treated by stochastic simulation. Model uncertainty on the other hand, is not considered explicitly. Some authors point out that model uncertainty generally is more important than parameter uncertainty (Morgan and Henrion, 1990). However, if this is the case in the presented approach is not known. One way of handling model uncertainty is to perform multiple estimations with a set of different models (see section 2.5). The models could include both analytical models, as well as different conceptualisations of a particular site. Such an approach would supply more information about the importance of model uncertainty but make practical applications complicated and tiresome. However, the uncertainty in the presented framework will not be significant as long as it influences the decision alternatives in similar ways (Rosén, 2002), which is believed to be the case for many problems. The uncertainty will be unacceptable only when it is so large that a change in decision may occur. Therefore, large uncertainties do not automatically imply a problem.

Experience from applications of the methodology indicates that it is important how failure is defined. In principle, there are many ways of defining failure for a water resource. In practice, two ways are dominating; (1) using a *transport time* criterion, or (2) using a *concentration* criterion. A transport time criterion is the most commonly used failure criterion. Failure occurs if the transport time of the contaminant from the point of accident to a compliance boundary is shorter than a specified time limit, i.e. the contaminant transport is so rapid that there is not enough time for remedial actions to take place. The compliance boundary can for example be located at a certain distance from the railway or at the abstraction wells of a water supply. It is often more difficult to use a concentration criterion in the analysis because this means that a certain level of contamination is accepted. Often, this approach is not met with sympathy, especially if the water resource is used for supplying drinking water to the public.

Another uncertainty in the methodology that will need additional study is the estimation of failure cost, especially the *in situ* values (National Research Council, 1997). It should be mentioned that methods for estimation of *in situ* values (environmental values) exist. What is needed is more practical experience of these methods.

The framework for selection between alternative protective actions has proven to be useful for hydrogeological problems along roads and railways. The same principles could also be used for other environmental problems, with minor changes to the framework. One such application could be selection between alternative mitigating and remedial actions for contaminated land. A similar framework for selection between alternative remediation strategies at a municipal landfill has also been presented (Norrman, 2000b). It is our belief that there are numerous other decision problems that could benefit from a similar approach to the presented one.

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.

Back, P.-E., 2001. Sampling Strategies and Data Worth Analysis for Contaminated Land. B 486, Department of Geology, Chalmers University of Technology, Göteborg.

Back, P.-E. and Rosén, B., 2001. Riskanalys av områden där järnvägstrafik berör vattentäkter och andra vattenresurser - Metodikutveckling. Varia 513, Swedish Geotechnical Institute, Linköping.

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.

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.

Burmaster, D.E., 1996. Benefits and Costs of Using Probabilistic Techniques in Human Health Risk Assessments - With an Emphasis on Site-Specific Risk Assessments. Human and Ecological Risk Assessment, 2(1): 35-43.

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.

Deutsch, C.V. and Journel, A.G., 1998. GSLIB, Geostatistical Software Library and User´s Guide. Oxford University Press, New York, 369 pp.

Eklund, H.S. and Rosén, L., 2000. Hydrogeological decision analysis for selection of alternative road stretches and designs. In: S.e. al. (Editor), Groundwater: Past Achievements and Future Challenges. Balkema, Rotterdam, pp. 923-928.

Ellison, S.L.R., Rosslein, M. and Williams, A., 2000. Quantifying Uncertainty in Analytical Measurement. EURACHEM / CITAC Guide CG 4, QUAM:2000.1, EURACHEM / CITAC.

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., 1993. Sampling Strategy Guidelines for Contaminated Land. 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, 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., 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.

Gleser, L.J., 1998. Assessing Uncertainty in Measurement. Statistical Science, 13(3): 277-290.

Goodman, D., 2002. Extrapolation in Risk Assessment: Improving the Quantification of Uncertainty, and Improving Information to Reduce the Uncertainty. Human and Ecological Risk Assessment, 8(1): 177-192.

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., Bourgine, B., Dubois, D., Fargier, H. and Côme, B., 2003. Hybrid Approach for Addressing Uncertainty in Risk Assessment. Journal of Environmental Engineering, 129(1): 68-78.

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.

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. Kostecki (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/hugintro/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.

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.

Kaplan, S. and Garrick, B.J., 1980. On The Quantitative Definition of Risk. Risk Analysis, 1(1): 11-27.

Karlström, J.M., 2001. Provtagningsmetodik för förorenad mark - med jämförelse mellan hollow stem auger och skruvprovtagare. Master Degree Thesis, Lund Institute of Technology, Lund, 76 pp.

Keith, L.H. (Editor), 1996. Principles of Environmental Sampling. American Chemical Society, Washington, DC, 848 pp.

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.

LeGrand, H.L. and Rosén, L., 2000. Systematic Makings of Early Stage Hydrogeological Conceptual Models. Ground Water, 38(6): 887-893.

Lindley, D.V., 1997. The choice of sample size. The Statistician, 46(2): 129-138.

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.

McMahon, A., Heathcote, J., Carey, M., Erskine, A. and Barker, J., 2001. Guidance on Assigning Values to Uncertain Parameters in Subsurface Contaminant Fate and Transport Modelling. NC/99/38/3, National Groundwater & Contamination Land Centre, Environmental Agency, Olton.

Minkkinen, P., 1987. Evaluation of the Fundamental Sampling Error in the Sampling of Particulate Solids. Analytica Chimica Acta, 196: 237-245.

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.

National Research Council, 1997. Valuing Ground Water. Economic Concepts and Approaches. National Academy Oress, Washington DC, 189 pp.

Naturvårdsverket, 1997a. Bakgrundshalter i mark. Halter av vissa metaller och organiska ämnen i jord i tätort och på landsbygd. Efterbehandling och sanering. Rapport 4640 (in Swedish), Stockholm.

Naturvårdsverket, 1997b. Generella riktvärden för förorenad mark. Beräkningsprinciper och vägledning för tillämpning. Rapport 4639 (in Swedish), Stockholm.

Naturvårdsverket, 1998. Requirements for site remediation. Guidelines for practical achievement of acceptable residual concentrations and quantities - methods and quality aspects. Report 4808, Stockholm.

Naturvårdsverket, 1999. Metodik för inventering av förorenade områden. Bedömningsgrunder för miljökvalitet. Vägledning för insamling av underlagsdata. Rapport 4918 (in Swedish), Stockholm.

Norrman, J., 2000a. Decision Analysis under Risk and Uncertainty at Contaminated Sites. A literature review. Publ. B 485, Department of Geology, Chalmers University of Technology, Göteborg.

Norrman, J., 2000b. Risk-based Decision Analysis for the Selection of Remediation Strategy at a Landfill. In: O. Sililo (Editor), Groundwater: Past Achievements and Future Challenges. A.A. Balkema, Capetown, South Africa, pp. 797-802.

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.

Rai, S.N., Krewski, D. and Bartlett, S., 1996. A General Framework for the Analysis of Uncertainty and Variability in Risk Assessment. Human and Ecological Risk Assessment, 2(4): 972-989.

Ramsey, C.A. and Suggs, J., 2001. Improving Laboratory Performance Through Scientific Subsampling Techniques. Environmental Testing & Analysis(March/April 2001).

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.

Rautman, C.A., McGraw, M.A., Istok, J.D., Sigda, J.M. and Kaplan, P.G., 1994. Probabilistic comparison of alternative characterization technologies at the Fernald Uranium-in-Soils Integrated Demonstration Project, Waste Management -94, Tucson, Arizona, pp. 2117-2124.

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., 2002. Decision Analysis for Managing Contamination Risks from Petroleum Transport on Roads. In: E. Bocanegra, D. Martinez and H. Massone (Editors), Groundwater and Human Development. IAH & ALHSUD, Mar del Plata, Argentina, pp. 1784-1793.

Rosén, L. and Gustafson, G., 1996. A Bayesian Markov Geostatistical Model for Estimation of Hydrogeological Properties. Ground Water, 34(5): 865-875.

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 Tunnelling. Dissertation Thesis, Royal Institute of Technology, Stockholm, 191 pp.

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.

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, 1996. Soil Screening Guidance: User´s Guide. 9355.4-23, Second Edition, Office of Solid Waste and Emergency Response, Washington, D.C.

U.S. EPA, 1997a. Guiding Principles for Monte Carlo Analysis. EPA/630/R-97/001, Risk Assessment Forum, Washington, D.C.

U.S. EPA, 1997b. 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.

Wagner, B.J., 1999. Evaluating Data Worth for Ground-Water Management under Uncertainty. Journal of Water Resources Planning and Management, 125(5): 281-288.

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.

