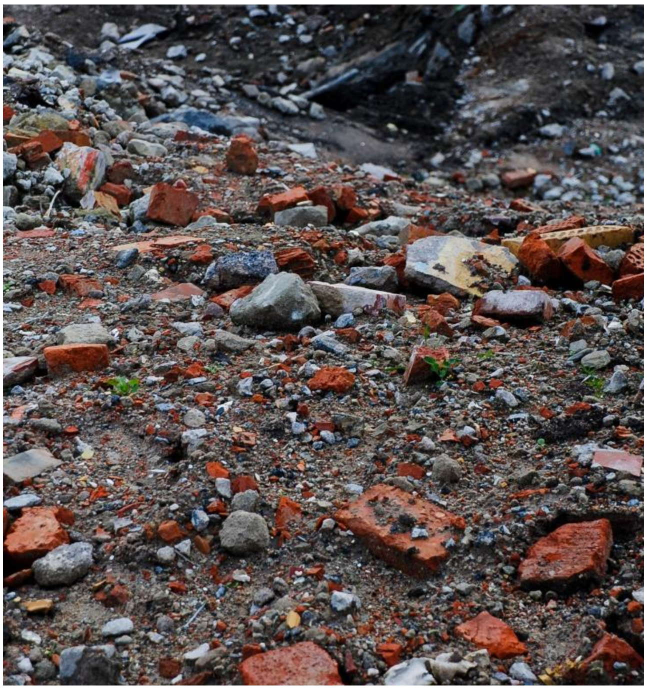
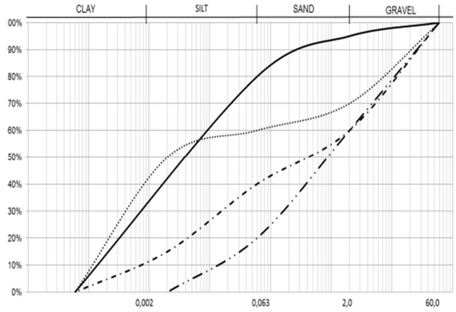
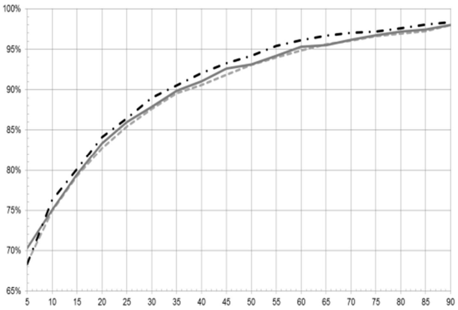






Optimized sampling schemes for filling material

Applied on contaminated sites through statistical analysis

Master of Science Thesis in the Master Degree Programme Geo and water engineering

ANDREAS JOHANSSON
MAJA ÅNELIUS

Department of Civil and Environmental Engineering
Division of GeoEngineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2011
Master's Thesis 2012:9

MASTER’S THESIS 2012:09

Optimized sampling schemes for filling materials

Applied on contaminated sites through statistical analysis

Master of Science Thesis in the Master’s Programme Geo and Water Engineering

A. JOHANSSON

M. ÅNELIUS

Department of Civil and Environmental Engineering
Division of GeoEngineering
Engineering Geology Research Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2011

Optimized sampling schemes for filling materials
Applied on contaminated sites through statistical analysis
A. JOHANSSON
M. ÅNELIUS

Examensarbete / Institutionen för bygg- och miljöteknik,
Chalmers tekniska högskola 2012:9

Department of Civil and Environmental Engineering
Division of GeoEngineering
Engineering Geology Research Group
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: + 46 (0)31-772 1000

Cover:
Left: Filling materials from one of the reference sites, see page 19.
Upper right: Grain size distributions for different soil classes, see page 21.
Lower right: Confidence levels for one of the studied sites, see page 50.

Chalmers Reproservice / Department of Civil and Environmental Engineering
Göteborg, Sweden 2011

Optimized sampling schemes for filling materials
Applied on contaminated sites through statistical analysis
Master of Science Thesis in the Master's Programme Geo and Water Engineering
A. JOHANSSON & M. ÅNELIUS
Department of Civil and Environmental Engineering
Division of GeoEngineering
Engineering Geology Research Group
Chalmers University of Technology

ABSTRACT

An important part of contaminated site investigations is the initial soil survey. There are many well working systems about the spatial pattern design of the soil samples during the survey. However, as it is uncommon to submit all collected soil samples for chemical analysis it is important that the selection of which samples to send is performed in a structured and statistically sound manner.

This thesis aim is to contribute to the research by providing a tool that will increase both the environmental and economic sustainability for environmental analysis of contaminated sites. The specific aim of the thesis is to provide a tool that helps to select which samples to send to chemical analysis by analyzing the variance in contamination levels in filling materials.

The thesis is focused on the contaminant Polycyclic Aromatic Hydrocarbons, PAHs, which are common pollutants that can be found at many contaminated sites and are hazardous to both human health and the environment.

Five reference sites in western Sweden have been chosen for the study, all consisting of anthropogenic soil. A soil classification system has been developed and used to divide the soil samples into different classes, or strata, for which statistical analysis has been performed. The focus of the statistical analysis has been to compare three different sampling schemes; simple random sampling, proportional stratified random sampling and optimal stratified random sampling. From this, a recommendation on how to optimize the soil survey using a statistical analysis on the sample selection procedure will be formulated.

The result indicates that there is a correlation between soil class and contamination level and that the number of chemical analysis can be decreased if using an optimal stratified random sampling scheme. However, this only applies to sites contaminated by a diffuse source. Since some information about the basic statistics of the strata is necessary in order to perform an optimal stratified sampling scheme the method is more suitable if surveys have previously been performed.

Finally, the recommendation from this thesis is that designing a sampling scheme using the optimal stratified random sampling is a tool that is well suited if sampling has been performed in the area before and the materials are non-landfill materials with a diffuse contamination. The thesis also suggests a methodology on how to design a sample scheme using optimal stratified random sampling. For the most reliable result the field survey should not consist of auger drilling only. An excavation pit will give a better result since it provides a better view of the soil stratification and includes all grain sizes.

Key words: Contaminated land, PAH, filling material, stratified sampling, soil survey, statistical simulation, sampling scheme, survey design

Table of Contents

Figures, tables and images

Appendices

Preface

This master's thesis has been written as the final part of the Civil and Environmental Engineering program at Chalmers University of Technology during the autumn of 2011. The thesis has been written in collaboration with Norconsult AB, who has also provided financially, as well as NCC AB.

We wish to thank a number of people for their support and guidance in this project. First, we are grateful to our supervisor at Chalmers Lars Rosén who has helped us take a loose idea and turn it into a real project. We are also thankful to the personal at the Civil and Environmental laboratories, especially Peter Hedborg and Mona Pålsson, for their assistance during the practical part of the project and to Tommy Norberg for the advice on statistics.

We are also grateful to our supervisor at Norconsult, Ulf Johansson, who took time from his busy schedule and gave us input and guidance and together with Annelie Loberg and Katarina Lennartsdotter Parkkonen helped us formulate the aim and scope of the project. We also wish to thank the rest of the staff at the Environmental Analysis team for their support and encouragement during the completion of the thesis, and that provided the financial resources required for the project.

We would like to thank Malin Norin at NCC who contributed with criticism and input on the thesis. We are also thankful of the help given to us by Johannes Senning and Petra Brinkhoff at NCC in locating data for our project.

Finally we wish to thank the environmental administrators at Västra Götaland County who answered our questions about contaminated sites in the region, Mikael Johansson who read the thesis and gave us advice on the disposition as well as Pär-Erik Back and Jenny Norrman who also read the thesis and gave us input. Finally, we wish to thank Ale municipality, Bengtsfors municipality and the Traffic Authority in Gothenburg who gave us permission to use their data for analysis.

With hopes of a receding storm and cold and white winter,

Andreas Johansson and Maja Ånelius,

Gothenburg, December 2011

Norconsult

NCC

Notations

General statistic variables:

[illegible]	Mean value
[illegible]	Data value *i*
[illegible]	Standard deviation
[illegible]	Variance
[illegible]	Number of sample data values
[illegible]	Size of population
[illegible]	Coefficient of variation
[illegible]	Standard error of the mean
[illegible]	Estimated variance of the mean
[illegible]	Sample fraction for stratum a
[illegible]	Population size for stratum a
[illegible]	Weight of stratum a
[illegible]	Optimal sampling fraction for stratum a
[illegible]	Estimated variance of the mean with proportional stratified random sampling
[illegible]	Estimated variance of the mean with optimal stratified random sampling
[illegible]	The a’th stratum
[illegible]	Number of strata

Thesis Specific Variables

[illegible]	Mean concentration for PAH-group *J**
[illegible]	Standard deviation of concentration for PAH-group *J**
[illegible]	Variance of concentration for PAH-group *J**
[illegible]	Number of soil samples within soil class *K′*
[illegible]	Total number of soil samples in the survey
[illegible]	Coefficient of variation of concentration for PAH-group *J**
[illegible]	Sample fraction weight with proportional stratified random sampling
[illegible]	Sample fraction weight for PAH-group *J** with optimal stratified random sampling
[illegible]	Standard error of the mean concentration with simple random sampling for PAH-group *J**
[illegible]	Standard error of the mean concentration with proportional stratified random sampling for PAH-group *J**
[illegible]	Standard error of the mean concentration with optimal stratified random sampling for PAH-group *J**
[illegible]	Number of data samples in simulation
* [illegible]	PAH-group, L = Low, M = Mean, H = High
′ [illegible]	Soil class, *I* to *V*

Glossary and abbreviations

Chemical analysis	Specific soil samples are analyzed for contamination levels
Data value	A specific value within a sample
Population	All possible data values representing all the soil samples collected in a soil survey.
Sample	The data of all the soil samples within the survey that are chemically analysed.
Soil Class	Soil determination based on system devised for this thesis
Soil sample	A specific soil sample
Soil survey	The field procedure in which soil samples are taken
Survey	Collecting data values for a sample

EPA	Environmental Protection Agency
PAH	Polycyclic Aromatic Hydrocarbon
TOC	Total Organic Carbon
LMW	Low Molecular Weight
MMW	Medium Molecular Weight
HMW	High Molecular Weight
SRS	Simple Random Sampling
PSRS	Proportional Stratified Random Sampling
OSRS	Optimal Stratified Random Sampling

1 Introduction

After almost 200 years of industrial activity a substantial portion of the soil in Sweden, especially in our cities, is contaminated. These activities were going on without environmental regulations for a major part of this time, and even if some control and supervision was introduced in the late sixties it was not until January 1999 that the Swedish Environmental Code came into effect (Miljöbalken, 1999, p. 12:3).

In the same year, the Swedish Environmental Protection Agency (Swedish EPA) started an inventory of possible contaminated areas and released handbooks on environmental quality criteria. Since then 80 000 sites have been identified as potentially contaminated and prioritized for further actions. A number of these have been remediated.

These sites have been contaminated by a variety of sources, but three common sources of contamination are from industries that have released hazardous materials, sites that have been used more or less as landfills and filling out of low lying areas using contaminated anthropogenic soils. The last is especially true for Gothenburg, where the current shape of the central parts of the city has been formed by filling up marshes of the estuary of the river Göta Älv to create habitable land.

The Swedish Environmental Code states that the party responsible for the contamination also has to pay for the remediation, something that is known as the polluter pays principle. While this might seem to be a straightforward law, it is not always that easy. The company that caused the contamination can be defunct or untraceable, and in the case of filling materials where the soil was contaminated when put in place there is not always one specific responsible party.

For the cases where the pollution source and the responsible party are unknown, the present owner has the responsibility to ensure that the contaminations does not result in unacceptable risks towards human health and the environment (Miljöbalken, 1999, pp. 3a:7-3a:12). Because of this it is of importance that a soil survey is performed before a purchase of possibly contaminated land, since the liability of the contamination is also transferred to the new owner. For the highly prioritized contaminated areas where there is no responsible party the Swedish EPA contributes governmental financial resources to the remediation.

To investigate the contamination levels in an area, a soil survey has to be performed. In the soil survey, the goal is to by means of drilling and excavation pits find out what contamination levels are representative for an area and how those contaminants spread to the surroundings. There are several methods to perform these surveys, but a common goal for all of them is to find a balance between a sufficient amount of data and the cost of the survey.

The next step, after the soil survey, is to remove any contamination levels exceeding the guideline values through remediation. The method that is most common today is the so called “dig-and-dump” approach which is performed by excavating all material that is assumed to be contaminated and transport everything to a waste disposal and/or treatment facility. The main advantage of the “dig-and-dump” approach is the relatively low cost and the low risk of leaving contaminations at the site. The disadvantage of the method is that it requires extensive groundwork, a large amount of transportation and there are considerable risks associated with removing an excess of material. While several others more advanced and sustainable, methods exist they are still uncommon in practice (Brinkhoff, 2011).

It is apparent that contaminated soil is a large problem in construction work today, and will be so for many years forward. While the focus in construction work is still usually on legal requirements and low cost, it should be remembered that removing chemicals and pollutants from the soil is something positive. The focus should be on finding methods of remediation that removes all the pollutants, but in a way that is both sustainable and cost effective.

1.1 Problem definition

There are still several areas where the methodologies for both soil survey and remediation can be improved. The knowledge can be expanded, both regarding survey and remediation, to improve the economic and environmental sustainability of the process.

An important part of the soil investigation process is the initial soil survey. There has been extensive research into drilling pattern design, but this has mostly been aimed on the spatial locations of the drilling points. As important as the pattern design is the selection of which soil samples to submit for chemical analysis. The analyzed soil samples should commonly represent the entire area in regard to contamination and should be selected in a way that they can still be considered random in order to be able to perform statistical analysis on the data. To be able to perform a statistical analysis it is also important that the studied area has homogenous geology, contamination history and other key features in order for the soil samples to be comparable. In other words, the soil samples should originate from an area that has statistically homogenous features.

There has been extensive research into the connection between soil type and contamination levels to find which soil fractions the contaminants bind to. The result is often that the finer the fraction the higher the contamination concentrations, but less analysis has been focused on studying the variance of contamination levels in different soil fractions. If there is knowledge about how the variance differs between the fractions, it can be applied to find ways to design a survey that will reduce the uncertainty by sampling the soil fractions where the variance is high to a larger extent.

Under ideal conditions, all soil samples that have been extracted are also submitted for chemical analysis. This is seldom true, however, since the chemical analyses are costly and can be time-consuming. The problem is to decide how many soil samples that should be sent for chemical analysis and which once to choose. The goal is to find a balance between the number of analyses and the uncertainty of the result which will make the process more cost efficient. Another goal is to identify how the methodology can be applied for other parts of the soil survey and the remediation process.

1.2 Aim

The overall aim of this thesis is to describe a method that will increase both the environmental and economic sustainability for environmental analysis of contaminated sites. This can be achieved by decreasing the uncertainty in the analysis without additional costs compared to present procedures.

The specific purpose is to improve the methods used in survey and remediation of contaminated soil based on a statistical analysis of the variance of contamination levels in filling materials. From this a recommendation on how to optimize the soil survey using a statistical analysis on the sample selection procedure will be formulated.

1.3 Delimitations

The study will focus on five locations in western Sweden which means that the results will only be valid for locations with similar geology, contamination history and climate. Since almost 1/8 of all contaminated sites in Sweden are estimated to be located in Västra Götaland County[1], it is of interest to concentrate on this area. In addition the result can be used where similar conditions are found for sites outside the studied region. The soils we build on today are often anthropogenic and therefore this thesis will focus on filling materials rather than natural soils.

The study will be performed on already analyzed sites and no new soil samples will be taken. The five sites that will be considered are Surte 2:38 in Ale Municipality, EKA Chemicals in Bengtsfors Municipality, the Göta Älv Bridge in City of Gothenburg, Kvillebäcken in City of Gothenburg and Hexion in Mölndal Municipality. These sites are suitable for analysis since they are polluted from different types of activities and have different contaminant migration conditions.

The substances that will be studied are polycyclic aromatic hydrocarbons, PAHs. PAHs are common pollutants that can be found in most places where human activity has been performed and they are hazardous to both human health and the environment. This together with the fact that the information on their behavior in nature is not as abundant as for example for heavy metals makes them highly interesting to study.

1.4 Method

Literature studies will be made continuously during the process. These will be used as a support for the thesis while the main results will be from the statistical analyses. Areas where literature studies are applied are: background information about PAHs, the reference sites and statistical theory, as well as previous research within the area of study.

The next step will be to create a soil classification system based on grain size. The soil types in the data are based on an ocular determination that was performed by field technicians during the soil surveys. Since this is made without any scientific methods, grain size distributions will be compared to the ocular analysis for a selected number of soil samples to verify their quality. Where distributions are not available in the source material new analyses will be performed at the laboratories of the Department of Civil and Environmental Engineering, Chalmers University of Technology. The soil type will be analyzed using dry sieving and hydrometer analysis.

The first step in the analysis part of the thesis is to collect data from the chosen sites. Relevant information will be selected and summarized into a database. The data will be obtained from soil surveys performed by Norconsult AB (Norconsult) and NCC AB (NCC). The soil types in the data will be classified according to the classification system. The soil types from the field survey will also be compared with the result from the sieving.

When the data has been processed a basic statistical analysis can be performed. The methodology will be applied with the goal of finding the most representative mean value for the areas. Parameters of interest are mean value, standard deviation, variance and the coefficient of variation. The statistical analysis will be performed on each site as a unit as well

---

[1] Uffe Schultz Environmental Protection Department Västra Götaland County, email 28'Th of September 2011.

as for the different soil classes. No sub-division into smaller areas with a common characteristic will be performed.

More in-depth statistical analyses will be made using a method known as stratified random sampling. The results will then be analyzed using two different randomized sampling simulations, one sampling the data and one using probabilistic distributions. The statistics and calculations will be described in more detail in later chapters.

Finally the results from the statistical analyses will be applied to give recommendations on ways to optimize the procedure for selecting which, as well as how many, soil samples to perform chemical analysis on. The results will also be used to find applications for increasing the cost efficiency in the remediation processes.

While no classification such as the one with grain sizes will be performed for total organic carbon (TOC), a minor analysis of the correlation between TOC and contamination level will be carried out. This minor analysis will be used as a comparison to the other results.

The database will be compiled in Excel, which will also be used for most of the statistical analyses. To simulate the sampling processes, the software Crystal Ball will be used. Histograms will be drawn with the ProUCLsoftware developed by the US EPA specifically for statistical analysis of contaminated soil (US EPA, 2011).

1.5 Disposition

The thesis is divided into four main steps, which can be seen in Figure 1.1, starting with the pre-conditions that include research that has already been performed, statistical background and summaries of site assessments. The next part is a description of the methodology that will be used to arrive at the results and finally it is all integrated into conclusions and recommendation on how to proceed with the results.

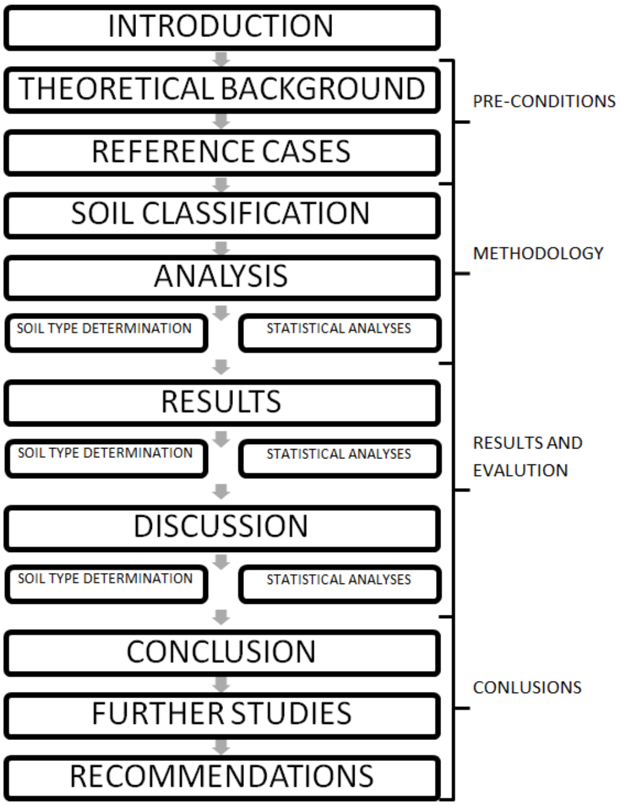


Figure 1.1: Disposition of the thesis.

1.6 Permit to use data

Permission to use the sample data has been given by the owners, which are: Ale Municipality for Surte 2:38, Bengtsfors Municipality for EKA, NCC AB for Hexion and Kvillebäcken and the Transport Authority in Gothenburg City for the Göta Älv Bridge. Permission to use the data is only valid for this master’s thesis and it cannot be reproduced or used without permission from the owners.

2 Theoretical background

This chapter will describe the preconditions for the analyses made in the study, including background on what research already has been performed in the area of study, field survey procedures relevant to the thesis, the behavior and characteristics of PAH and finally theory on the statistics used in the analysis.

2.1 Previous research

Several researchers have studied uncertainties in soil surveys for contaminated land and in what part of the process the main uncertainties originate. A common approach is based on a system devised by French mining-engineer Pierre Gy in the 1960's for mining purposes and can be seen in Figure 2.1, where several possible causes for uncertainties are identified. These include sampling errors, caused by the procedure used to design and perform the survey, and analytical errors which are errors caused by insufficient precision in the chemical analysis. Analysis of the sampling procedure uncertainties have found that the error can commonly be in the range of 30 – 40% and most of this is caused by the sample selection process (Back, 2003, pp. 43-84).

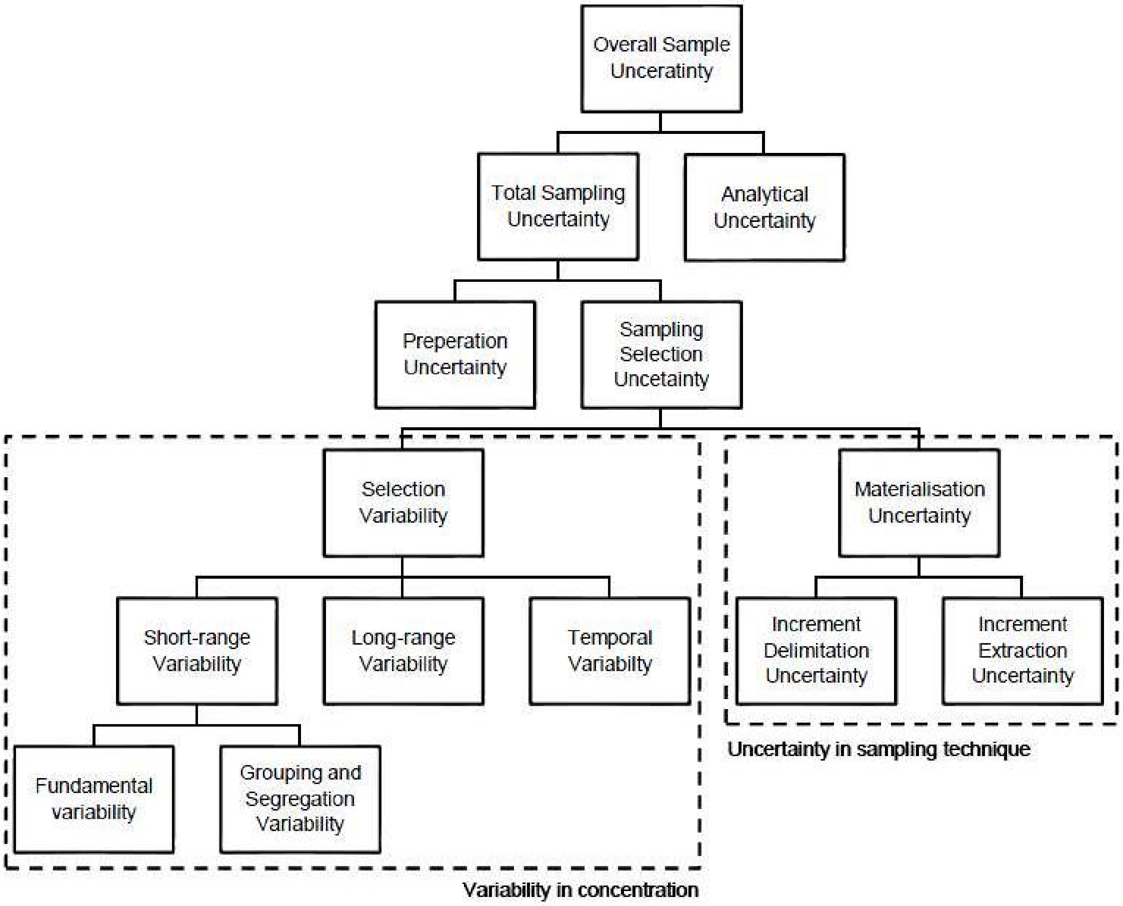


Figure 2.1: Approach to sampling uncertainty from Back (2003).

Some of the components in the approach to sampling uncertainty have an extra relevance for the thesis. The parameters in the Selection Variability are all related to the concentrations, where the Short-range Variability is connected to variations caused by the geology, the Long-range Variability on variations caused by spatial location and Temporal Variability is connected to effects on the concentrations over time. Of the Materialization Uncertainties, the Increment Extraction Uncertainty is the most interesting as it is connected to an erroneously extracted soil sample. For full descriptions on the uncertainties and calculations of the total uncertainty, see Back (2003).

Most of the previous research into soil survey design is aimed at the spatial location of the drilling points. In many guidelines, for example from both the US EPA (2002) and the Norrman and Back et al. for the Swedish EPA (2009), several methods of sample design are described including a stratified random sampling method but with the soil samples selected before geological surveys at the site have been performed. As it is common that not all extracted soil samples are sent for analysis, a stratification applied after the field survey can be beneficial as it will be based on a more accurate data.

Creating a standardized method for selecting soil samples to submit for chemical analysis also has statistical benefits. As was stated before much of the previous research has been aimed at optimizing the field part of the soil survey. If the drilling points have been placed in a random pattern and all collected soil samples are then sent for chemical analysis, there will be no statistical errors. However, if not all the soil samples are sent for analysis, and the selection of which to send is not performed with a random method, there is a risk of bias, a systematic error (Rice, 2007, pp. 199-200).

Systematic errors have been found to be one of the most dominant types of errors, and the problem is further increased by the fact that it cannot be reduced by taking more samples. Compared to the random error, which is the error caused by variations in the data, systematic errors will remain with an increased sample size. In addition, the systematic errors are hard to estimate and are entirely dependent on the source of the error (Back, 2003, p. 8).

Some of the analysis in the study will be aimed at the value of additional samples, something that several researchers have studied previously, for example Back (2006) and Norrman (2004). The studies performed in their papers are of a Bayesian approach, which means that the value of additional samples are calculated based on previous sampling and assumptions on the information gained by additional samples. What will be done in this thesis is instead to analyze how different sampling schemes can reduce the number of samples but retain the same confidence. The results could then be used in the Bayesian approach, but this will not be performed in this thesis.

Another approach for determining the optimal number of soil samples is presented by Norrman et al. (2009), where the sample size is calculated based on the coefficient of variation. While this is related to what is aimed at in this thesis, the focus in that method is not to reduce the sample size but to determine it. Their method could, however, be used as a reference when designing a soil sample using the stratified random sampling presented here.

2.2 Field Survey Procedure

Field surveys in Sweden are commonly performed according to a guideline released by the Swedish Geotechnical Society (2004) which standardizes the working process for soil, water and pore gas sampling. The standardized soil survey entails a number of steps related to preparations, documentation, health hazards, sampling technique and calibration of instruments.

One of the steps in the preparations is a study of geological and hydrological conditions at the site. If no soil surveys, either geotechnical or environmental, have been performed at the site before the available maps are commonly those released by the Geological Survey, which have a scale of 1:50 000 at best. This means that details smaller than approximately 100 m will not be visible (Geological Survey of Sweden, 2011).

During the field survey, the guideline stipulates that the extracted soil will be documented. This documentation should include an ocular analysis of the soil type, observations on groundwater and other notable aspects of the soil. These other aspects can be anthropogenic materials, which could indicate that the material is a filling material, smell and other deviations from what can be expected in a natural, undisturbed soil (Swedish Geotechnical Soceity, 2004). The field classification of the soil type is not intended to be more than a support for analysis of the stratigraphy and cannot be considered to be a correct determination of the soil type (Swedish Geotechical Society, 1996, p. 8:16).

Sampling of contaminated soil in Sweden is commonly performed with two methods: auger drilling, as seen in Figure 2.2, and excavation pits. The auger drilling is performed using a drilling rig and the excavation pit can be performed either by an excavator or if only a small pit is required by hand. The advantage with the excavation pit is that it gives a good overview of the soil stratigraphy and works for most soil types while the auger drilling has the advantage of requiring only a small hole, the auger is commonly 20 – 40 mm in diameter, and thereby having a low impact on the survey site. The disadvantage with the auger drilling is that the samples will be disturbed as the layering can change when the auger goes up or down through the soil and that it cannot extract large soil fractions such as stones and boulders (Swedish Geotechical Society, 1996) .

Figure 2.2: Soil sampling with a drilling rig, © Norconsult.

2.3 Polycyclic aromatic hydrocarbon, PAH

In this section, the properties of the substance chosen for analysis will be described. This is both to provide some background information on why PAHs are of interest when analyzing contaminated soil but mainly to give a description of their properties in nature.