Vose, D., 1996. Quantitative Risk Analysis. A Guide to Monte Carlo Simulation Modelling. John Wiley & Sons, Chichester, 328 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 1

Paper:

Risk Management of Groundwater Supplies Exposed to Railroad Transport of Dangerous Goods

GROUNDWATER AND HUMAN DEVELOPMENT
Bocanegra, E - Martínez, D - Massone, H (Eds.) 2002 - ISBN 987-544-063-9

RISK MANAGEMENT OF GROUNDWATER SUPPLIES EXPOSED TO RAILROAD TRANSPORT OF DANGEROUS GOODS

Pär-Erik Back
Department of Environmental Technology, Swedish Geotechnical Institute, SE-581 93 Linköping
Department of Geology, Chalmers University of Technology, SE-412 96 Göteborg, Sweden
E-mail: par-erik.back@swedgeo.se

Abstract
The Swedish National Rail Administration has a responsibility to protect public water supplies from contamination due to potential accidents with dangerous goods on railroads but protective actions can be quite costly. A methodology for risk management of water supplies in the vicinity of roads has been developed by the Swedish National Road Administration but several conditions differ between railroads and roads. In this paper, an extended framework for railroads is presented, consisting of six parts: 1) a decision model, 2) a system for classification of dangerous goods, 3) a hydrogeological conceptual model, 4) an accident and spill probability model, 5) a number of hydrogeological probability models, and 6) a consequence model. Risk-cost-benefit decision analysis is used to identify cost-efficient protective actions. The risk for a water supply is expressed as an annual expected cost. Consequences are expressed in monetary terms and arise when a failure criterion is met. Analytical contaminant transport models are used for estimation of failure probabilities and uncertainty is handled by stochastic simulation. The methodology can be used to 1) structure complex problems, 2) to identify the most cost-efficient protective measure among a set of alternatives, or 3) for risk communication purposes.

Keywords: dangerous goods, railroad, risk management, groundwater protection

Introduction

The Swedish National Rail Administration (SNRA) administrates almost 12,000 km of railroads in Sweden. It is estimated that well over 300 public water supplies are situated along this network of railroads, the private water supplies not counted (Löwegren, SNRA, personal communication, 2002). According to Swedish environmental legislation, the Environmental Code, SNRA has a responsibility to protect groundwater resources from contamination caused by the railroad system. This responsibility includes contamination during the construction phase of the railroad, as well as during operation and maintenance.

The focus of this paper is on public groundwater supplies. We define "water supply" as the abstraction wells in the aquifer supplying the water. In Sweden, there is a system with well head protection zoning around the water supply, with the inner well head protection area corresponding to a advection transport time of less than 60 to 100 days. Restrictions apply in the well head protection area in order to avoid contamination of the water supply.

About ten of the public water supplies have been protected by the installation of geo membranes along the railroad in the well head protection area. Such measures protect the water supply in the case of an accident with dangerous goods but the economic investment is high. These investments are principally based on demands made by the environmental authorities, with little or no consideration of the cost-efficiency.

The Environmental Code states that protective measures should be carried out if they cannot be considered unreasonable. It is also stated that attention should be paid to the benefits of the protective measures in relation to their costs. Estimation of cost-efficiency requires a tool to compare the cost with the resulting reduction in risk for the water supply. In this way, the most cost-efficient protective measures can be selected for a particular site, and unreasonable measures avoided.

A methodology for risk management of water supplies exposed to petroleum transport on roads has been developed by the Swedish National Road Administration (1998) and is described by Rosén (2002). In this methodology a Risk-Cost-Benefit (RCB) approach is used for risk and decision analysis. The framework for roads has been used for selection of alternative road stretches and protective measures along roads (Eklund and Rosén, 2000). It has also been used in risk management for more than 30 water supplies along roads, in various consulting reports. Rosén (1998) describes a similar framework for analysis of the pollution risk from deicing of roads. These risk management methodologies for roads are based on the review paper on

hydrogeological RCB decision analysis by Freeze et al. (1990).

Several conditions differ between railroads and roads, e.g. the probability of accidents occurring is much lower on railroads and the dangerous goods being transported are more diverse. Therefore, the methodology for roads has been developed into a risk-management framework comprising railroad-specific issues like accident statistics and transport models accounting for the physical properties of different liquids (Back and Rosén, 2001). The main objectives of this paper are (1) to describe the railroad-specific aspects of the methodology, and (2) to discuss possibilities and limitations of the methodology.

The Risk Management Framework

A complete risk management framework for railroads and water supplies should encompass at least four risk objects: 1) accident with transport of dangerous goods, 2) contamination during construction work, 3) contamination during maintenance work, and 4) diffuse contamination from railroad installations and railroad traffic during operation. This paper considers the first risk object, i.e. the risk associated with accidental spills of dangerous goods on the railroad.

The framework was developed to handle contamination risks for public groundwater supplies, but groundwater resources in general and surface water resources can also be handled, with slight modifications. The methodology can be used to 1) structure complex problems, 2) to identify the most cost-efficient protective measure among a set of alternatives, or 3) for risk communication purposes.

The risk management framework can be divided into: 1) a decision model, 2) a system for classification of dangerous goods, 3) a hydrogeological conceptual model, 4) an accident and spill probability model, 5) a number of hydrogeological probability models, and 6) a consequence model. Figure 1 illustrates the interaction between the different parts of the framework. Probability estimations are performed with the accident and spill probability model and the hydrogeological probability models, whereas the consequences are estimated in monetary terms with the consequence model. The consequences of an accident are assumed to be independent of the released substance, which of course is a simplification of reality. However, the framework can easily be extended to take this substance-dependence into account, as indicated by the dashed arrow in Figure 1.

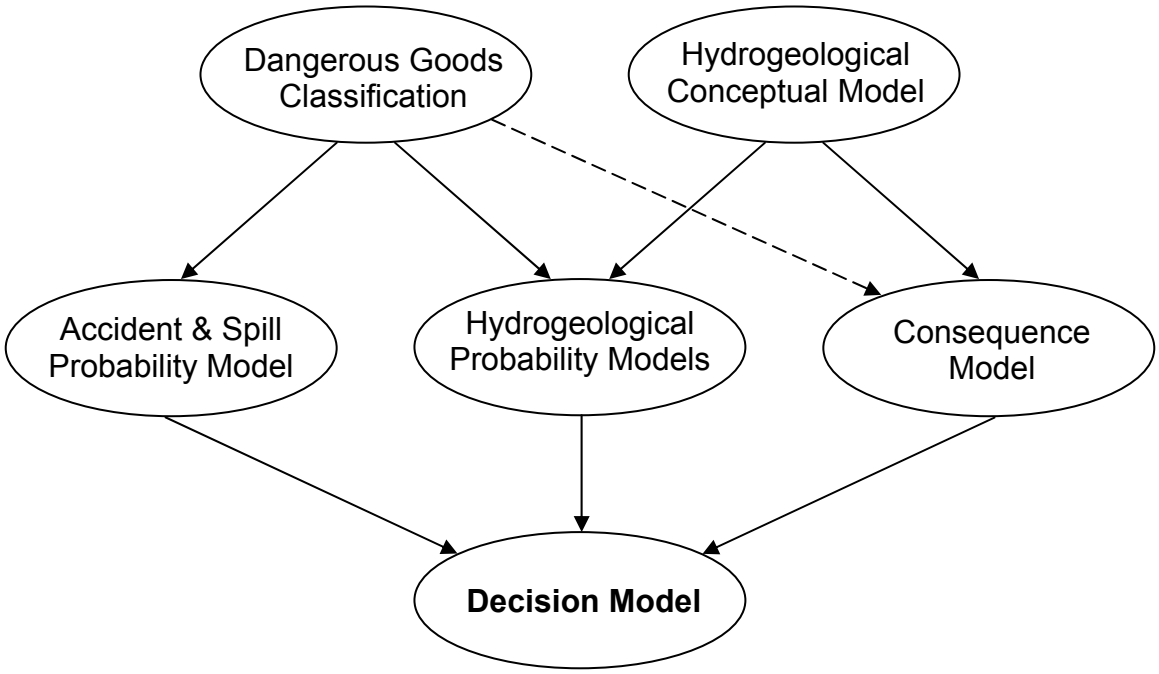


Figure 1. The framework for risk management of groundwater supplies in the vicinity of railroads.