PAHs are a large group of organic compounds that consists of 100 different molecules. In connection with contaminated soil 16 of these molecules are used as indicators due to their common occurrence in pollution sources and their harmful effects. The molecular structure of PAHs consists of aromatic rings, see figure 2.3, which defines their characteristics. They are formed during incomplete combustion of organic material, especially at temperatures around 600° C (Sterner, 2003).

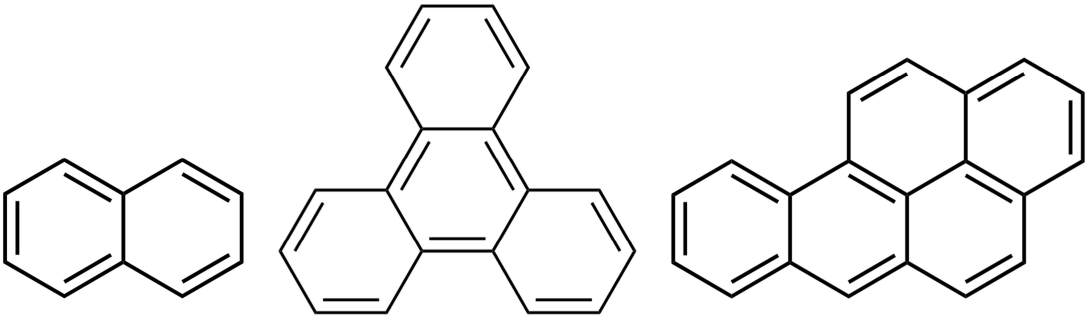


Figure 2.3: Example of three PAH molecules, from the left: Naphthalene, Trephynelene and Benzo[a]pyrene (Wikipedia, 2011).

Common sources for PAHs are wooden fires, combustion engines, tobacco smoke as well as barbequed and smoked provisions. They can also be found already formed in coal tar, Creosote, asphalt predating 1975 (Miljöförvaltningen, 2006) and car tires from before 2010, when an international law banned the use of PAH in the tires (Health and Safety Authority, Ireland, 2010).

The main health concern is carcinogenic effects, and PAHs are estimated to cause 10 – 100 lung cancer cases in Sweden every year (Bernes, 1998, p. 112). Other possible health effects are genetic alterations and hormone disturbances. PAHs are ranked in the highest hazard class, by the Swedish EPA, together with for example arsenic, lead and mercury (Swedish EPA, 2000, p. 20).

PAHs have a low volatility, which means that when they are released from a source the deposition usually occurs locally (Bernes, 1998, p. 47). In addition, the solubility with water is low which means that after the deposition the compounds are immobile compared to other organic pollutants. The result is that the substances are often accumulated at a location close to the source, which means that if the contamination source is not removed the concentrations in the soil will increase over time. PAHs can often be found in the organic material of the soil (Sterner, 2003).

PAHs are hard to separate from each other and are therefore often handled as a group (Sterner, 2003). Previously they were divided into carcinogenic and non-carcinogenic but now they are usually grouped into three classes based on the molecular weight. These are low, middle and high molecular weight. The new division is used since it is considered to give a more accurate representation of their behavior in nature (Swedish EPA, 2011) and will henceforth be used in this thesis. The molecules included in the new as well as the old groups can be found in Appendix 1.

One important difference between the groups is that the 3 molecules included in the low molecular weight group have a higher migration potential than the remaining 13 and some of them are even considered volatile. The molecules with lower weight also have a much lower half-life, up to 1/10 of the denser molecules. As an example, the low molecular weight PAH naphthalene has a half-life in soil of 70 days, while the same for high molecule weight Benzo[a]pyrene is 710 days (Swedish EPA, 2007).

2.4 Statistics

This section will provide the theory behind the statistics that will be used to analyze the data. It is divided into four parts: a section about the statistical parameters used for basic analyses, a section about the optimization method stratified random sampling, a description of the sampling simulation and finally a section on sensitivity analysis.

The statistical calculations will later be used to perform two variance analyses, first on data from the site surveys and in a later stage on data from sampling simulations performed in the thesis. The statistics described below will be used both to analyze and compare the reference sites and to evaluate the efficiency of the different sample selection procedures evaluated in the thesis. Some of the parameters are also required for the calculations and some of these may not be presented in the results section, but they are still crucial for the analysis performed in the thesis.

2.4.1 Statistical parameters

In this thesis the measured values are used to make estimations of the true values for the statistical parameters. Since determining the true mean values, standard deviations and variance requires an infinite amount of data, these are not known and the calculations will be performed using estimates the parameters since the amount of data is limited.

The mean, or the expected value, is one of the most basic properties of a sample but also one of the most important, as it is the basis of many statistical calculations. The mean value, $\bar{x}$, is defined as the most likely data value and is calculated using equation 2.1 (Rice, 2007).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i) \tag{2.1}$$

While the mean describes the most likely data value, the standard deviation is used to describe the scatter of the sample. The mean value plus/minus one standard deviation covers 68% of all data values. This means that a low standard deviation indicates that the data in general have values close to the mean and vice versa. The standard deviation, s, is calculated with equation 2.2 (Rice, 2007).

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.2}$$

To compare the standard deviation with the mean the coefficient of variation, C_x, can be used. It is calculated as the standard deviation divided by the mean, as seen in equation 2.3. The value is used to evaluate the data, if for example C_x is larger than 1 it indicates that if the data would be normally distributed there would be a high probability of negative values which is not valid for concentrations and consequently the data likely follows another distribution, for example lognormal (Rice, 2007).

$$C_x = \frac{s}{|\bar{X}|} \tag{2.3}$$

While the standard deviation is a useful tool to describe the distribution of the sample, most calculations will be performed using the variance, s^2, as seen in equation 2.4. The variance is the square of the standard deviation and does not have a unit that can be compared with the sample data values (Rice, 2007).

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{2.4}$$

In the same way as the variance is used to describe the variation in the data values of a sample the so-called estimated variance shows the variation of the calculated mean. To explain the estimated variance of the mean, $s_{\bar{X}}^2$, it is easier to describe the properties of its square root, the standard error. The standard error can be described as the difference in mean value between two or more surveys of the same size from a population. For example, if a contaminated area is surveyed several times the mean values will differ between each survey. The standard error is an estimate of the magnitude of this difference. However, multiple surveys are not actually taken; instead equation 2.5 is used to calculate the theoretical estimated variance based on only one survey (Rice, 2007).

$$s_{\bar{X}}^2 = \frac{s^2}{n} \tag{2.5}$$

2.4.2 Stratified random sampling

Stratified random sampling is the tool that will be used to optimize the sampling. The concept is to divide the population into several subpopulations, strata, from different properties and survey each group individually. It is often used in opinion polls when it is desired to have a survey group that is representative for the entire population (Rice, 2007). In this thesis however, geological properties will be used to define the strata.

When performing stratified random sampling there are two common approaches: proportional and optimal stratified random sampling. A proportional stratified random sampling, PSRS, scheme is used when it is desired that the strata are surveyed in accordance to their size. For example, if a poll is performed on a population with 60% women and 40% men the same percentages will be applied when designing the sampling survey (Rice, 2007).

When performing a PSRS the first step is to calculate the number of data values to be taken from each stratum, the sample fraction n_a, using equation 2.6 where the formula for the a'th stratum is given. As seen in the equation a variable W_a is used which is the weight of the stratum, i.e. the size of the stratum divided by the size of the population. The theoretical estimated variance of the mean can then be calculated as seen in equation 2.7 (Rice, 2007).

$$n_h = n\frac{N_h}{N} = nW_h \qquad (2.6)$$

$$\sigma^2_{\bar{x},p} = \frac{1}{n}\sum_{h=1}^{L} W_h \sigma_h^2 \qquad (2.7)$$

The other method that will be used is optimal stratified random sampling, OSRS. The goal for the OSRS is to find the allocation within the strata that will minimize the uncertainty in the estimated mean value for the population. It is achieved by analyzing the variance in each stratum and design the survey based on this information. As an example, if you have a population consisting of two strata of equal weight and know that the first has a higher variance then the second; the majority of the survey should be allocated to the former stratum (Rice, 2007).

Determining the OSRS sampling fraction is slightly more complicated than the proportional. As seen in equation 2.8, it is calculated by multiplying the total number of data values with the product of the weight and variance of the individual stratum divided with the sum of the products of the weight and variance for all strata. The theoretical estimated variance of the mean is then computed with equation 2.9.

$$n_h = n\frac{W_h \sigma_h}{\sum_{h=1}^{L} W_h \sigma_h} = nW_{h,o} \qquad (2.8)$$

$$\sigma^2_{\bar{x},o} = \frac{\left(\sum_{h=1}^{L} W_h \sigma_h\right)^2}{n} \qquad (2.9)$$

2.4.3 Statistical sampling simulation

Sampling simulations are methods used to analyze possible outcomes from mathematical distributions. The principle is to perform several repeated iterations with random variables taken from the distributions and analyze the outcomes. Statistical evaluations such as those described in section 2.4.1 can then be performed on the results to investigate their variability. The probability mass functions that will be used for the data are lognormal and uniform distributions, equation 2.10 and 2.11 respectively. The reasons behind using these two functions will be explained in section 5.2.3.

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-(\ln(x)-\mu)^2/(2*\sigma^2)} \qquad (2.10)$$

$$f(x) = \begin{cases} 0 & \text{for } x < a \\ \dfrac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for } x > b \end{cases} \tag{2.11}$$

2.4.4 Sensitivity Analysis

To evaluate the mathematical properties of the statistical calculations, a sensitivity analysis will be performed on the three formulas for the standard error, equations 2.5, 2.7 and 2.9. A sensitivity analysis aims to evaluate how the result of the calculation changes as the parameters increase or decrease in value. The sensitivity is calculated using equation 2.12 where a derivate for the function is calculated in respect to the parameter of interest and multiplied with the parameter through the formula. For the function in the equation, a second sensitivity to y can be calculated (Norberg, 2011):

$$Sensitivity\ to\ x = \frac{x}{f(x,y)} * \frac{\partial f}{\partial x} \tag{2.12}$$

The result will be a number that indicates how the result of *f(x,y)* changes when x is increased. A positive number will indicate that *f(x,y)* becomes larger and a negative number that it decreases, with larger value indicating a faster change. As an example, a sensitivity of 1 is a linear relationship (Norberg, 2011).

3 Reference Cases

This section will provide an overview of the five reference sites. The descriptions will give a short history of the sites activities but are mainly aimed to provide information about the contaminant transportation conditions and an overview of the geology and nature of the filling materials.

For each site a geologic profile showing the general soil stratification and groundwater table, an overview map and an image of a typical filling material can be found. The profiles and maps are not in scale and for Site E no groundwater table is marked as it varies too much to set at a certain level. A map showing the locations of the sites within Västra Götaland County can be seen in Figure 3.1. All the sites are contaminated by a variety of pollutants in addition to PAHs but no information about these will be provided.

Figure 3.1: Map with the reference sites marked (Google, 2011 with edits).

3.1 Site A: Surte 2:38, Ale

Site A is located approximately 15 km north of Gothenburg. It lies at the eastern shore of River Göta Älv and covers an area of circa 32 000 m^2, the site is shown in Figure 3.2 (Ale kommun/Norconsult, 2009). The contaminations at the site originate from a period of 30 years starting in the 1930's when the area was used as an industrial waste disposal, mainly from the wharfs in Gothenburg. After this, a boat club has been active at a portion of the site since 1987 while the remaining area is a meadow (Ale kommun/SWECO VIAK, 2007).

The natural soils at the site are mostly postglacial fine clay, mud and fluvial sediments with a thick layer of glacial clay beneath. These are overlaid by the filling materials, which in have been found to contain large amounts of waste such as oil products, scrap metal, wood, glass slag, plastics and other construction materials. The contaminations are spread over the entire site and have a magnitude of 0.5 to 2.5 meters, with high contamination levels over the entire area but several peak points (Ale kommun/Norconsult, 2009).

Since the area has a low topography and is located close to the river, the groundwater table can be found close to the ground surface. The materials at the site have a high hydraulic conductivity, and the river level regulates the groundwater table. The result is that a large portion of the contaminated volume is below the water table. The direction of groundwater flow is towards River Göta Älv (Ale kommun/SWECO VIAK, 2007).

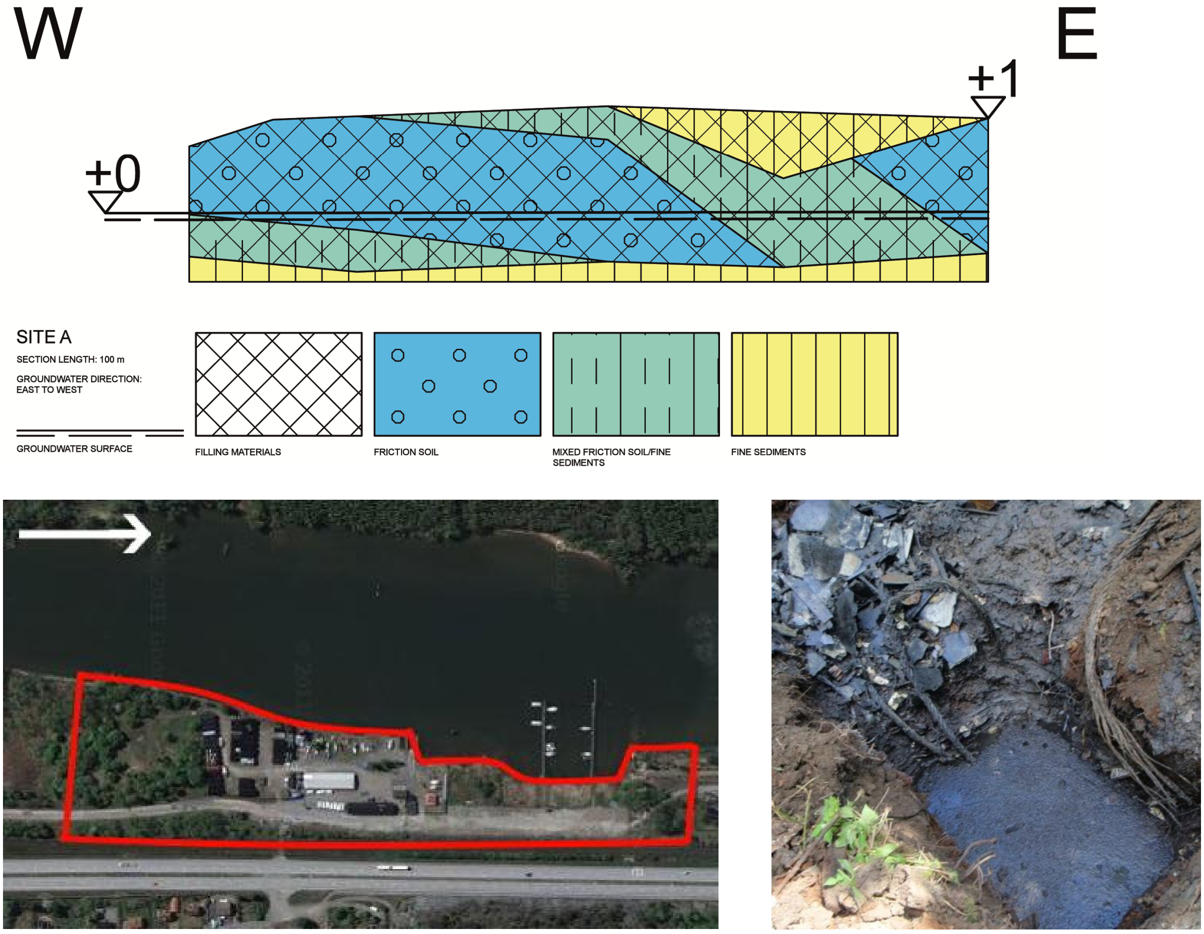


Figure 3.2: Geologic profile, overview and filling materials for Site A. Image © Norconsult.

3.2 Site B: EKA Chemicals, Bengtsfors

Site B is located in Bengtsfors Municipality 170 km north of Gothenburg. The area is circa 30 000 m^2 and lies at the outskirts of the town at the shore of lake Bengtsbrohöljen. There is a long history of hazardous activities at the site, starting with a chlor-alkali factory that was active between1897 and 1923. Thereafter several other polluting industries have followed such as wood preservation, sawmill and dry cleaning. Today the site has been remediated and is occupied by several light industries (Bengtsfors kommun/Geo Innova AB, 2003b). The data on the site is from the pre-remediation assessments.

Exploratory drillings before the remediation showed that the geology in the area consisted of a bottom layer of till followed by sorted gravel and sand as well as peat. Materials consisting of waste from industries such as bark, sludge, metal- and sawdust combined with boulders had been used to fill up parts of the site. These fillings were found to be up to seven meters thick (Bengtsfors kommun/Geo Innova AB, 2003a).

The groundwater table is closely connected to the water level in the lake. The direction of the groundwater flow is towards the closest shore, see map in Figure 3.3. The groundwater table varies, but is generally 1 to 2 meters below the ground surface, which means that the fillings were partially saturated (Bengtsfors kommun/Geo Innova AB, 2003a).

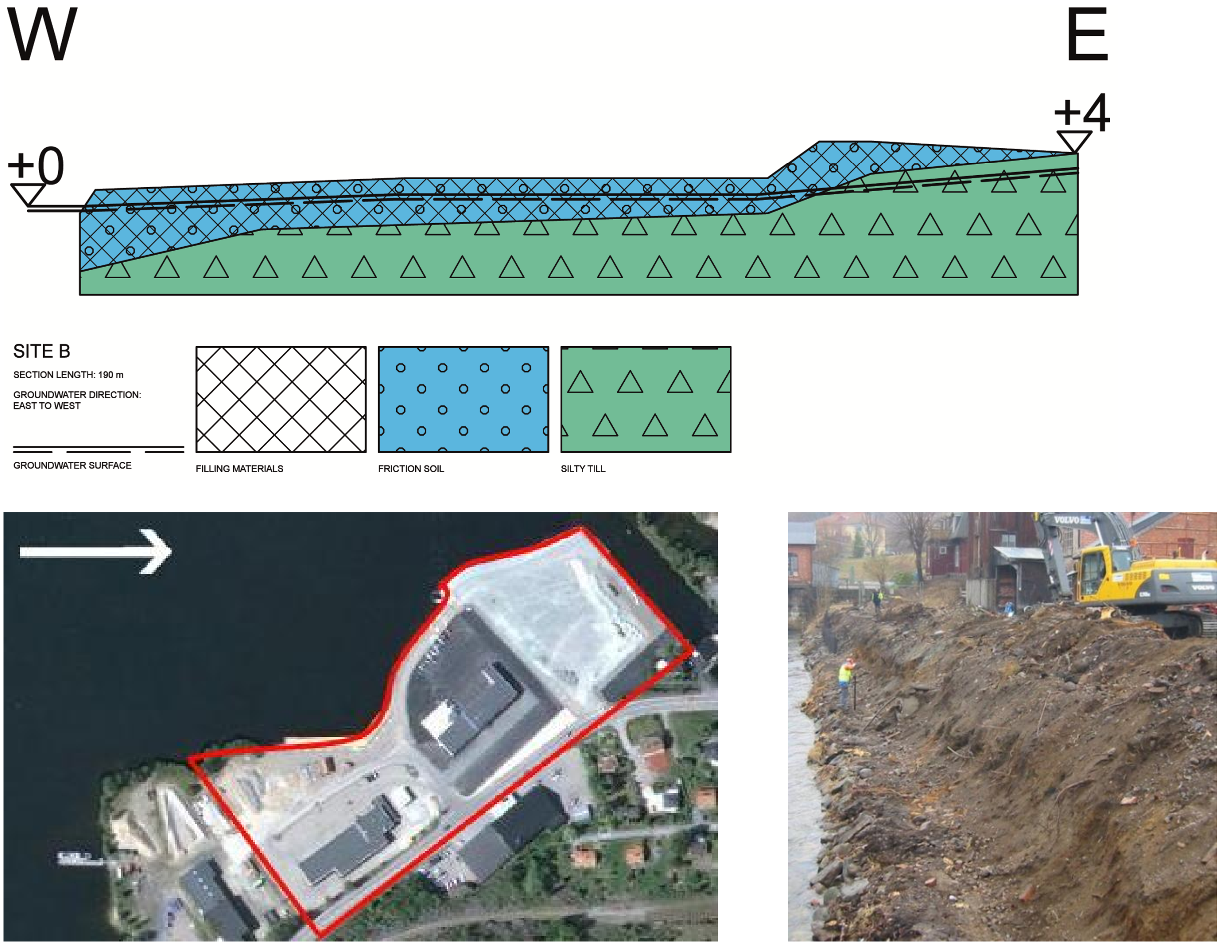


Figure 3.3: Geologic profile, overview and filling materials for Site B. Image © Norconsult.

3.3 Site C: Göta Älv bridge, Gothenburg

The site is located in central Gothenburg and covers an area of 360 000 m^2, seen in Figure 3.3. The data is from a survey that was made in the vicinity of the current Göta Älv Bridge as a part of the pre-study for a new bridge. The site was originally marshes but at the middle of the 19th century it was filled up to enable the city to expand. At first, the filling materials were mostly dredged from the harbor but overtime the amount of other waste increased. The source of this waste is hard to determine, but it is often contaminated (Göteborgs stad/Norconsult AB, 2011). Today the area is a part of the inner city and contains a variety of activities including a highway, train station and several smaller industries.

The natural soils consist of fluvial sediments overlying clay, often with lacustrine sediments in between. Above these there is a layer of filling materials which consists of dredging materials, such as silty clay, in the bottom and then up to three meters of coarser materials containing bricks, glass, wood, slag-products and construction waste on top.

As with Site A, this location has a low topography. Since the geology is mostly clay, the groundwater table will follow the topography and is found 1 to 2 meters below the ground surface. The soil has a low permeability, which contributes to a low groundwater flow (Geological Survey of Sweden, 2011).

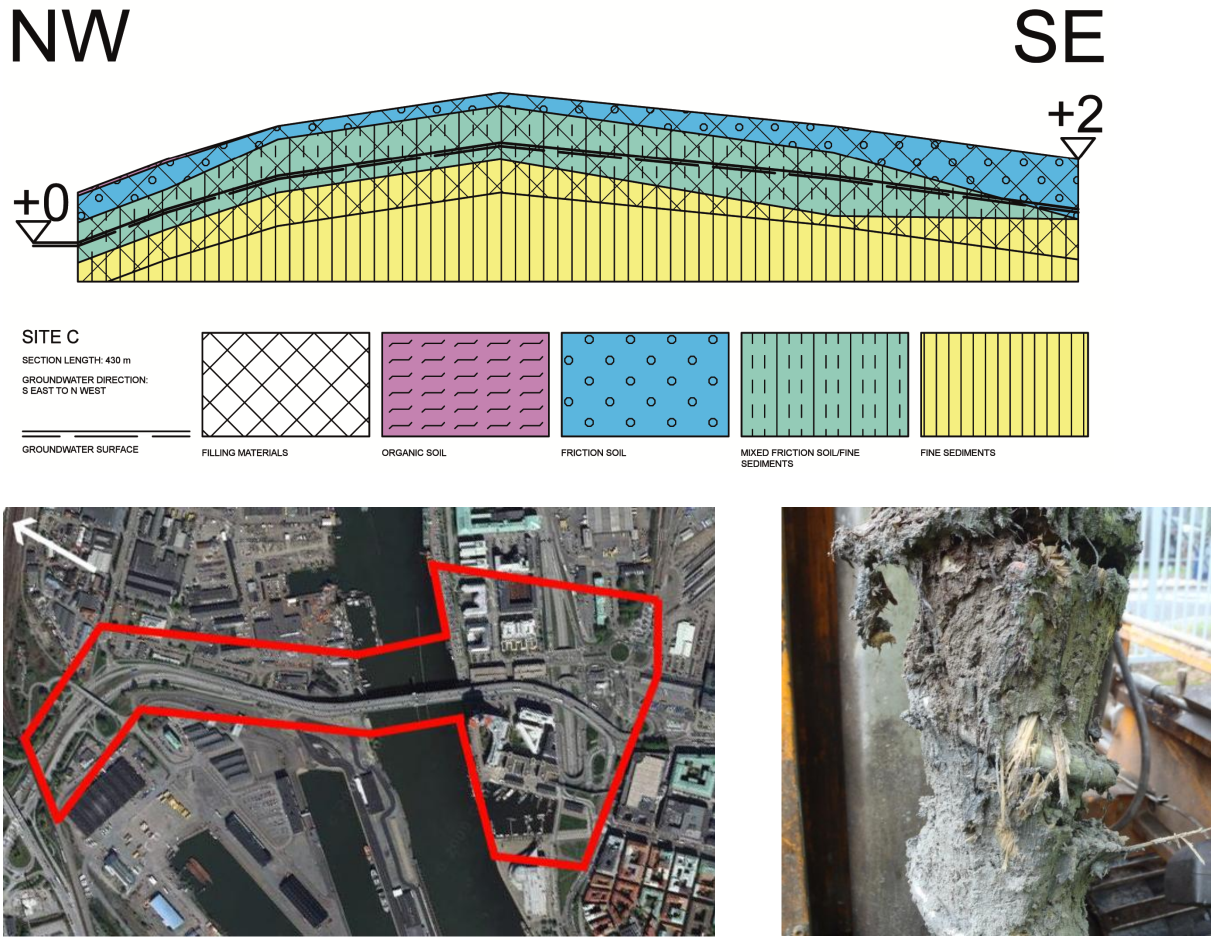


Figure 3.4: Geologic profile, overview and filling materials for Site C. Image © Norconsult.

3.4 Site D: Kvillebäcken, Gothenburg

Site D is a 100 000 m^2 large area in the central part of Gothenburg, at Hisingen. Different types of industries have occupied the area for a long time, which has contributed to the contaminations at the site. In addition, the ground level has been adjusted by means of filling materials similar to those at Site C, which may have been contaminated (Norra Älvstranden Utveckling AB/SWECO, 2009). The area is currently being developed into a residential area by several different companies.

The soil stratigraphy at the site consists of a thin, approximately 0.5 m thick, layer of organic soils. Beneath these there is a layer of filling materials with a thickness varying between 1 and 2.5 meters. The filling materials contain a variety of materials such as slag, wood, bricks and concrete. The fillings are underlain by a clay layer with a magnitude of 30 to 40 meters (Norra Älvstranden Utveckling AB/SWECO, 2009).

The site is adjacent in the east to the stream Kvillebäcken, which runs across Hisingen Island and has its outlet in River Göta Älv see Figure 3.5. The groundwater flow in the area is considered to be directed towards the stream. The filling materials are partly below the groundwater table (Norra Älvstranden Utveckling AB/SWECO, 2009).

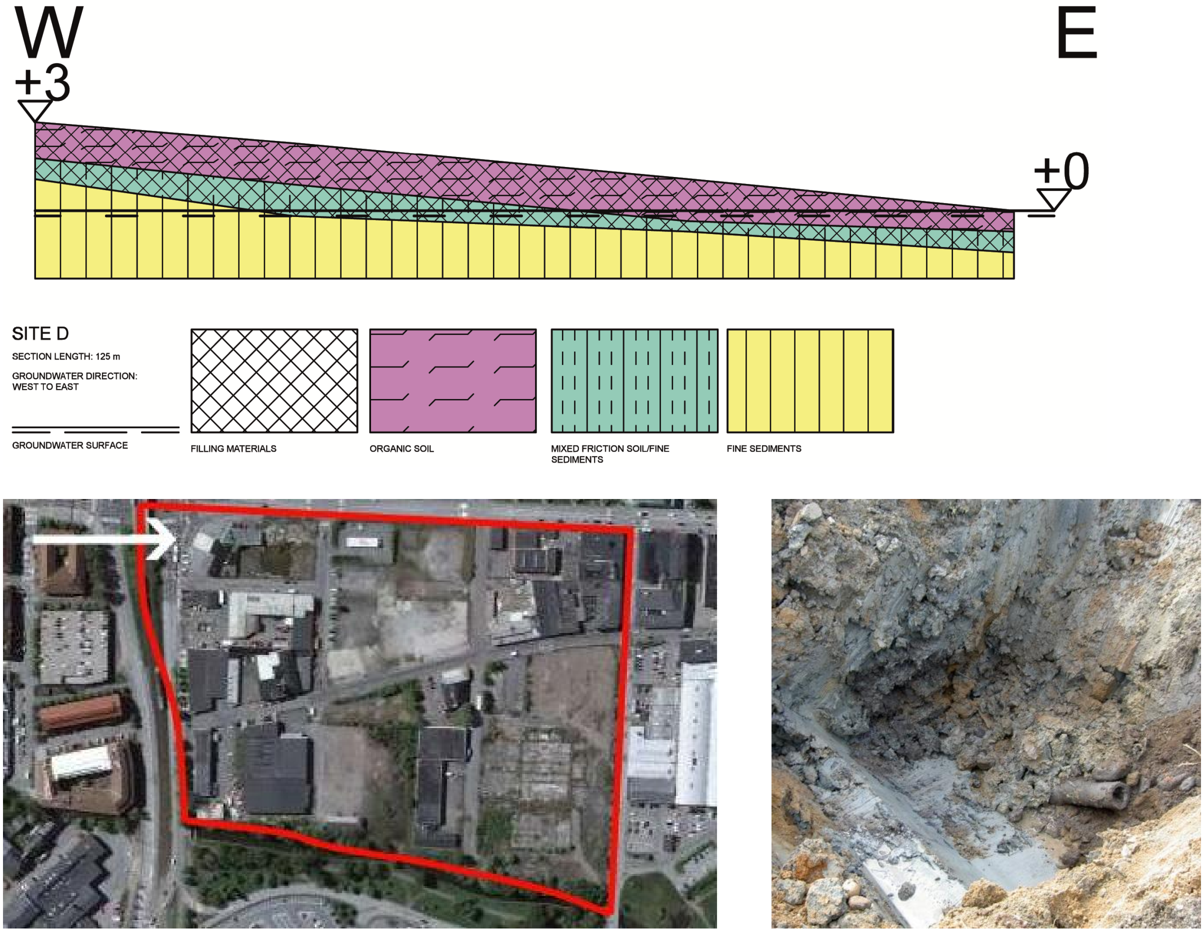


Figure 3.5: Geologic profile, overview and filling materials for Site D. Image © NCC.