Decision model

In the decision model, the risk is quantified as an *expected annual cost* (risk cost) by multiplying the probability of an event and the consequence of that event. The consequence is expressed as a monetary cost. Figure 2 illustrates an event tree for accidents with dangerous goods on railroads. The probabilities in Figure 2 are defined as:

- $P_{r,\,j}$ = the probability of an accident leading to spill of a liquid *j*.
- P_i = the probability of infiltration of the liquid.
- $P_{v,\,j}$ = the probability of unsuccessful remediation in the vadose zone with respect to the released liquid *j*. Unsuccessful remediation is defined by a failure criterion, e.g. a specified transport time to the groundwater table.
- P_d = the probability of unfavorable groundwater flow direction from the spill, i.e. flow towards the water supply.
- $P_{s,\,j}$ = the probability of unsuccessful remediation in the saturated zone with respect to the released liquid *j*. Unsuccessful remediation is defined by a failure criterion, e.g. a specified transport time to the water supply.

The consequence costs in Figure 2 are defined as:

- C_g = the cost for remediation at the ground surface.
- C_v = the cost for remediation in the vadose zone, including the ground surface.
- C_s = the cost for remediation in the saturated zone.
- C_e = the loss of extractive values for the water resource (see *Consequence model* section).
- C_i = the loss of *in situ* values for the water resource (see *Consequence model* section).

Since a number of different substances are handled in the framework, each substance will have a unique set of probabilities in Figure 2, i.e. the risk cost is estimated separately for each substances *j*. The total annual risk cost *R* for the water supply is equal to the sum of risk costs for all substances:

$$R = \sum_{j=1}^{N} \left(P_{r,j} \left(C_g + P_i \left(C_{v,j} - C_g + P_{v,j} \left(C_s + P_d P_{s,j} \left(C_e + C_i \right) \right) \right) \right) \right) \quad (1)$$

where *N* is the number of substances included in the analysis.

In the decision model, RCB decision analysis is applied in order to identify the most cost-efficient protection strategy among a set of alternative actions. This is performed by comparing the risk reduction for a protective action to its investment cost. The *risk-cost minimization objective function* is defined as the sum of risks and costs over the time horizon, e.g. the life span of the railroad. The most cost-efficient alternative is identified as the one with lowest value of the objective function.

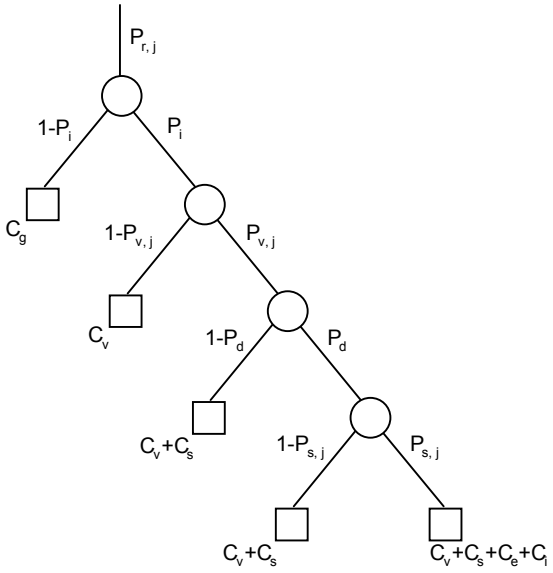


Figure 2. Event tree for the risk analysis. Circles symbolize probability nodes (chance nodes) and squares indicate consequence nodes (terminal nodes).

The fundamentals of hydrogeological RCB decision analysis are described in detail by Freeze et al. (1990). The decision model in this framework for railroads is similar to the one for roads, as described by Rosén (2002).

Dangerous goods classification

A wide range of chemical substances is transported on railroads in the vicinity of water supplies, e.g. petroleum products, acids, ammonia, phenol and several other toxic substances. The purpose of the dangerous goods classification model is to limit the number of substances in the analysis to a reasonable number, without distorting the risk estimate. All substances are placed in one of two classes: A) substances that pose a hazard to groundwater, and B) substances that do not pose a hazard to groundwater. Only liquids are placed in class A, whereas condensed liquefied gases and solid materials belong to class B and are disregarded. Water-soluble condensed liquefied gases like ammonia is considered a special case that can pose a hazard to water supplies under unfortunate circumstances, e.g. if the gas is released during rainy weather.

Substances in class A are classified as either aqueous phase liquids (APLs) or non-aqueous phase liquids (NAPLs) because of their different behavior in the subsurface. All substances with solubility lower than 5 % are roughly classified as NAPLs. Different substances with similar properties are handled as a homogeneous group to facilitate the analysis. Also, a flexible "10 percent rule" is applied in such a way that only substances, or groups of substances, contributing to more than 10 percent of the total load of dangerous goods in class A are considered in the analysis. The total load of the selected substances are adjusted upwards (upscaled) to account for the neglected substances, so that the total load of goods is taken into account.

Hydrogeological conceptual model

The hydrogeological conceptual model constitutes the hydrogeological base for the risk analysis. In the framework, a set of hydrogeological type settings is presented, representing typical hydrogeological conditions where groundwater is abstracted. Aller et al. (1987) define a hydrogeological type setting as a composite description of all the major geologic and hydrologic factors, which affect and control groundwater movement into, through, and out of an area. The hydrogeological setting description allows inference of information from well-known areas to areas with limited information (Eklund and Rosén, 2000).

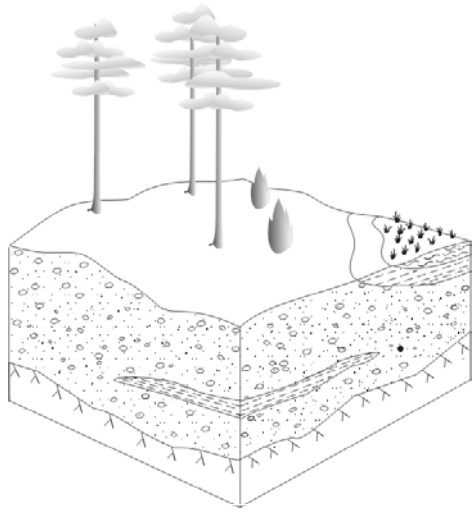


Figure 3. Unconfined aquifers in sub-aquatic glacio-fluvial deposits; an example of a hydrogeological type setting. The illustration is taken from Eklund (2002).

There are 18 predefined hydrogeological type settings in the current framework for railroads, identical to the ones in the framework for roads (Rosén, 2002). A conceptual description and an illustration (Figure 3) are presented for each type setting. Most likely values, uncertainty intervals, and probability density function types (PDF-types) are listed for several model parameters for each type setting (Table 1).

Table 1. Model parameters for aquifers in sub-aquatic glacio-fluvial deposits. K_v and K_h are the vertical and horizontal hydraulic conductivity respectively, n_e is the effective porosity, i is the hydraulic gradient, D_v is the depth to the groundwater table, and R_c is the retention capacity as defined by CONCAWE (1981).

Parameter	Most likely value	Interval	PDF-type
K_v [m/s]	10^{-4}	10^{-6} - 10^{-2}	Lognormal
K_h [m/s]	10^{-3}	10^{-5} - 10^{-1}	Lognormal
n_e [m^3/m^3]	0.25	0.15 - 0.35	Normal
i [m/m]	0.005	0.0005 - 0.02	Triangular
D_v [m]	10	1 - 30	Triangular
R_c [m^3/m^3]	0.015	0.006-0.03	Triangular

The section of the railroad passing through the well head protection area may encounter several geological formations. In such circumstances, the railroad should be divided into subsections, each subsection representing only one hydrogeological type setting. The risk is determined individually for each subsection and the total risk for the water supply is estimated as the sum of all subsections.

Accident and spill probability model

An accident and spill probability model is used to estimate the probability of a railroad accident, leading to a spill of dangerous goods within the well head protection area. The model is based on accident statistics on Swedish railroads, collected from different sources and representing a time period between approximately 1981 and 1999 (Fredén, 2001). The output from the model is an estimate of the probability $P_{r,j}$ in Figure 2.

Two kinds of railroad accidents are considered in the model: 1) derailment of freight trains, and 2) collisions between freight trains and heavy road vehicles on railroad crossings. A third type of accident, collision between a freight train and an opposing train on a double track, can also be taken into account but the probability of such an event is so low that it can be neglected.

Hydrogeological probability models

Failure criteria and compliance boundaries

Contaminant transport models are used to estimate the probability of unsuccessful remediation of the vadose zone and the saturated zone respectively. The remediation is considered unsuccessful if a failure criterion is met at a compliance boundary. At least four types of failure criteria can be considered: 1) concentration criterion, 2) mass flux criterion, 3) transport time criterion, and 4) transport distance criterion. The compliance boundary can be of different types, for example 1) the groundwater table, 2) a specified distance downstream from the spill, or 3) the water supply itself. The framework is flexible enough to handle different failure criteria and compliance boundaries, but recommendations are listed in Table 2, for the vadose zone and the saturated zone respectively.