3.5 Site E: Hexion, Mölndal

Site E is located in the city of Mölndal, 9 km south of Gothenburg city center and is approximately 45 000 m^2, see Figure 3.6. The site has been industrialized for a long time, but the activities contributing most to the contamination started in 1827 when a company producing flaxseed oil and paint established. Over time the production turned towards other chemicals, and when the last owner closed the factory in 2007 the main products were alkyd and polyester. The area is presently being converted into a residential area by NCC (Mölndals stad, 2010).

The site is a part of the Gothenburg moraine, which is a recessional moraine, and the geology is composed of sand, gravel and till as wells as sections of fine sediments. Bedrock can generally be found at depths of 5 to 15 meters. Parts of the area have been filled up with anthropogenic soils with a thickness of up to 5 meters (NCC Boende AB/SWECO, 2009). As with the other sites, the filling materials contain several anthropogenic substances such as bricks, concrete, scrap metal and coal slag[1].

Site E is located in a western facing slope and due to the mixed hydraulic conditions of the geology the water pressure varies over the area. This leads to large variations in the water table, between 2 and 8 meters below the ground surface, and the fillings are partially saturated.

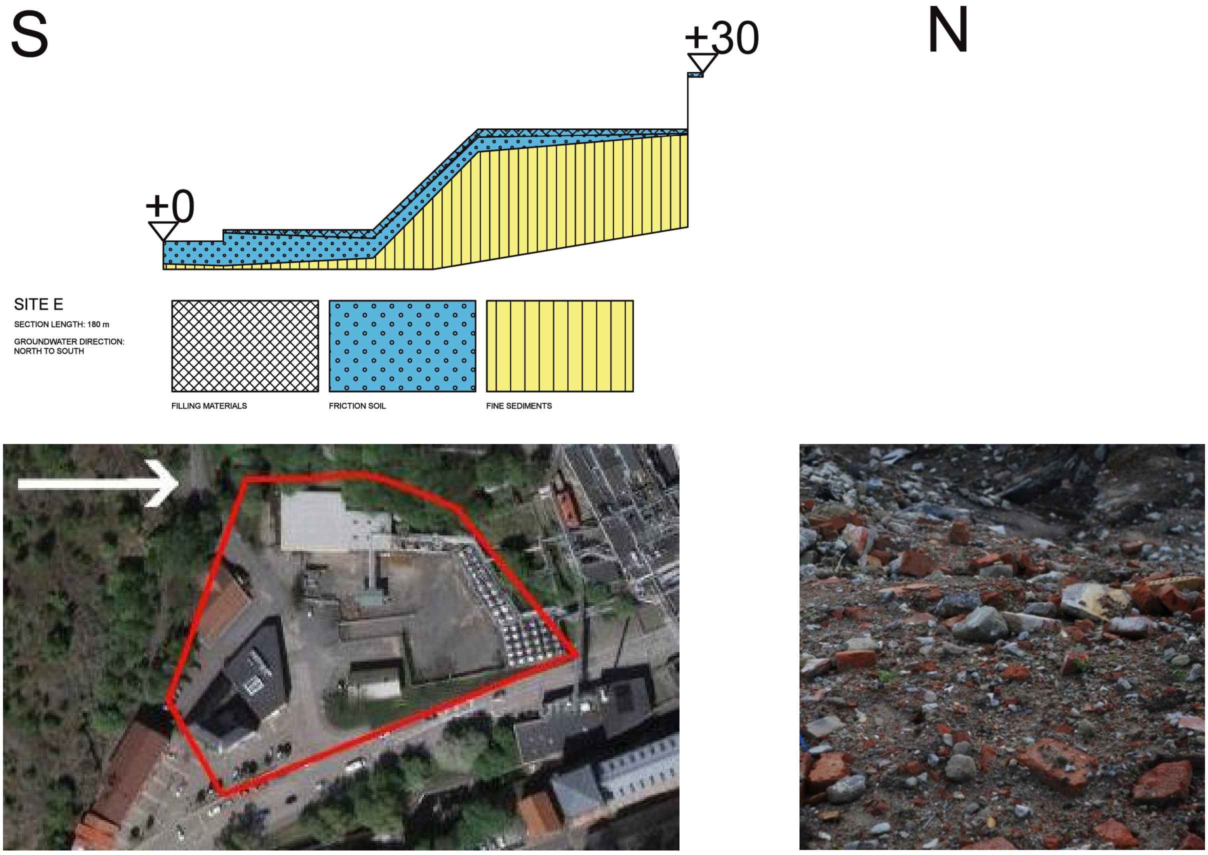


Figure 3.6: Geologic profile, overview and filling materials for Site E. Image © A. Johansson.

---

[1] Malin Norin NCC Teknik, Renare Mark Symposium: Studiebesök Hexion, 13 October 2011

3.6 Overview of site conditions

The basic conditions for all the sites are described in Table 3.1 below. As can be seen, two of the sites have contamination sources with a point release, such as a factory, while the remaining three have more diffuse contamination sources due to the use of contaminated materials as fillings. The migration potential is classified according to the Swedish EPA guideline (2000, pp. 31-43), for both transportation rate in the ground and the average time for the groundwater to reach surface waters based on the average conductivity of the soil and distance to surface water. While this classification is simple, for example it only considers one transportation mechanism and is commonly based on tabulated values and not site soundings, it can be used as a rough estimate of the migration potential. The migration potential for Site A was estimated by GF Konsult (2006) while the authors, aided by site surveys, have classified the remaining sites.

Table 3.1: Summary of the conditions at each reference site. The migration potential for Site A (GF Konsult, 2006), Site B to E has been classified by the authors.

Area	**Source**	**Size**	**Migration potential**	
			In soil and groundwater	*To surface water*
Site A	Diffuse	3 ha	Very Large	Large
Site B	Diffuse/Point	3 ha	Very Large	Large
Site C	Diffuse	36 ha	Moderate	Slight
Site D	Diffuse	10 ha	Moderate	Slight
Site E	Point	5 ha	Large	Moderate

4 Soil Classification

This chapter describes the soil classification system that has been developed for the filling materials. The classification of grain sizes will be adapted to the guideline from the Swedish Geotechnical Society, which is designed to be compatible with international standards (Karlsson & Hansbo, 1984). *The standardized Swedish soil fraction classes for natural soils, the texture triangle used for classification and explanation of abbreviations can be found in Appendix 2.*

In the following sections the five different grain size classes will be introduced and the motive behind the classification will be provided. Last in this section there is also a sub-section on how organic soils have been treated during classification. The different soil classes are exemplified with grain size distributions in Figure 4.1. All the percentages are in respect to the dry weight of the fractions below 20 mm of the soil sample.

4.1 Class I: Waste

Class I includes all soil samples that are mainly composed of materials other than natural soil. These are most common in areas which have been used as landfills, for example study Site A. Due to the mixed nature of the materials, the properties of the soils in this class are highly variable, and for example the total organic carbon content can differ greatly. Note that no grain size distributions can be determined for this class.

4.2 Class II: Friction Soils

Class II comprises the grain sizes known as friction materials, which includes boulders, stones, gravel and sand. If this is compared to the grain size distribution guideline, this mean that all grain sizes above 0.063 mm falls in this category (Karlsson & Hansbo, 1984). Fine fractions can be included in the sample if the content is below 15%. Friction soils are commonly used as filling material for construction purposes due to their technical properties with high conductivity and low susceptibility to frost heave. The high hydraulic conductivity also normally results in a fast migration of contaminants from the materials (Burden & Sims, 1999).

4.3 Class III: Friction soils with fine sediments

If a soil sample consists of mostly friction soils but in addition features a higher content of fine sediments, it will be categorized as Class III. For a soil to be in Class III, the fine fraction (< 0.063 mm) should amount to between 15% and 40% (Karlsson & Hansbo, 1984). As fine sediments have a much lower conductivity than the friction materials their presence in a soil sample will alter the transportation conditions. It is also common for pollutants to attach more to the fine fractions, especially to clay (Stevens, 2010).

4.4 Class IV: Fine sediments with friction soils

If the content of clay or silt is above 40%, the fine sediment will be the primary grain size. Soils containing above 40% but less than 80% of fine sediment will be categorized as Class IV. Soil types in Class III and Class IV are not typically used in technical applications, but they have historically been used to alter the topography (Göteborgs stad/Norconsult AB, 2011, pp. 8-11).

4.5 Class V: Fine sediments

If the friction material content is below 20%, these will not be included in the name of the soil sample and will be categorized as Class V, which includes silt and clay, in other words fine sediments with grain sizes below 0.063 mm (Karlsson & Hansbo, 1984). As mentioned before, fine sediments have a very low conductivity and can even have capillary properties. This results in a slow groundwater flow through the soil, which results in a slow contaminant transportation.

4.6 Organic content

As several soil samples contain not only mineral soil but also has an organic content, it should be clarified how this is considered during classification. If the organic part is not the primary soil type, it will not be considered during the classification. If the organic soil is the primary soil type however, the soil is classified as Class III as the organic soils share some of the characteristics of clay with lower hydraulic conductivity (Swedish EPA, 2000) and increased tendency to bind contaminants, especially PAHs (Bernes, 1998, p. 35).

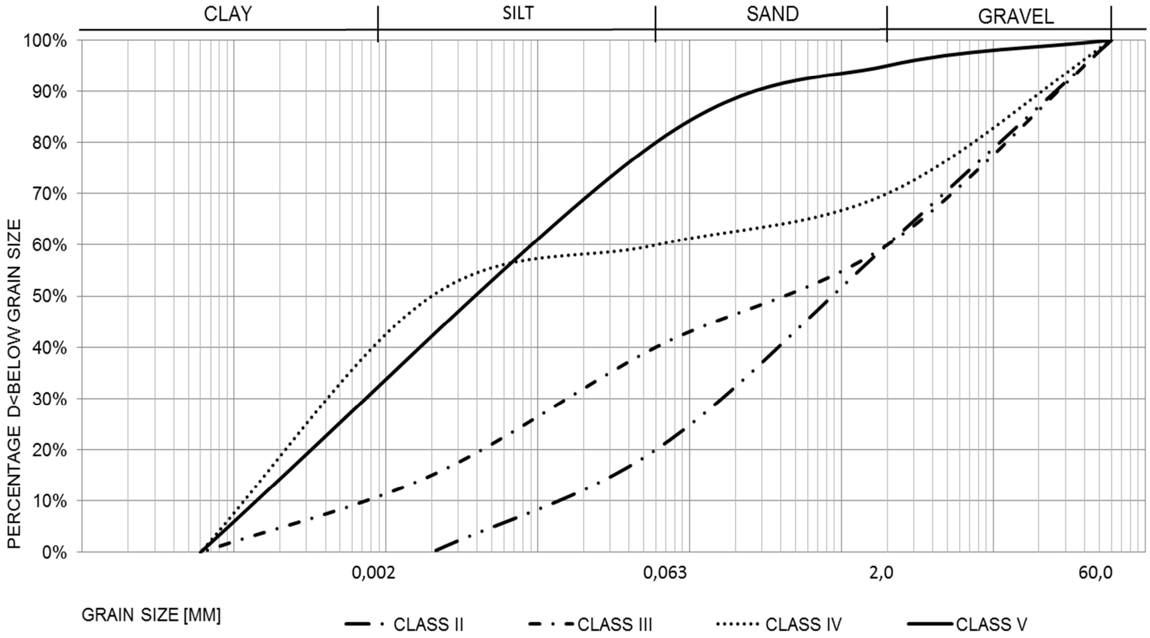


Figure 4.1: Examples of grain size distributions for Class II, III, IV and V. As Class I consist of materials that are not natural soil, there are no grain size distributions for this class.

5 Analysis

This chapter describes the calculations underlying the analyses and results. The focus will be on the calculation steps and not actual numbers. The aim is to give an overview of the work process, seen in Figure 5.1. The chapter is divided into sections for the work steps that have been used in the analysis for the thesis.

5.1 Soil type determination

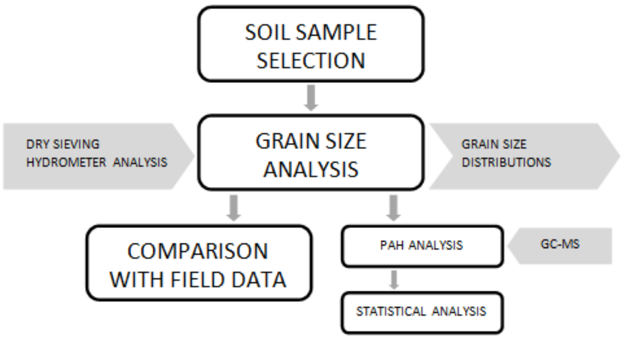


Figure 5.1: Work process for the soil type determination.

As mentioned before, the grain size classification for the soil samples will be based on the ocular analysis that was performed by field technicians during the soil survey. Since this is done without any scientific methods grain size distributions for each site will be determined by dry sieving and hydrometer analysis and will be compared to the ocular analysis to verify their accuracy.

Since dry sieving will bed used, the soil can form aggregates during drying which will make the sieving unreliable. Under those circumstances only hydrometer analysis will be performed, which will mean that the proportion between gravel and sand will be unknown.

The problem with aggregation while drying would not occur if wet sieving was used instead of dry sieving. On the other hand, wet sieving can break the material instead, especially in this case when the material consists of anthropogenic soil. In addition, the wet sieving produces a contaminated wastewater which can not be treated at the laboratory used for the analysis and these factors makes dry sieving a better alternative.

If the fine sediment content is less than 15 % no hydrometer analysis will be performed since it will not be necessary for the determination of the samples name or class using the soil texture triangle.

When both dry sieving and hydrometer analysis is preformed the hydrometer analysis will be considered more reliable for the correct amount of fine sediment and the grain size distribution will be re-calculated if the fine sediment content in the two analyses is not equal. As an example; if the fine sediment is 10 g heavier according to the hydrometer analysis than the dry sieving, the extra grams will be added to the fine sediment content from the dry sieving and the difference subtracted evenly from the other grain sizes before drawing the grain size distribution.

If wet sieving could be carried out, this problem would not occur but since the dry sieving will not separate smaller fractions that have adhered to the larger, this crude correction is carried out. The actual distribution of smaller particles is likely more complex than this, but the simple method is used rather than making assumptions regarding unknown adherence factors.

After the grain size distributions have been drawn both the name and class of the soil sample will be compared with the result from the soil survey. As mentioned before, PAH analysis is performed on the sieved samples where the PAH content is unknown.

As an additional analysis for Site E, the results from the laboratory sieving will be compared with coarser sieving performed at the site during remediation. The aim of this comparison will be to estimate the amount of material that is not included in the field survey as it is too large to be extracted during drilling.

5.2 Sampling scheme analysis

The work process for the sampling scheme analysis is presented in Figure 5.2.

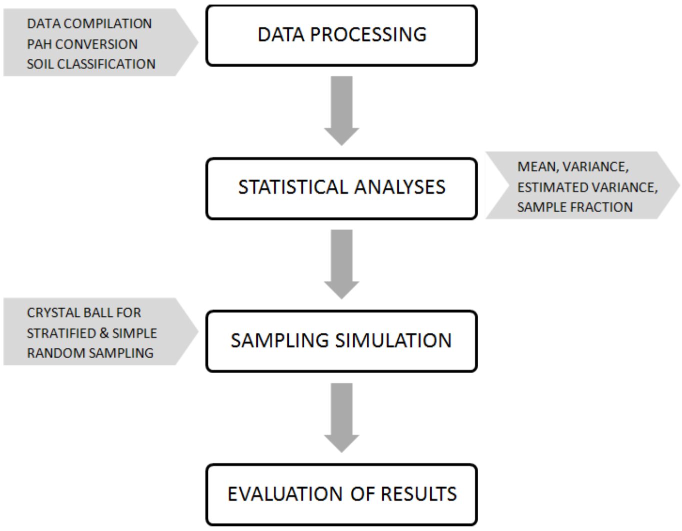


Figure 5.2: Work process for the sampling scheme analysis, which is the main part of the thesis.

5.2.1 Data processing

The first step in the conceptual model is the data processing. The main task here is to compile the data and prepare for further calculations. The methodology for soil classification is described in chapter 4 and the PAH-conversion in Appendix 1.

5.2.2 Statistical analyses

As a first step to investigate the properties of the standard error equations, a sensitivity analysis will be performed. For the simple random sampling, the equations described in section 2.4.4 will be used. For the stratified random sampling, where the standard error equations contain sums, it becomes more difficult. A derivate of a sum cannot be calculated, and therefore a numerical sensitivity analysis will be performed. This is done by altering the parameters in the formula and studying the change. As the error for the stratified sampling schemes are dependent on the individual standard errors for the strata, the sensitivity analysis will be aimed at evaluating how the standard error changes depending on the magnitude of the difference between the strata.

As mentioned before, the basic statistical parameters described in section 2.4.1 will be calculated as a first statistical analysis of the sites. The results from this analysis will be used in further calculations as well as analyzed as a first assessment of the conditions at the sites. The mean value, coefficient of variation as well as the other basic statistical parameters will be presented further on.

When the basic statistical analysis has been performed, the next step will be the stratified random sampling. The goal is to find a method that will optimize the number of samples sent for chemical analysis, the sample fraction n_m, based on the contamination levels in each site aided by equation 2.6 and 2.8 for proportional and optimal stratified random sampling respectively.

As a first evaluation of the calculated sampling fractions a standard error will be calculated for three different sample designs: simple random sampling (SRS), proportional random sampling (PSRS) and optimal random sampling (OSRS). Simple random sampling is a basic sampling scheme where a value is randomly selected with no consideration of strata and can be compared with tossing a dice. It is used as a reference method to compare the stratified sampling with a strictly random one.

The standard error $\boxed{?}_{\boxed{?}}$ is calculated using the equations described in chapter 2.4, and will be calculated for several sample sizes, starting at $n_T = 5$ up to n of the original soil survey in increments of 5. The results will be visualized in graphs, with $\boxed{?}_{\boxed{?}}$ on the y-axis and n_T on the x-axis. To evaluate the difference between the different standard errors, the quotient between the standard error for simple random sampling and proportional and optimal stratified random sampling respectively will be calculated.

5.2.3 Sampling simulation

To verify the results from the calculation of the theoretical standard error for the different sampling schemes, two sampling simulations will be performed. The theoretical standard error cannot be used to analyze if there is a correlation between the soil class and contamination level, only to estimate how much lower the standard error will be if there is a correlation. The sampling simulation is required to see if there actually is a correlation.

The first simulation will be to sample the data values from the original survey 10000 times with different sample sizes, n_T, starting at 5 and up to the number of data values in the original soil survey. The simulations will be performed with replacement, which means that the same data value can be chosen several times per iteration.

During the data simulations, the uniform distribution from section 2.4.3 will be applied. The uniform distribution will be used to randomly select one of the values in the survey data. For SRS, each soil sample in the survey will have an equal probability of being sampled and for the stratified random sampling methods every value in a particular stratum will have the same probability of being sampled.

The second method that will be used to analyze the data is a Monte Carlo simulation, where the mean and standard deviation calculated from the soil survey data is used to estimate a mass function for each soil class, and then sample from the mass function. This will be performed with a starting sample size of 10 and then in steps of 10 up to 60.

In the probabilistic simulation the lognormal distribution in section 2.4.3 will be used. The reason for using the lognormal mass function is due to an analysis of histograms for the data, which indicated that the data had a lognormal distribution for all sites. See appendix 3 for histograms for all sites and molecular weights. Strictly positive numbers, a high density of samples at low values and a few outliers with high values characterize the lognormal distribution[1].

For both the random sampling procedures, a mean value will be calculated for each sample size and the mean value and standard deviation for the randomized mean will be used to calculate the lognormal confidence levels of the mean from the survey being covered in the sample. As the calculation of a lognormal confidence is too complicated to perform by hand, it will be calculated using computer software. The results will be presented in graphs where the confidence level is plotted against the sample size n_T. As with the theoretical standard error, confidence levels will be drawn for SRS, PSRS and OSRS for both sampling simulations and each molecular weight respectively.

In addition to the confidence level diagrams, an additional comparison will be performed for the data sampling simulation. It will consist of a table like the example in Table 5.1, where the mean confidence level for the different sampling schemes will be divided with each other. This is only performed for the data sampling simulation due to the larger amount of data for this simulation. This comparison will give a direct indication on how much better a sampling scheme is compared to another.

Table 5.1: Example of a comparison between contamination levels.

	SRS	PSRS	OSRS
SRS	NA	=SRS/PSRS	=SRS/OSRS
PSRS		NA	=PSRS/OSRS
OSRS			NA

5.2.4 Evaluation of results

The results from the sampling scheme analysis will be evaluated in two steps. First, each site will be evaluated individually and conclusions and observations for each site will be described. In the next step, the results from all sites will be studied to find general results that are shared by several or all sites. The aim is to find similarities that can be used to give recommendations on an optimized sample design for filling materials.

There are some important aspects that will be studied during the evaluation of the results. First, as the goal with the thesis is to study the correlation between soil class and contamination level, signs of this will be analyzed. The criteria for determining if there is a correlation or not will be if the results from the standard error, data sampling and probabilistic sampling are similar. If the stratification methods are well suited for all analysis steps, the

---

[1] Tommy Norberg, Docent in mathematical statistics, Chalmers University of Technology, lecture 2011-01-25

conclusion will be that there is a correlation. If one of the analysis steps has a deviating result, the analysis will be that there is no correlation.

During the analysis of the confidence level, the aim will be to find out if fewer samples could have been analyzed for the sites and if one of the sampling schemes would require fewer samples than the others. An important number of interests will be at how many samples a 100 % confidence is reached, but as 100 % confidence seldom is required lower confidence levels will also be of interest. The main focus will be on a 95 % confidence, which is a commonly used confidence limit during analysis of contaminated sites (Norrman, Jenny; Purucker, Tom; Back, Pär-Erik; Engelke, Fredric; Stewart, Robert, 2009).

5.3 TOC analysis

As mentioned in section 2.3 PAHs tend to bind to organic materials. One assumption made when performing the statistical analysis is that the correlation between soil classes and contamination level is the essential one rather than the correlation with TOC. To verify this assumption an analysis on the correlation between TOC and contamination levels for some of the sites will be performed as well.

The first step is to evaluate if there is any correlation between the two parameters, which is done by plotting them against each other with TOC content on the x-axis and contamination level on the y-axis. This will be performed for the three sites where data on TOC is available; Site A, B and C. The amount of data on Site B and C is insufficient for any further analysis due to a low number of data values and low spread in TOC content.

For Site A the theoretical standard error will be calculated as well, but instead of using soil classes like in the previous analysis the stratification will be based on TOC content. The quotient between SRS and PSRS as well as between SRS and OSRS will be calculated in the same way as for the soil class analysis. Since the standard error for SRS is independent of the stratification, it will be the same for the TOC analysis as for the soil class analysis. This means that the quotients will be comparable between the two analysis methods and conclusions on which type of analysis that is more efficient can be drawn.

6 Results

The results are divided into three sections. The first presents the outcome of the soil type determination, the second the statistical analysis of the correlation between contamination level and grain size and the third the results from the comparison between contamination level and TOC content.

6.1 Results, Soil Type Determination

As has been described before, the aim of this part of the study is to compare the field classification of the soil samples with the grain size distribution as determined by dry sieving. The grain size distributions can be found in Appendix 3. The correctness of both the name and the soil class will be analyzed, but the correctness of the soil class is the most important as it will affect the result of the sampling scheme analysis.

6.1.1 Results for Soil Type Determination, Site A

For Site A 10 grain size distributions have been determined and compared to the field classifications in Table 6.1, with the ones that were incorrectly classified in the field marked with bold. The definitions are incorrect for all sieved samples. It should be noted that according to the field classification many of the samples contain stone, i.e. particles larger than 60 mm. These are not included in the sieved samples, and may be the cause for several of the naming errors. When comparing the class for the sieved samples with the class from the field determination, the result is that 2 of the samples are classified incorrectly.

Table 6.1: Comparison of field classification and results from sieving for Site A.

Sample ID	*Soil type, field*	*Soil class, field*	*Soil type, sieving*	*Soil class, sieving*
Site A: 1	F(grSa)	2	**F(Sa)**	2
Site A: 2	F(grstSa)	2	**F(grSa)**	2
Site A: 3	F(grstSa)	2	**F(grSa)**	2
Site A: 4	F(grstSa)	2	**F(saGr)**	2
Site A: 5	F(grstclSa)	3	**F(clgrSa)**	3
Site A: 6	F(clgrSa)	3	**F(grSa)**	**2**
Site A: 7	F(clGr)	3	**F(grSa)**	**2**
Site A: 8	(Cl)	5	**(siCl)**	5
Site A: 9	(orCl)	5	**(siCl)**	5
Site A: 10	(orCl)	5	**(siCl)**	5

6.1.2 Results for Soil Type Determination, Site B

For Site B, the field classifications have been compared with grain size distributions for 12 soil samples. The result from the comparison, as seen Table 6.2, is that the field determination is entirely correct for 6 of the soil samples and incorrect for the other 6. The classification is affected for 2 of these samples, where it was classified as Class III but the sieving indicated Class II for 1 soil sample and vice versa for another.

Table 6.2: Comparison of field classification and results from sieving for Site B.

Sample ID	*Soil type, field*	*Soil class, field*	*Soil type, sieving*	*Soil class, sieving*
Site B: 1	F(grSa)	2	F(grSa)	2
Site B: 2	F(grSa)	2	**F(saGr)**	2
Site B: 3	F(grSa)	2	F(grSa)	2
Site B: 4	F(grSa)	2	**F(saGr)**	2
Site B: 5	F(grSa)	2	F(grSa)	2
Site B: 6	F(grSa)	2	**F(sigrSa)**	**3**
Site B: 7	F(grSa)	2	F(grSa)	2
Site B: 8	F(grSa)	2	F(grSa)	2
Site B: 9	F(grSa)	2	**F(saGr)**	2
Site B: 10	F(grSa)	3	**F(saGr)**	2
Site B: 11	F(grSa)	2	F(grSa)	2
Site B: 12	F(clgrsiSa)	3	**F(grSa)**	**2**

6.1.3 Results for Soil Type Determination, Site C

For Site C at least 6 out of the 10 compared soil samples had the correct soil class in comparison to the field determination, see Table 6.3. However, only 1 of the soil samples has the entire correct definition. For 4 of the samples with high content fine sediments the exact name could not be determined since only hydrometer analysis could be performed due to aggregation. For 2 of these samples the uncertainty only affects the name and not the class while the other 2 samples will either belong to Class IV or V depending on the proportion between gravel and sand.

Table 6.3: Comparison of field classification and results from sieving for Site C.

Sample ID	*Soil type, field*	*Soil class, field*	*Soil type, sieving*	*Soil class, sieving*
Site C: 1	F(grSa)	2	**F(saGr)**	2
Site C: 2	F(saCl)	4	**F(xCl)**	4
Site C: 3	F(Sa)	2	F(Sa)	2
Site C: 4	F(grsaCl)	4	**F(xCl)**	4
Site C: 5	F(grsiSa)	3	**F(saGr)**	**2**
Site C: 6	F(clgrSa)	3	**F(grclSa)**	3
Site C: 7	F(clgrSa)	3	**F(grsacl S)**	3
Site C: 8	(siCl)	5	**F(xSi)**	**4**/5
Site C: 9	(saCl)	5	**F(sasiGr)**	**3**
Site C: 10	(siCl)	5	**F(xSi)**	**4**/5

6.1.4 Results for Soil Type Determination, Site E

For Site E, only 2 of the soil samples had different names after the sieving compared to the field classification, see Table 6.4. For 1 of these, the class was incorrect as a soil sample that would have been categorized as Class III based on the field classification was in reality a Class II sample. For this sample however, the fine fraction was only 0.2% lower than the limit for Class III.

For Site E the grain size distributions have been compared with data from sieving at site as well, see Appendix 4, where the entire soil volume was sieved and fractions larger than 20 mm was included. This comparison indicated that on average approximately 25% of the total weight was above 20 mm. If these fractions were included in the field classification, they would not affect the classification of the soil for this site, but likely the name.

Table 6.4: Comparison of field classification and results from sieving for Site E.

Sample ID	*Soil type, field*	*Soil class, field*	*Soil type, sieving*	*Soil class, sieving*
Site E: 1	F(grSa)	2	F(grSa)	2
Site E: 2	F(grSa)	2	F(grSa)	2
Site E: 3	F(grSa)	2	F(grSa)	2
Site E: 4	F(Sa)	2	**F(grSa)**	2
Site E: 5	F(grSa)	2	F(grSa)	2
Site E: 6	F(Sa)	2	F(Sa)	2
Site E: 7	F(grSa)	2	F(grSa)	2
Site E: 8	F(Sa)	2	F(Sa)	2
Site E: 9	F(Sa)	2	F(Sa)	2
Site E: 10	F(grsiclSa)	3	**F(grSa)**	**2**

6.1.5 Results for Soil Type Determination, Combination

Looking at the results from all the sites in Table 6.5 a majority, approximately 80%, of the samples is categorized in to the correct soil class based on the field determination. The number of samples where the name is also correct is significantly lower; approximately 33% of the soil samples were given the correct name in the field survey.