Table 2. Recommended failure criteria and compliance boundaries.

	Failure criterion	Compliance boundary
Vadose zone	Transport time	Groundwater table
Saturated zone	Transport time	1) Distance from railroad 2) Groundwater supply

A transport time criterion is most suitable for practical applications, e.g. remediation is considered a failure if the transport time to the compliance boundary is shorter than the time criterion. It is believed that it will be difficult to get acceptance for a maximum allowable concentration criterion or a mass flux criterion at a water supply, because these criteria imply a certain acceptable contamination of the water supply. Such criteria are more realistic to use in situations where risks to ecosystems are of concern, not when human water consumption is addressed. The transport distance criterion listed above is of minor interest.

Two complexity levels are defined for the choice of transport models, with the more complex level requiring more information. For the less complex level, all substances are assumed to behave like water on the ground and in the subsurface. In this paper, only the more complex level is addressed, taking the properties of different liquids into account.

Multiphase flow is considered in the vadose zone and contaminant transport with dispersion and sorption in the saturated zone. Parameter uncertainty is handled by assigning PDFs to uncertain variables and the uncertainty is propagated through the analytical models by stochastic simulation (Monte Carlo).

Infiltration of liquid

The probability of ground infiltration of the spilled liquid is denoted P_i (Figure 2) and is estimated subjectively based on available information. If low-permeable materials like clay or geo membranes protected the ground surface, this can be taken into account by assigning P_i a value less than 1. In such a case P_i can be estimated as the probability that a damaged tank car would end up on a low-permeable area after an accident. With no protection from infiltration and high-permeable soil, P_i will be equal to 1.

Unsuccessful remediation of the vadose zone

The probability of unsuccessful remediation of the vadose zone ($P_{v,j}$ in Figure 2) is estimated as:

$$P_{v,j} = P_{g,j} \cdot P_{T,j} \tag{2}$$

where $P_{g,j}$ is the probability that the released volume of a liquid is so large that it will reach the groundwater table, and $P_{T,j}$ is the probability that the transport time to the groundwater is shorter than the transport time criterion, i.e. there is not enough time available to remediate the vadose zone. Subscript j indicates that the probability or parameter of interest is substance dependent. Such probabilities must be estimated for each substance in accordance with (1).

It is assumed that $P_{g,j} = 1$ for APLs because these liquids mix with percolating water. This assumption may lead to an overestimation of the risk because it implies that even small spills of APLs are mobile, and not retained, in the vadose zone. The validity of this assumption depends on the water content in the vadose zone. For NAPLs, an expression of the probability $P_{g,j}$ can be derived from CONCAWE (1981):

$$P_{g,j} = P\left(\frac{V_u}{A_{u,j} \cdot D_v \cdot R_{c,j} \cdot c_{f,j}} \geq 1 \right) \tag{3}$$

where V_u is the released volume of liquid [m^3], $A_{u,j}$ is the area of the pool of liquid on the ground surface [m^2], D_v is the depth to the groundwater table [m], $R_{c,j}$ is the retention capacity for the liquid [m^3/m^3], and $c_{f,j}$ is an approximate correction factor for viscosity [dim.less]. If the depth of the pool is assumed to be constant, which is a more or less reasonable assumption depending on the roughness of the ground surface, the following solution for the

pool area can be derived from a mass flux differential equation (Back and Rosén, 2001):

$$A_{u,j} = \frac{q_u \cdot \nu_j}{K_m \cdot i \cdot \nu_w}\left[1 - \exp\left(-\frac{K_m \cdot i \cdot \nu_w \cdot V_u}{\nu_j \cdot h \cdot q_u}\right)\right] \quad (4)$$

where q_u is the constant outflow rate of the liquid from the damaged tank car [m^3/s], K_m is the hydraulic conductivity of the superficial soil layer [m/s], i is the hydraulic gradient [m/m], h is the constant depth of the pool of liquid [m], ν_j is the kinematic viscosity for the liquid [m^2/s], and ν_w is the kinematic viscosity for water [m^2/s]. Before applying (4), it is necessary to consider the site-specific conditions like ditches and the track structure.

The probability $P_{T,j}$ can be estimated by a simple two-phase infiltration model for steady state conditions, derived from Mull (1971):

$$P_{T,j} = P\left(\frac{D_v \cdot n \cdot S_j \cdot \nu_j}{K_v \cdot i \cdot k_{r,j} \cdot \nu_w} \le T_v\right) \quad (5)$$

where T_v is the transport time criterion [s], n is the porosity [m^3/m^3], S_j is the saturation of the liquid in the pore space [m^3/m^3], K_v is the hydraulic conductivity in the vertical direction [m/s], and $k_{r,j}$ is the relative permeability of the liquid [dim.less]. If the steady state assumption cannot be considered justified a more realistic and complex model has to be applied, e.g. the Green and Ampt NAPL infiltration model described by Charbeneau (2000).

Unfavorable groundwater flow direction

Depending on the hydrogeological conditions at the site, it will be more or less certain that the direction of groundwater flow from the spill is towards the abstraction wells. From a risk point of view, this is an unfavorable flow direction. This probability (P_d in Figure 2) is estimated from hydrogeologic expertise based on available information. Numerical groundwater models can be used as tools for the estimation.

Unsuccessful remediation of the saturated zone

When a spill of NAPL has occurred, the liquid will be retained in the pore space of the vadose zone and in the saturated zone. This NAPL will act as a contaminant source for the groundwater for a long time. If remediation cannot be performed fast enough, the NAPL will be a threat to the water supply since water-soluble species in the NAPL will dissolve and be transported away. The same principles apply to APLs but to a more limited degree. APLs in the pore spaces will dissolve relatively fast and be washed away by the groundwater.

The probability of unsuccessful remediation of the saturated zone ($P_{s,j}$ in Figure 2) is estimated by a 1D analytical solution of the advection-dispersion solute transport equation. Dispersion is taken into account in order not to underestimate the risk (Baca, 1999). Sorption is taken into account for organic substances because neglecting it would lead to overly conservative risk estimates for certain substances. Biodegradation is not considered because of scarcity of site-specific information regarding the controlling factors.

The choice between a pulse source model and a continuous source model depends on the expected scenario at the point of the spill, i.e. how fast complete remediation can take place. As mentioned, residual NAPL in the saturated zone is likely to act as a contaminant source for a long time. Suitable equations abound in the literature and can be found in for example van Genuchten and Alvares (1982), Domenico and Schwartz (1990), and Fetter (1999). As an example, the retardation equation with a continuous source can be used for accidental spills of NAPLs (Domenico and Schwartz, 1990):

$$C_j = \frac{C_{0,j}}{2}\left[erfc\left(\frac{R_{f,j} \cdot L - v \cdot T_L}{2 \cdot \sqrt{\alpha_L \cdot v \cdot T_L \cdot R_{f,j}}}\right)\right] \quad (6)$$

where C_j is the concentration of the contaminant at the compliance boundary [kg/kg], $C_{0,j}$ is the concentration in the groundwater under the spill [kg/kg], $R_{f,j}$ is the retardation factor for the contaminant [dim.less], L is the distance to the compliance boundary [m], α_L is the longitudinal dispersivity [m], v is the velocity of the water [m/s], and T_L [s] is the transport time to the compliance boundary. Since a transport time criterion is used, it is the arrival time of the plume front that is of interest (Baca, 1999). Therefore, (6) is solved for T_L when C_j deviates slightly from zero (arrival of the plume front).

It is important to note that for NAPLs like oil or gasoline, which are mixtures of different substances, the transport calculations in the saturated zone should be made for one substance in the mixture. This substance should be selected with regard to its mobility and how hazardous it is. The probability of unsuccessful remediation is given by:

$$P_{s,j} = P\left(T_L \le T_s\right) \quad (7)$$

where T_s [s] is the transport time criterion defining failure.

Consequence model

All consequences in Figure 2 are quantified as costs. Three principally different types of costs are considered: 1) costs for remedial actions, 2) loss of extractive values of the water resource (C_e), and 3) loss of *in situ* values of the water resource (C_i). The remediation costs consist of costs for remediation of the ground surface (C_g), the vadose zone (C_v), and the saturated zone (C_s), including costs for excavation and restoration of the track structure. Secondary costs caused by standstill on the railroad during remedial actions should also be included.