Table 6.5: Summary of the results from the grain size analysis for all the sites.

Site	*Sieved samples, ∑*	*Correct name, ∑*	*Correct class, ∑*	*Correct class, %*
Site A	10	0	8	80
Site B	12	6	10	83,3
Site C	10	1	6 – 8	60 – 80
Site E	10	8	9	90
Total	42	15	33 – 35	78,6 – 83,3

6.2 Results, Sampling Scheme Analysis

The results from the sampling scheme analysis will start with the results from the sensitivity analysis for the standard error, for additional calculation data see Appendix 6. Then the results from the statistical analysis, comparing the three different sampling schemes for each site will be presented individually. All calculated parameters underlying the results for Section 6.2.2 to 6.2.6 can be seen in Appendix 7, the diagrams for the standard error in Appendix 8 and the data from the sampling simulations in Appendix 9.

6.2.1 Results, Sensitivity Analysis

The result from the calculated sensitivity for the SRS standard error, seen in Table 6.6 indicates that if the sample size is increased the error decreases but if the standard deviation is increased the error is increased as well. In other words, the SRS standard error will be low for a survey where the soil samples have similar values or for a survey with a large sample size.

For the stratified sampling schemes, no sensitivity for the standard error could be calculated. The numerical sensitivity analysis indicates that the standard error for both proportional and optimal stratified random sampling decreases if the sample size is increased.

The error for proportional sampling decreases if the standard deviation for the strata has similar values or if the size of the strata is equal. For the optimal sampling the result indicated that it has a decreased standard error compared to the proportional sampling if there is a large difference between the values of the standard deviation in each stratum, especially when one small stratum has a significantly higher standard deviation. This is in line with what is stated in Rice (2007), i.e. the proportional sampling is most efficient for strata where the mean value is variable while the optimal sampling is more efficient if the standard deviation is variable.

Table 6.6: Result from the sensitivity analysis on the standard error for simple random sampling.

Parameter	n	S
[illegible]	$-n/2$	1

6.2.2 Result for Sampling Scheme Analysis, Site A

As seen in Table 6.7, the soil samples for Site A are spread among all the soil classes with a majority of them in Class I and II. Class II is in addition to being the class with most samples also the class with the highest mean and coefficient of variation for all molecular weights. Class I and III have mean values of almost the same magnitude as Class II, but with significantly lower coefficient of variation. While having a comparatively low mean, the coefficient of variation for Class V is higher than all other except for Class II.

It should be noted that the concentrations for middle and high molecular weight are above the guideline values for less sensitive land use, which are 20 and 10 mg/kg dry substance respectively, but not even above the guideline for sensitive land use, 3 mg/kg, for low molecular weight.

Table 6.7: Summary statistics for Site A.

STATISTICS	n_K	[illegible]	[illegible]	[illegible]	$C_{x,L}$	$C_{x,M}$	$C_{x,H}$
CLASS I	39	1,1	17,2	27,0	1,4	1,7	1,5
CLASS II	44	2,2	27,7	38,3	4,0	3,3	3,0
CLASS III	20	2,1	21,2	28,8	1,4	1,5	1,9
CLASS IV	10	0,5	6,3	7,7	0,9	1,1	1,2
CLASS V	7	0,5	10,5	13,5	2,2	2,3	2,4
ENTIRE AREA	*120*	*1,6*	*20,4*	*29,1*	*3,4*	*2,9*	*2,6*

Studying the sample fractions in Table 6.8, most of the samples will be allocated to Class I and II for a proportional sampling scheme. For all sample sizes larger than 8 all classes will be represented. The optimal sampling fractions are quite different, as the combination of many data values and a high variance results in a large portion of the sample being allocated to Class II. Class IV and V will contribute only to a small amount, as a result of the low weight of their respective strata and the low coefficient of variation in Class IV.

Table 6.8: Sampling weights for Site A.

STATISTICS	n_K	W_P	$W_{O,L}$	$W_{O,M}$	$W_{O,H}$
CLASS I	39	33%	12%	19%	19%
CLASS II	44	37%	74%	67%	63%
CLASS III	20	17%	12%	10%	14%
CLASS IV	10	8%	1%	1%	1%
CLASS V	7	6%	1%	3%	3%
ENTIRE AREA	*120*	*100%*	*100%*	*100%*	*100%*

The theoretical standard error quotients that have been calculated for Site A can be seen in Table 6.9. These indicate that the SRS and PSRS would have a similar reliability, while an OSRS scheme would have a higher reliability, between 15 and 27% fewer samples would have to be taken with the OSRS for a certainty equal to that of PSRS and SRS

Table 6.9: Comparison of standard errors for the different sampling schemes for Site A.

MOLECULAR WEIGTH/ QUOTIENT	LMW	MMW	HMW
▯▯,▯/▯▯,▯	1,00	1,00	1,00
▯▯,▯/▯▯,▯	1,27	1,18	1,15

.

The results of the data sampling simulation, seen in Table 6.10 and Figure 6.1, are more even then the theoretical analysis implied, and also differs somewhat in which sampling scheme that is the most reliable. For low molecular weight the OSRS is still the most reliable overall, but with a very small margin. For the other two molecular weights however the PSRS seems to have the highest confidence and the OSRS the lowest.

Table 6.10: Comparison of confidence levels from the data simulation for Site A.

		LMW			MMW			HMW		
		SRS	PSRS	OSRS	SRS	PSRS	OSRS	SRS	PSRS	OSRS
LMW	SRS	NA	0,993948	0,996316						
	PSRS		NA	1,002383						
	OSRS			NA						
MMW	SRS				NA	0,997014	1,002618			
	PSRS					NA	1,005621			
	OSRS						NA			
HMW	SRS							NA	0,998502	1,002365
	PSRS								NA	1,003869
	OSRS									NA

The probabilistic sampling simulations have a similar trend as the data sampling simulation, which can be seen in Figure 6.2. The optimal sampling has the highest confidence for all but one sample size for low molecular weight. For middle and high molecular weight OSRS has the lowest confidence and PSRS has the highest at low sample sizes and SRS at larger.

Studying the values of the confidence levels, the results from the sampling simulations indicates that for the middle and high molecular weight approximately 70 samples, or 60% of the survey sample size, would have been sufficient for a 100% confidence estimate of the mean. The trend is similar for low molecular weight but the confidence never reaches 100%.

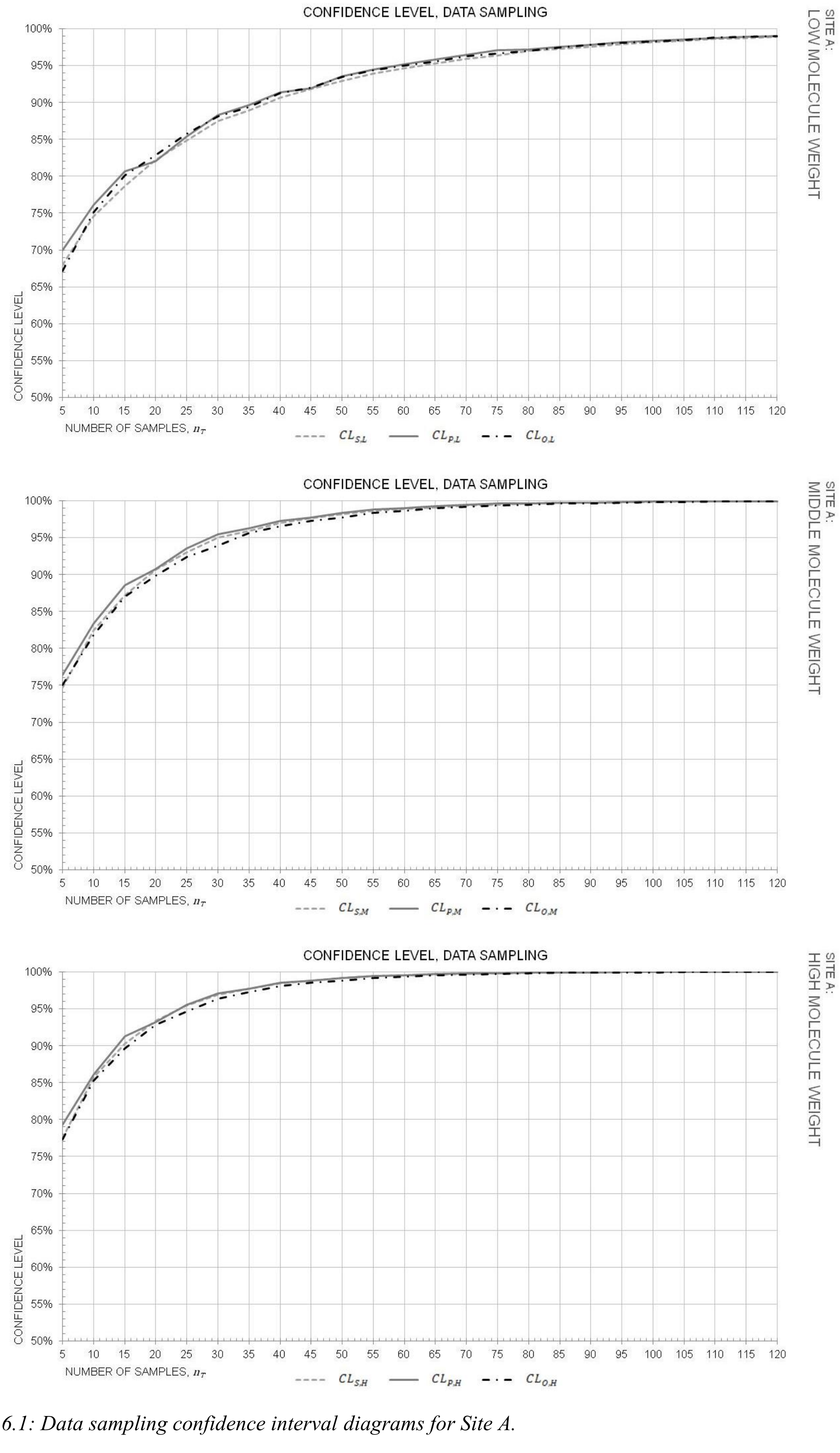


Figure 6.1: Data sampling confidence interval diagrams for Site A.

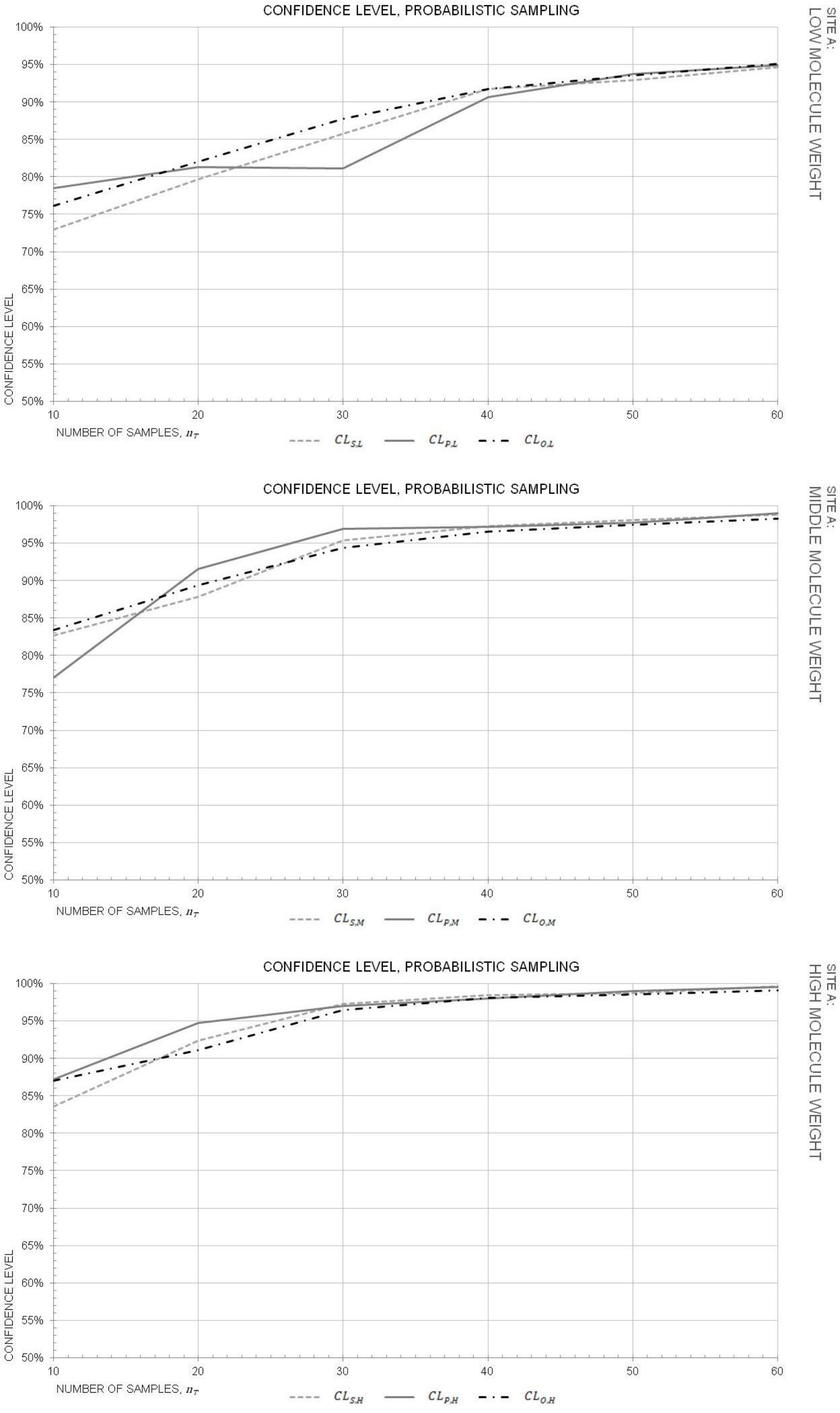


Figure 6.2: Probabilistic sampling confidence interval diagrams for Site A.

6.2.3 Result for Sampling Scheme Analysis, Site B

The samples for site B are concentrated to Class II, friction soils, and there are few or no samples in the other classes, as seen in Table 6.11. The mean concentrations at the site are high for PAHs with middle and high molecular weight, above the general guidelines for less sensitive land use. The highest mean concentrations are found in Class I, consisting of only one data value, and the lowest in Class III. For PAHs with low molecular weight the mean concentration is below the guidelines for sensitive land use.

The coefficient of variation is approximately four times as high for Class II as for Class III. Since only 1 out of 47 samples is categorized as Class I, no coefficient of variation can be calculated for this soil class even if there are soil samples in the class. The implications of this will be discussed later on.

Table 6.11: Summary statistics for Site B.