The extractive values are values related to the use and exploitation of the water resource. The loss of extractive values is estimated by quantifying 1) costs of arranging temporary water supply, 2) loss of industrial production due to water shortage, 3) cost of labor for private consumers arranging water distribution, and 4) costs for establishing a new water supply. These costs must be estimated site-specifically. Swedish National Road Administration (1998) and Rosén (2002) give more detailed presentations of these costs.

The *in situ* values are connected to the presence of groundwater in the aquifer and such values are often difficult to estimate. Rosén (2002) discusses these values and their influence on a decision. NRC (1993) provides a detailed description of *in situ* values, e.g. loss of ecological values.

Applications of the methodology

The experience from practical applications of the methodology for railroads has so far been limited. However, the framework has partially been applied to the Bjästatjärn public water supply along the planned Botnia Railroad on the north East Coast of Sweden. The Bjästatjärn groundwater supply is located in an esker of glacio-fluvial deposits of sand and gravel and has three abstraction wells. The capacity of the water supply is about 2,300 m^3/day. The load of dangerous goods liquids on the railroad is assumed to be about 100,000 tons per year.

Probability estimations were performed for two decision alternatives; with and without the installation of protective geo membranes. Five different scenarios regarding consequences where evaluated. No strict RCB decision analysis was performed. Instead, probabilities and costs were presented for each scenario, as a basis for risk communication and decisions of acceptable risk levels and protective actions (J&W, unpublished material, 2002).

Currently, the framework is applied at Högby public water supply in the municipality of Mjölby, where a double track section of 1,700 m is planned through the well head protection area. The capacity of the water supply is over 2,000 m^3/day. Suggested protective measures include the installation of geo membranes. The load of dangerous goods liquids (class A) is estimated to more than 140,000 tons per year, constituting about 60 different substances, with APLs dominating. Four groups of liquid solutions are included in the analysis: cyanide, acids, hydrogen peroxide, and ammonia. Separate analyses are carried out for each group of substances. Three subsections of the railroad are also analyzed separately because of differing hydrogeological conditions.

Failure is defined to occur if the contaminant transport time to a compliance boundary 100 m from the railroad is shorter than a specified time criterion. Consequences are estimated considering costs of remedial actions and loss of extractive values. Preliminary results indicate that the estimated risk is very low (Johansson, SWECO VIAK, personal communication, 2002). Costly protective actions, like the installation of geo membranes, are difficult to justify with a strict RCB decision-analysis approach under these circumstances.

Conclusions and discussion

The framework presented in this paper has relatively simple parts but as a whole it is quite complex in all its details. The team performing the analysis will require competence in several fields, such as hydrogeology, risk and decision analysis, accident and freight statistics, chemistry, economy, and engineering. Applied correctly, it is believed that the framework will be an important tool for cost-efficient management of risks to water supplies. It should be emphasized that there are uncertainties in the methodology, which must be taken into consideration.

Often a distinction is made between uncertainty in model structure and uncertainty in parameter values, although the difference can be rather delicate (Morgan and Henrion, 1990). In the presented framework, parameter uncertainty is managed by PDFs and stochastic simulation. Model uncertainty must be handled in other ways.

Two types of model uncertainty exist; uncertainty in conceptual models and uncertainty in quantitative models (McMahon et al., 2001). Both these uncertainties are important to consider in the presented framework but they are extremely difficult, if

not impossible, to quantify. This is a problem, since uncertainty about model structure generally is more important than parameter uncertainty (Morgan and Henrion, 1990). Furthermore, it has been shown that uncertainty often is underestimated and that bias may be substantial (Hammitt and Schlyakhter, 1999). Therefore, it is wise to apply the framework with some degree of conservatism in order not to underestimate the risk, especially when the problem is conceptualized. However, it is important to note that although some model uncertainties are large this may be of minor importance in the context of decision-making, as long as the uncertainties are not large enough to affect the decision.

It is unavoidable that a framework off this type involves portions of subjectivity, which may lead to slightly different risk estimates between analysts. One example of an important aspect that may affect the result is the choice of failure criteria. In the case of a transport time criterion, an optimistic analyst might believe that remedial actions could be carried out within a few hours, while another person might believe several days or weeks would be needed. One way to handle this is to define the time criterion as a stochastic variable.

An additional example of a conceptual uncertainty will be given. The described framework considers primary effects caused by the released substance, whereas secondary effects are ignored. However, in certain circumstances secondary effects may be the most important, e.g. when a large volume of acid is released. This will lower the pH-value and dissolve metals. In this case, the metals may be the hazard to the water supply, not the acid itself.

Although experience from practical application of the methodology is limited, it is believed that it will result in low risk estimates. This is mainly due to the low probability of accidents with dangerous goods on Swedish railroads. However, the statistical basis for the accident and spill model is rather weak, and the probability estimates therefore uncertain.

A low estimated risk implies low cost-efficiency for costly protective actions. Under such circumstances there must be strong political or legal reasons to motivate these actions. In this context it is important to understand that other considerations than strict economical can be embraced in the RCB decision process, e.g. legal and political aspects. Such factors can be allowed to influence the analysis if desired, for example by including different perspectives on environmental and ecological values in the analysis.

The framework presented in this paper provides a valuable tool for cost-efficient management of water supplies along railroads and will be used by SNRA in the planning stage for future construction and re-investment projects. In future work, the framework will be extended to encompass three other risk sources in addition to transport of dangerous goods, i.e. risks for groundwater supplies associated with 1) the construction of railroads, 2) the maintenance of railroads, and 3) diffuse contamination from railroad installations and railroad traffic.

Acknowledgement

This methodology was developed through the financial support of SNRA. The help provided by the personnel at SNRA, Swedish Geotechnical Institute, Chalmers University of Technology, and a number of Swedish consultant companies, has been invaluable in the development of the methodology. Many thanks to all the persons who contributed with fruitful discussions and constructive comments.

References

Aller, LT, Bennet, T, Lehr, JH, Petty, RJ, and Hackett, G. 1987. *DRASTIC: A Standardized System for Evaluating Groundwater Pollution Potential Using Hydrogeological Settings*. U.S. Environmental Protection Agency/600/2-87/035. Washington D.C.

Baca, E., 1999: On the misuse of the simplest transport model. Technical Commentary. *Ground Water, 37(4)*, 483.

Back PE, and Rosén B. 2001. Riskanalys av områden där järnvägstrafik berör vattentäkter och andra vattenresurser – Metodutveckling [Risk analysis of areas where railroad traffic may affect water supplies and other water resources - Development of method]. Swedish Geotechnical Institute, Varia 513.

Charbeneau, RJ. 2000. *Groundwater Hydrauliscs and Pollutant Transport*. Prentice Hall, Upper Saddle River, NJ.

CONCAWE, 1981. *Revised Inland Oil Spill Clean-up Manual*. Report no. 7/81. Den Haag, Bryssel.

Domenico, PA, and Schwartz, FW. 1990. *Physical and Chemical Hydrogeology*. John Wiley & Sons, New York.

Eklund, HS. 2002. Hydrogeologiska typmiljöer. Verktyg för bedömning av grundvattenkvalitet, identifiering av grundvattenförekomster samt under-

lag för riskhantering längs vägar [Hydrogeological type settings. Tool for assessment of groundwater quality, identification of groundwater resources, and a basis for risk management along roads]. Chalmers University of Technology, Department of Geology. Publ. A101.

Eklund, HS and Rosén L. 2000. Hydrogeological decision analysis for selection of alternative road stretches and designs. **In:** Proceedings of the XXX IAH Congress on Groundwater: Past Achievements and Future Challenges, 923-928.

Fetter, CW. 1999. *Contaminant Hydrogeology*, 2nd edn. Prentice Hall, Upper Saddle River, NJ.

Freden, S. 2001. Modell för skattning av sannolikheten för järnvägsolyckor som drabbar omgivningen [Model for estimation of the probability of railway accidents affecting the surroundings]. Swedish National Rail Administration, Environmental Section, report 2001:5.

Freeze, RA, Massman, J, Smith, JL, Sperling, T, and James, BR. 1990. Hydrogeological decision analysis, 1. A framework. *Ground Water, 28(5)*, 738-765.

Hammitt, JK, and Shlyakhter, AI. 1999. The expected value of information and the probability of surprise. *Risk Analysis, 19(1)*, 135-152.

McMahon, A, Heathcote, J, Carey, M, Erskine, A, and Barker, J. 2001. Guidance on assigning values to uncertain parameters in subsurface contaminant fate and transport modelling. Environment Agency (UK), National Groundwater & Contamination Land Centre, report NC/99/38/3.

Morgan, MG, and Henrion, M. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press.

Mull, R. 1971. The migration of oil-products in the subsoil with regard to ground-water-pollution by oil. **In:** Jenkins, 1971. HA-7(a), 1-8.

NRC (National Research Council). 1997. *Valuing Ground Water. Economic Concepts and Approaches*. National Academy Press, Washington DC.

Rosén, L. 1998. Risk Assessment and Pollution Prevention. Keynote Lecture. **In:** Proceedings to the International Symposium on Deicing and Dustbinding - Risk to Aquifers, D&D98. Nordic Hydrological Programme. NHP Report No 43, 93-110.

Rosén, L. 2002. Decision analysis for managing contamination risks from petroleum transports on roads. In this volume.

Swedish National Road Administration. 1998. Förorening av vattentäkt vid vägtrafikolycka. Hantering av risker med petroleumutsläpp [Contamination of water supplies after road accidents. Management of risks associated with petroleum spills]. VV Publ 98:064.

van Genuchten, MT, and Alvares, WJ. 1982. *Analytical solutions of the one-dimensional convective-dispersive solute transport equation*. US Department of Agriculture. Agricultural Research Service, Technical Bulletin 1661, Washington DC.

Appendix 2

MathCad Application for Estimation of Data Worth
in Sampling Programs

MathCad Application for Estimation of Data Worth in Sampling Programs

Prior information expressed as a PDF

Input and result sheet on last page.

The minimum value of the mean concentration is *a* and the maximum is *b*.
The most likely value, *m*, is equal to the mode for the triangular, log-normal, and normalised normal distributions.

Percentile below b: $P_{95} := 0.95$ P(x<b)=0.95 for normal and log-normal distributions, alt. 1

$P_{99} := 0.99$ P(x<b)=0.99 for normal and log-normal distributions, alt. 2

$AL_2 := AL$ AL is the level for D^+ due to sample uncertainty.

$Action\_Level := 0..2$ AL_2 is the level for C^+ on the prior distribution.

Uniform distribution:

$$f_{uni}(\mu) := \begin{cases} 0 & \text{if } \mu < a \vee \mu > b \\ \left(\frac{1}{b-a}\right) & \text{otherwise} \end{cases}$$

$$\sigma_{uni} := \sqrt{\frac{(b-a)^2}{12}} \qquad \sigma_{uni} = 274.2$$

$$\mu_{uni} := \frac{a+b}{2} \qquad \mu_{uni} = 525$$

Triangular distribution:

$$f_{tri}(\mu) := \begin{cases} \left[\frac{2(\mu-a)}{(b-a)\cdot(m-a)}\right] & \text{if } \mu \le m \wedge \mu \ge a \\ \left[\frac{2(b-\mu)}{(b-a)\cdot(b-m)}\right] & \text{if } \mu > m \wedge \mu \le b \\ 0 & \text{otherwise} \end{cases}$$

$$\sigma_{tri} := \sqrt{\frac{\left(a^2+b^2+m^2-a\cdot m-a\cdot b-b\cdot m\right)}{18}} \qquad \sigma_{tri} = 208.5$$

$$\mu_{tri} := \frac{a+b+m}{3} \qquad \mu_{tri} = 416.7$$

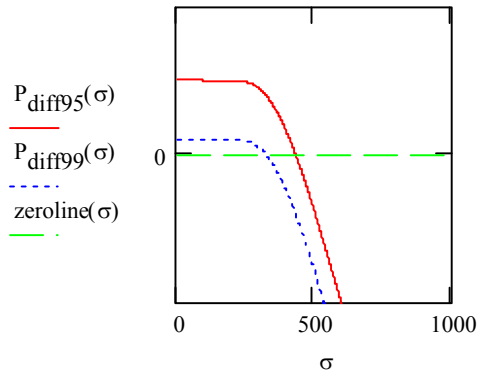
Normalised normal distribution:

$zeroline(\sigma) := 0$

$$P_{diff95}(\sigma) := \frac{pnorm(b,m,\sigma) - pnorm(a,m,\sigma)}{1 - pnorm(a,m,\sigma)} - P_{95}$$

$$P_{diff99}(\sigma) := \frac{pnorm(b,m,\sigma) - pnorm(a,m,\sigma)}{1 - pnorm(a,m,\sigma)} - P_{99}$$

The solutions are given by $P_{diff}(\sigma) = 0$



$\sigma := 300$ Initial trail only. Change initial value of σ if MathCad can't converge to a solution.

$\sigma_{N95} := root\left(P_{diff95}(\sigma), \sigma\right)$ $\sigma_{N95} = 431.3$ $\sigma_{N99} := root\left(P_{diff99}(\sigma), \sigma\right)$ $\sigma_{N99} = 324.1$

Normalised normal distribution:

$$f_{norN95}(\mu) := \begin{cases} 0 & \text{if } \mu < a \\ \dfrac{dnorm\left(\mu, m, \sigma_{N95}\right)}{1 - pnorm\left(a, m, \sigma_{N95}\right)} & \text{otherwise} \end{cases}$$

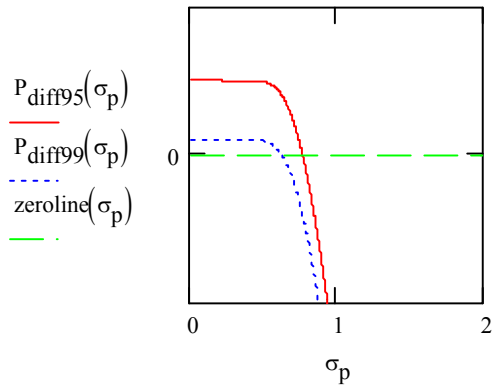
$$f_{norN99}(\mu) := \begin{cases} 0 & \text{if } \mu < a \\ \dfrac{dnorm\left(\mu, m, \sigma_{N99}\right)}{1 - pnorm\left(a, m, \sigma_{N99}\right)} & \text{otherwise} \end{cases}$$

$\mu_{N95} := \int_{-\infty}^{\infty} \mu \cdot f_{norN95}(\mu)\, d\mu$ $\mu_{N95} = 454.6$ $\mu_{N99} := \int_{-\infty}^{\infty} \mu \cdot f_{norN99}(\mu)\, d\mu$ $\mu_{N99} = 371.3$

Lognormal distribution:

$P_{diff95}(\sigma_p) := plnorm\left[b - a, ln\left[(m - a)e^{\sigma_p^2}\right], \sigma_p\right] - P_{95}$ $P_{diff99}(\sigma_p) := plnorm\left[b - a, ln\left[(m - a)e^{\sigma_p^2}\right], \sigma_p\right] - P_{99}$

The solutions are given by $P_{diff}(\sigma_p) = 0$



$\sigma_p := .7$ Initial trail only. Change initial value of σ_p if MathCad can't converge to a solution.

$\sigma_{p95} := \text{root}\left(P_{diff95}(\sigma_p), \sigma_p\right)$ $\sigma_{p95} = 0.766$

$\mu_{p95} := \ln\left[(m-a)e^{\sigma_{p95}^2}\right]$ $\mu_{p95} = 5.597$

$\sigma_{L95} := \sqrt{\left(e^{2\cdot\mu_{p95}+\sigma_{p95}^2}\right)\cdot\left(e^{\sigma_{p95}^2}-1\right)}$

$\mu_{L95} := e^{\mu_{p95}+\frac{\sigma_{p95}^2}{2}}$

$\sigma_{L95} = 322.8$ $\mu_{L95} = 361.5$

$\sigma_{p99} := \text{root}\left(P_{diff99}(\sigma_p), \sigma_p\right)$ $\sigma_{p99} = 0.625$

$\mu_{p99} := \ln\left[(m-a)e^{\sigma_{p99}^2}\right]$ $\mu_{p99} = 5.402$

$\sigma_{L99} := \sqrt{\left(e^{2\cdot\mu_{p99}+\sigma_{p99}^2}\right)\cdot\left(e^{\sigma_{p99}^2}-1\right)}$

$\mu_{L99} := e^{\mu_{p99}+\frac{\sigma_{p99}^2}{2}}$

$\sigma_{L99} = 186.6$ $\mu_{L99} = 269.7$

Lognormal distribution:

$$f_{log95}(\mu) := \begin{cases} 0 & \text{if } \mu < a \\ \text{dlnorm}(\mu - a, \mu_{p95}, \sigma_{p95}) & \text{otherwise} \end{cases}$$

$$f_{log99}(\mu) := \begin{cases} 0 & \text{if } \mu < a \\ \text{dlnorm}(\mu - a, \mu_{p99}, \sigma_{p99}) & \text{otherwise} \end{cases}$$

Estimation of prior probabilities P(state) = P(C)

$$P_{pri\_Cp} := \begin{pmatrix} \int_{AL_2}^{b} f_{uni}(\mu)\,d\mu \\ \int_{AL_2}^{\infty} f_{tri}(\mu)\,d\mu \\ \int_{AL_2}^{\infty} f_{norN95}(\mu)\,d\mu \\ \int_{AL_2}^{\infty} f_{norN99}(\mu)\,d\mu \\ \int_{AL_2}^{\infty} f_{log95}(\mu)\,d\mu \\ \int_{AL_2}^{\infty} f_{log99}(\mu)\,d\mu \end{pmatrix}$$

$$f_{pri}(\mu) := \begin{cases} f_{uni}(\mu) & \text{if PDF} = 1 \\ f_{tri}(\mu) & \text{if PDF} = 2 \\ f_{norN95}(\mu) & \text{if PDF} = 3 \\ f_{norN99}(\mu) & \text{if PDF} = 4 \\ f_{log95}(\mu) & \text{if PDF} = 5 \\ f_{log99}(\mu) & \text{otherwise} \end{cases}$$

$P(\mu > AL_2) = P(C^+) =$ $P_{Cp} := P_{pri\_Cp_{PDF}}$ $P(\mu < AL_2) = P(C^-) =$ $P_{Cm} := 1 - P_{Cp}$

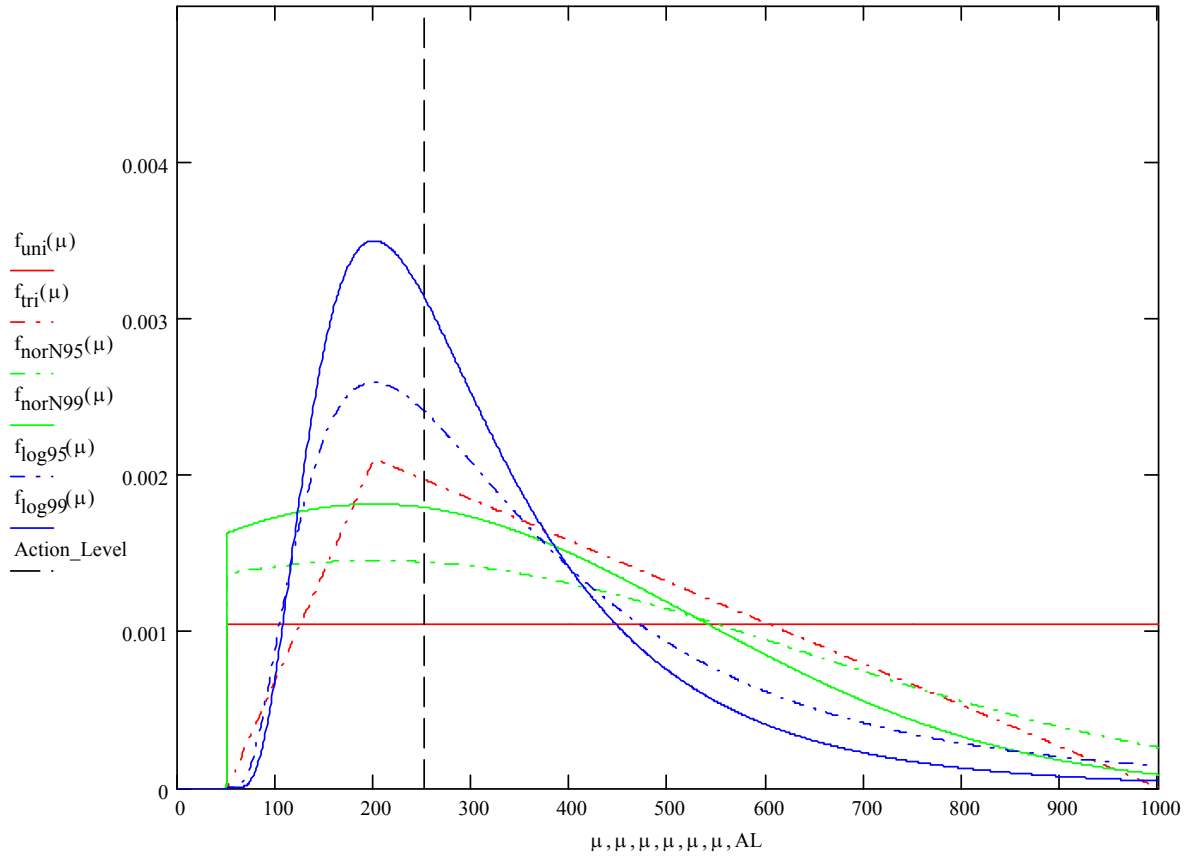
$$P_{prior} := \begin{pmatrix} P_{Cp} \\ P_{Cm} \end{pmatrix}$$ $P_{Cp} = 0.789$ $P_{Cm} = 0.211$

These are the **prior probabilities of state**, i.e. the probabilities for the true mean concentration to exceed or fall below the action level.

Prior distributions

Prior information:

Lowest value:	a = 50
Most likely value:	m = 200
Highest value:	b = 1000



	Mode	Mean	Stdv.	Prob. C+
1. Uniform PDF:	Not unique	$\mu_{uni} = 525$	$\sigma_{uni} = 274.2$	$P_{pri\_Cp_1} = 0.789$
2. Triangular PDF:	m = 200	$\mu_{tri} = 416.7$	$\sigma_{tri} = 208.5$	$P_{pri\_Cp_2} = 0.74$
3. Normalised normal PDF, 95%:	m = 200	$\mu_{N95} = 454.6$	$\sigma_{N95} = 431.3$	$P_{pri\_Cp_3} = 0.714$
4. Normalised normal PDF, 99%:	m = 200	$\mu_{N99} = 371.3$	$\sigma_{N99} = 324.1$	$P_{pri\_Cp_4} = 0.647$
5. Lognormal PDF, 95%:	m = 200	$\mu_{L95} = 361.5$	$\sigma_{L95} = 322.8$	$P_{pri\_Cp_5} = 0.652$
6. Lognormal PDF, 99%:	m = 200	$\mu_{L99} = 269.7$	$\sigma_{L99} = 186.6$	$P_{pri\_Cp_6} = 0.566$

Estimation of probabilities P(sample|state) = P(D|C)

Including sample uncertainty

x is the measured sample concentration and μ is the true mean concentration.
We assume that errors are normally distributed.
μ corrected for bias, is expressed as $\mu_k(\mu) := \mu + \Delta + k\cdot\mu$

$P(x>AL \mid \mu) = P(D+ \mid \mu) = P_{X\_al}(\mu) := 1 - pnorm\left(AL, \mu_k(\mu), \frac{CV\cdot\mu}{\sqrt{n}}\right)$

$P(x<AL \mid \mu) = P(D- \mid \mu) = P_{x\_AL}(\mu) := pnorm\left(AL, \mu_k(\mu), \frac{CV\cdot\mu}{\sqrt{n}}\right)$

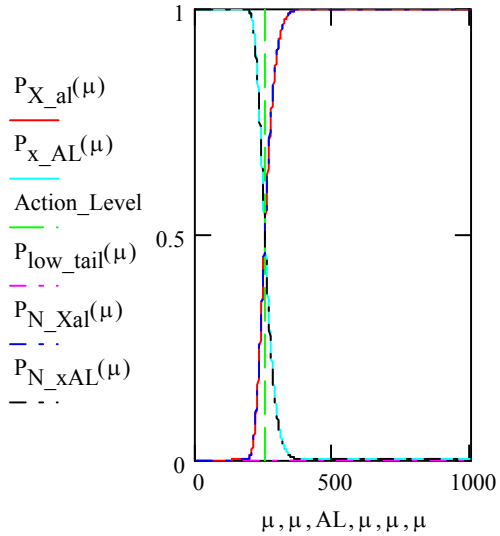
The functions above include probabilities for μ<0 (and even small probabilities for μ to exceed the highest concentration that can possibly be encountered). This can cause an error if the tail of the probability distribution reaches significantly below zero. Therefore, the functions are normalised. No normalisation is performed for the high tail exceeding the highest possible concentration, since the probability for this is very small for ordinary problems.

$P_{low\_tail}(\mu) := pnorm\left(0, \mu_k(\mu), \frac{CV\cdot\mu}{\sqrt{n}}\right)$

The normalised probability functions:

$P_N(x>AL \mid \mu) = P_N(D+ \mid \mu) = P_{N\_Xal}(\mu) := \frac{1 - pnorm\left(AL, \mu_k(\mu), \frac{CV\cdot\mu}{\sqrt{n}}\right)}{1 - P_{low\_tail}(\mu)}$

$P_N(x<AL \mid \mu) = P_N(D- \mid \mu) = P_{N\_xAL}(\mu) := \frac{pnorm\left(AL, \mu_k(\mu), \frac{CV\cdot\mu}{\sqrt{n}}\right) - P_{low\_tail}(\mu)}{1 - P_{low\_tail}(\mu)}$



The figure presents the probability of the measured average concentration to exceed (large X) and to fall below (small x) the action level AL (green line) as a function of the true mean concentration μ.

Weighing sample uncertainty with prior information

Below, **the probabilities P(sample|state)** are calculated by integration. Probabilities $P(D^+|C^+)$ and $P(D^-|C^+)$ are calculated by integration from the action level to infinity. Probabilities $P(D^+|C^-)$ and $P(D^-|C^-)$ are calculated by integration between zero and the action level. All integrations are performed on normalised probability functions, i.e. after elimination of the tails below zero. The probability in the high tail above realistic concentrations is ignored because this probability is usually extremely small.

$$P(D+|C+) = P(D+,C+)/P(C+) = P_{Dp\_Cp} := \int_{AL_2}^{\infty} P_{N\_Xal}(\mu) \cdot \frac{f_{pri}(\mu)}{P_{Cp}} \, d\mu \qquad P_{Dp\_Cp} = 0.982$$

$$P(D-|C+) = P(D-,C+)/P(C+) = P_{Dm\_Cp} := \int_{AL_2}^{\infty} P_{N\_xAL}(\mu) \cdot \frac{f_{pri}(\mu)}{P_{Cp}} \, d\mu \qquad P_{Dm\_Cp} = 0.018$$

$$P(D-|C-) = P(D-,C-)/P(C-) = P_{Dm\_Cm} := \int_{0}^{\infty} P_{N\_xAL}(\mu) \cdot \frac{f_{pri}(\mu)}{P_{Cm}} \, d\mu - \int_{AL_2}^{\infty} P_{N\_xAL}(\mu) \cdot \frac{f_{pri}(\mu)}{P_{Cm}} \, d\mu$$

$$P_{Dm\_Cm} = 0.949$$

$$P(D+|C-) = P(D+,C-)/P(C-) = P_{Dp\_Cm} := 1 - P_{Dm\_Cm} \qquad P_{Dp\_Cm} = 0.051$$

Estimation of probabilities P(sample) = P(D)

$$P(D^+) = P_{Dp} := P_{Cm} \cdot P_{Dp\_Cm} + P_{Cp} \cdot P_{Dp\_Cp} \qquad P_{Dp} = 0.786$$

$$P(D^-) = P_{Dm} := P_{Cm} \cdot P_{Dm\_Cm} + P_{Cp} \cdot P_{Dm\_Cp} \qquad P_{Dm} = 0.214$$

Estimation of probabilities P(state|sample) = P(C|D) by Bayesian updating

Probability of correct classification as contaminated:

$$P(C+|D+) = P_{Cp\_Dp} := \frac{P_{Cp} \cdot P_{Dp\_Cp}}{P_{Cp} \cdot P_{Dp\_Cp} + P_{Cm} \cdot P_{Dp\_Cm}} \qquad P_{Cp\_Dp} = 0.986$$

Probability of incorrect classification as contaminated (overestimation):

$$P(C-|D+) = P_{Cm\_Dp} := \frac{P_{Cm} \cdot P_{Dp\_Cm}}{P_{Cp} \cdot P_{Dp\_Cp} + P_{Cm} \cdot P_{Dp\_Cm}} \qquad P_{Cm\_Dp} = 0.014$$

Probability of incorrect classification as uncontaminated (underestimation, consumer's risk):

$$P(C+|D-) = P_{Cp\_Dm} := \frac{P_{Cp} \cdot P_{Dm\_Cp}}{P_{Cp} \cdot P_{Dm\_Cp} + P_{Cm} \cdot P_{Dm\_Cm}} \qquad P_{Cp\_Dm} = 0.067$$

Probability of correct classification as uncontaminated:

$$P(C-|D-) = P_{Cm\_Dm} := \frac{P_{Cm} \cdot P_{Dm\_Cm}}{P_{Cp} \cdot P_{Dm\_Cp} + P_{Cm} \cdot P_{Dm\_Cm}} \qquad P_{Cm\_Dm} = 0.933$$

Prior analysis

First we sum up the costs in the payoff table.

$$Tcost := \sum_{i=1}^{3} cost^{\langle i \rangle} \qquad Tcost = \begin{pmatrix} -1000 \\ -1000 \\ -2000 \\ 0 \end{pmatrix}$$

Prior objective function (expected cost):

$$EC_{pri} := \left| \begin{array}{l} EC \leftarrow \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \text{for } a \in 1..2 \\ \quad \text{for } s \in 1..2 \\ \quad\quad EC_a \leftarrow EC_a + P_{prior_s} \cdot Tcost_{2(a-1)+s} \\ \max(EC) \end{array} \right. \qquad EC_{pri} = -1 \times 10^3$$

Preposterior analysis

Preposterior probabilities of sample data: $P_D := \begin{pmatrix} P_{Dp} \\ P_{Dm} \end{pmatrix}$ $\qquad P_D = \begin{pmatrix} 0.786 \\ 0.214 \end{pmatrix}$

Preposterior probabilities of state given sample data:

$$P_{prep} := \begin{pmatrix} P_{Cp\_Dp} \\ P_{Cm\_Dp} \\ P_{Cp\_Dm} \\ P_{Cm\_Dm} \end{pmatrix} \qquad P_{prep} = \begin{pmatrix} 0.986 \\ 0.014 \\ 0.067 \\ 0.933 \end{pmatrix}$$

Preposterior objective function (expected cost):

$$EC_{prep} := \left| \begin{array}{l} EC \leftarrow \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \text{for } d \in 1..2 \\ \quad \text{for } a \in 1..2 \\ \quad\quad \text{for } s \in 1..2 \\ \quad\quad\quad EC_{a,d} \leftarrow EC_{a,d} + P_{prep_{2(d-1)+s}} \cdot Tcost_{2(a-1)+s} \\ P_{D_1} \cdot \max\left(EC^{\langle 1 \rangle}\right) + P_{D_2} \cdot \max\left(EC^{\langle 2 \rangle}\right) \end{array} \right. \qquad EC_{prep} = -814.403$$

Data worth estimation

Expected Value of Perfect Information, $EVPI := \sum_{i=1}^{2} P_{prior_i} \cdot \max\left(Tcost_i, Tcost_{i+2}\right) - EC_{pri}$ $EVPI = 210.526$

Expected Value of Sample Information, $EVSI := EC_{prep} - EC_{pri}$ $EVSI = 185.597$

Reliability (ratio between EVSI and EVPI), $Reliability := 100 \cdot \frac{EVSI}{EVPI}$ $Reliability = 88.159$

Input and result sheet for Data Worth Analysis

Prior estimation (PDF) of mean concentration

Reasonable minimum value: a ≡ 50
Most likely value: m ≡ 200
Reasonable maximum value: b ≡ 1000
Appropriate prior distribution: PDF ≡ 1

1. Uniform PDF
2. Triangular PDF
3. Normalised normal PDF, 95%
4. Normalised normal PDF, 99%
5. Lognormal PDF, 95%
6. Lognormal PDF, 99%

IMPORTANT! The selected PDF should reflect the prior information about the mean concentration, which is NOT equal to the distribution of individual measurements!

Action level

Action level: AL ≡ 250

Planned sampling program

Number of samples, n ≡ 12
Cost per sample, $C_S \equiv 4$
Sample uncertainty (random part), coefficient of variation, CV ≡ .4
Sample uncertainty (systematic part), two cases:

1. Additive-constant bias, Δ ≡ 0
2. Multiplicative bias, k ≡ 0

Δ and k are positive when the measured value is higher than the true value.

Payoff table (costs)

Column 1:	Investment cost	Row 1:	Remediation and C^+
Column 2:	Cost of failure	Row 2:	Remediation and C^-
Column 3:	Benefit	Row 3:	No remediation and C^+
		Row 4:	No remediation and C^-

$$\text{cost} \equiv \begin{pmatrix} 0 & -1000 & 0 \\ 0 & -1000 & 0 \\ 0 & 0 & -2000 \\ 0 & 0 & 0 \end{pmatrix}$$

Costs are negative and benefits are positive.

Data Worth

Expected Value of Perfect Information: EVPI = 210.526
Expected Value of Sample Information: EVSI = 185.597
Reliability (ratio between EVSI and EVPI): Reliability = 88.2 %
Expected Net Value, $ENV := EVSI - n \cdot C_S$ ENV = 137.597