STATISTICS	n_K	$\bar{?}_?$	$\bar{?}_?$	$\bar{?}_?$	$C_{x,L}$	$C_{x,M}$	$C_{x,H}$
CLASS I	1	20,8	216,7	219,3	-	-	-
CLASS II	40	2,7	28,8	16,6	4,6	4,7	4,1
CLASS III	6	0,8	8,3	8,5	1,5	1,4	1,2
CLASS IV	0	-	-	-	-	-	-
CLASS V	0	-	-	-	-	-	-
ENTIRE AREA	*47*	*2,8*	*30,2*	*21,2*	*4,1*	*4,2*	*3,2*

In Table 6.12 above the sampling weights for the stratified sampling schemes are presented. For the proportional sampling all classes containing data values are represented with Class II as the most sampled one. For the optimal sampling scheme only data values from Class II and III are represented. Already for the proportional sampling a large portion of the samples are allocated in Class II. In the optimal sampling scheme, when the variation is included as well, the proportion of samples allocated in Class II is increased greatly and almost all samples will be distributed to the class. As mentioned before there is no variation in Class I, since it only consists of one data value, which means that the sampling weight for the class will be zero.

Table 6.12: Sampling weights for Site B.

STATISTICS	n_K	W_P	$W_{O,L}$	$W_{O,M}$	$W_{O,H}$
CLASS I	1	2%	0%	0%	0%
CLASS II	40	85%	99%	99%	98%
CLASS III	6	13%	1%	1%	2%
CLASS IV	0	0%	0%	0%	0%
CLASS V	0	0%	0%	0%	0%
ENTIRE AREA	*47*	*100%*	*100%*	*100%*	*100%*

Studying the standard errors for Site B, the results are fairly conclusive, as seen in Table 6.13. The PSRS has a higher reliability then the SRS, 6 to 14 % better, and the OSRS is even more reliable and requires between 16 and 24 % fewer samples than the SRS for an equal certainty. The results for low and middle molecular weights are the same, but the stratified sampling appears to be slightly more effective for PAHs with high molecular weight.

Table 6.13: Comparison of standard errors for the different sampling schemes for Site B.

MOLECULAR WEIGTH/ QUOTIENT	LMW	MMW	HMW
[illegible]	1,06	1,06	1,14
[illegible]	1,16	1,16	1,24

The confidence interval diagrams for the data sampling simulation, seen in Figure 6.3, show a different result than the theoretical standard error. The diagrams for low and middle molecular weight once again have a similar trend, but are different from the standard error since the SRS is consistently the most reliable sampling scheme, which can also be seen in

Table 6.14. The PSRS appears to be the least reliable alternative for all sample sizes and the OSRS falls between them.

Table 6.14: Comparison of confidence levels from the data simulation for Site B.

		LMW			MMW			HMW		
		SRS	PSRS	OSRS	SRS	PSRS	OSRS	SRS	PSRS	OSRS
LMW	SRS	NA	1,064914	1,038671						
	PSRS		NA	0,975357						
	OSRS			NA						
MMW	SRS				NA	1,043961	1,028921			
	PSRS					NA	0,985593			
	OSRS						NA			
HMW	SRS							NA	1,045208	1,030078
	PSRS								NA	0,985525
	OSRS									NA

The confidence interval diagrams for the probabilistic sampling simulation indicate something entirely different. The general result from the probabilistic sampling is that the PSRS has the highest confidence, the OSRS the lowest with the SRS somewhere in-between. However, the confidence varies over the sample sizes.

The confidence level is low for all the molecular weights for the data sampling, only the SRS for high molecular weight exceeds 95%. The results from the probabilistic simulation are slightly better; especially for PSRS and SRS, which exceed 95% for all molecular weights.

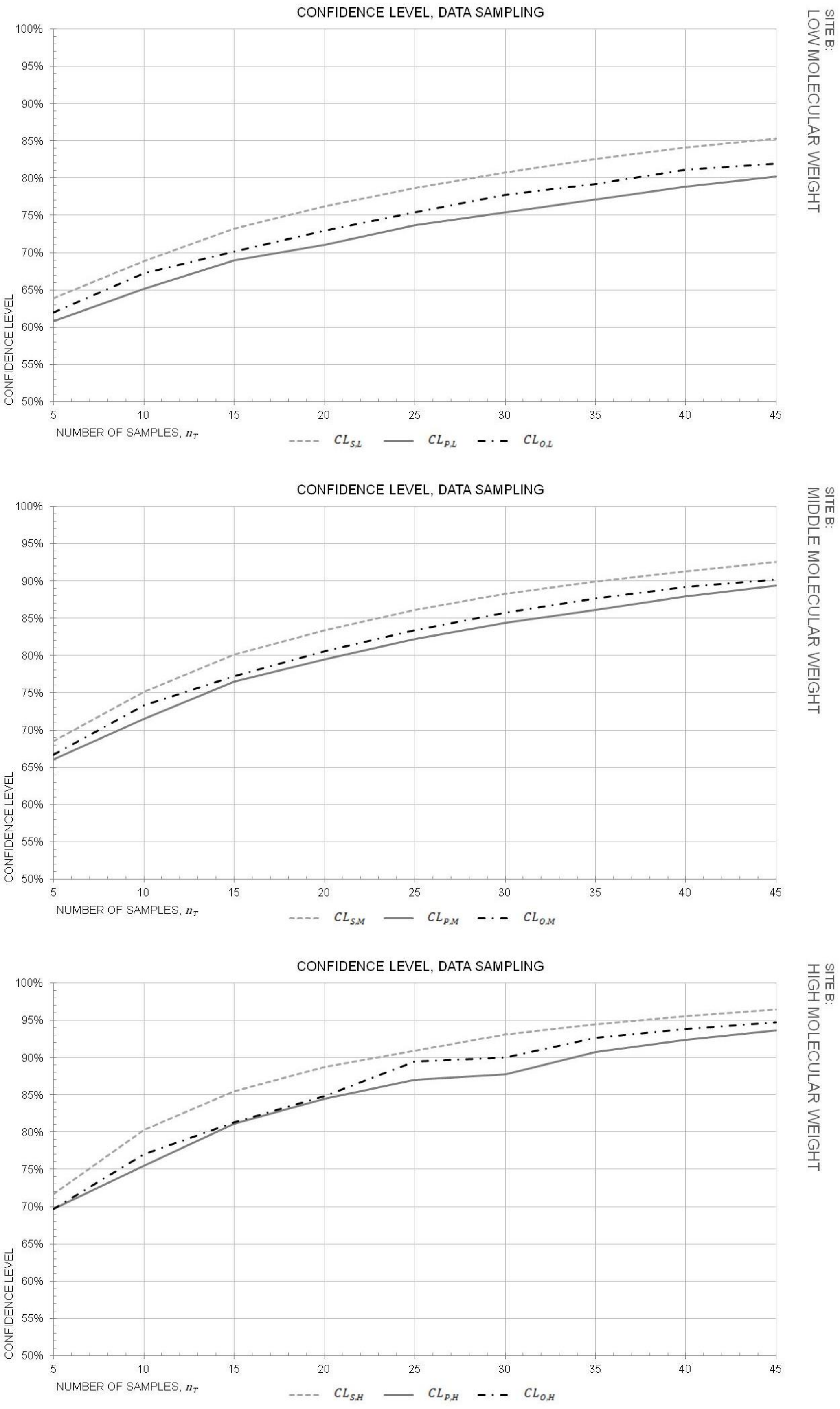


Figure 6.3: Data sampling confidence interval diagrams for Site B.

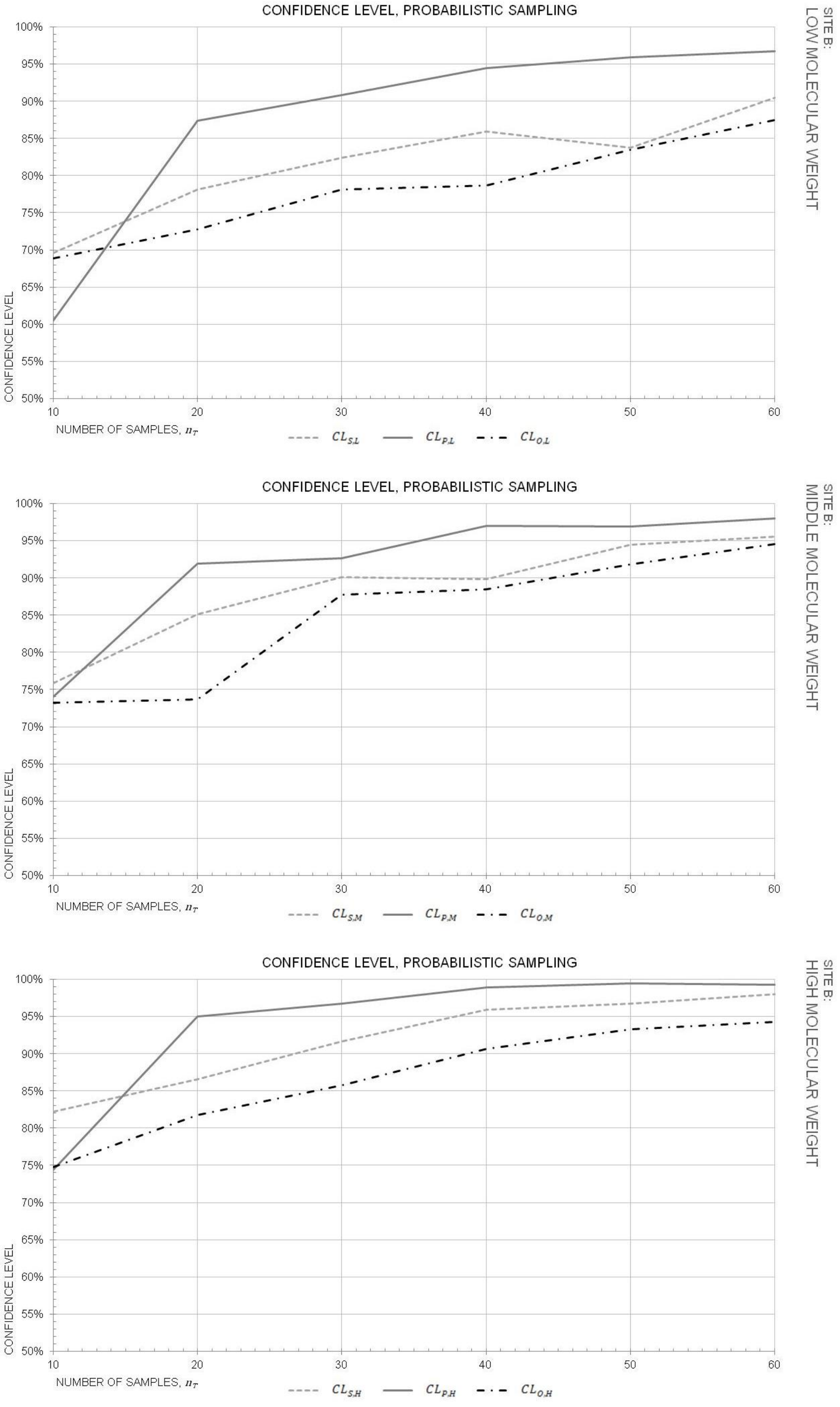


Figure 6.4: Probabilistic sampling confidence interval diagrams for Site B.

6.2.4 Result for Sampling Scheme Analysis, Site C

For Site C the results, presented in Table 6.15, is that Soil Class III is the one containing most soil samples, highest mean and the highest coefficient of variation for most molecular weights. Only Class IV has one value that is higher, the mean value for high molecular weight and it is also the only class with statistical values that are close to those of Class III. There are no samples in Class I.

In general, the values are low. For low molecular weight the overall mean concentration is not above the guideline for sensitive land use and for the other two molecular weights the mean concentration are below the guideline value for less sensitive land use. Some of the classes have mean values above this guideline, but for especially Class II and V the values are very low.

Table 6.15: Summary statistics for Site C.

STATISTICS	n_K	$?_?$	$?_?$	$?_?$	$C_{x,L}$	$C_{x,M}$	$C_{x,H}$
CLASS I	0	-	-	-	-	-	-
CLASS II	8	0,2	1,7	1,6	1,2	1,8	1,5
CLASS III	16	1,5	19,5	15,5	3,1	3,3	2,9
CLASS IV	12	0,7	11,9	18,8	1,2	1,5	2,2
CLASS V	4	0,1	0,5	0,9	1,1	1,0	0,9
ENTIRE AREA	40	0,9	11,8	12,3	3,5	3,5	3,0

The high number of samples and the high variance in Class III are reflected in the OSRS sample allocation, as seen in Table 6.16. Even though only 40% of the data values are in the class, more than 80% of the soil samples are allocated to it for low and middle molecular weight. For high molecular weight Class IV is also highly represented, but Class V is never included in the optimal sampling and Class II only for sample sizes exceeding 25.

Table 6.16: Sampling weights for Site C.

STATISTICS	n_K	W_P	$W_{O,L}$	$W_{O,M}$	$W_{O,H}$
CLASS I	0	0%	0%	0%	0%
CLASS II	8	20%	2%	2%	2%
CLASS III	16	40%	86%	81%	58%
CLASS IV	12	30%	11%	17%	40%
CLASS V	4	10%	0%	0%	0%
ENTIRE AREA	*40*	*100%*	*100%*	*100%*	*100%*

The result of the standard error analysis presented in *Table 6.17* is that the PSRS is equal in reliability with the SRS, while the OSRS is more reliable for high molecular weights and significantly more reliable for low and middle molecular weights. Between 17 and 37 % fewer samples would have to be taken using an OSRS scheme compared to a SRS or PSRS.

Table 6.17: Comparison of standard errors for the different sampling schemes for Site C.

MOLECULAR WEIGTH/ QUOTIENT	LMW	MMW	HMW
[illegible]/[illegible]	1,00	1,00	1,00
[illegible]/[illegible]	1,37	1,31	1,17

The data sampling simulation diagrams in Figure 6.5 also indicates that the OSRS is more reliable, which can be seen in Table 6.18 with the comparison of the mean confidence. For the probabilistic sampling simulation the results, which can be seen in Figure 6.6, confirms this. While the difference is slightly lower in the latter results, both sampling simulations indicate the same trend. OSRS is the most reliable, especially for middle and high molecular weights, and PSRS and SRS have almost the same confidence.

Table 6.18: Comparison of confidence levels from the data simulation for Site C.

		LMW			MMW			HMW		
		SRS	PSRS	OSRS	SRS	PSRS	OSRS	SRS	PSRS	OSRS
LMW	SRS	NA	0,996063	0,985627						
	PSRS		NA	0,989522						
	OSRS			NA						
MMW	SRS				NA	0,991573	0,950283			
	PSRS					NA	0,958359			
	OSRS						NA			
HMW	SRS							NA	0,992647	0,950464
	PSRS								NA	0,957504
	OSRS									NA

The confidence levels for the OSRS on Site C are high, and the results from the data sampling indicates that for example for high molecular weight the survey sample size could be reduced to half for a 95% certainty. The OSRS is in general approximately 10% more reliable for the data sampling simulation and 5% for the probabilistic simulation.

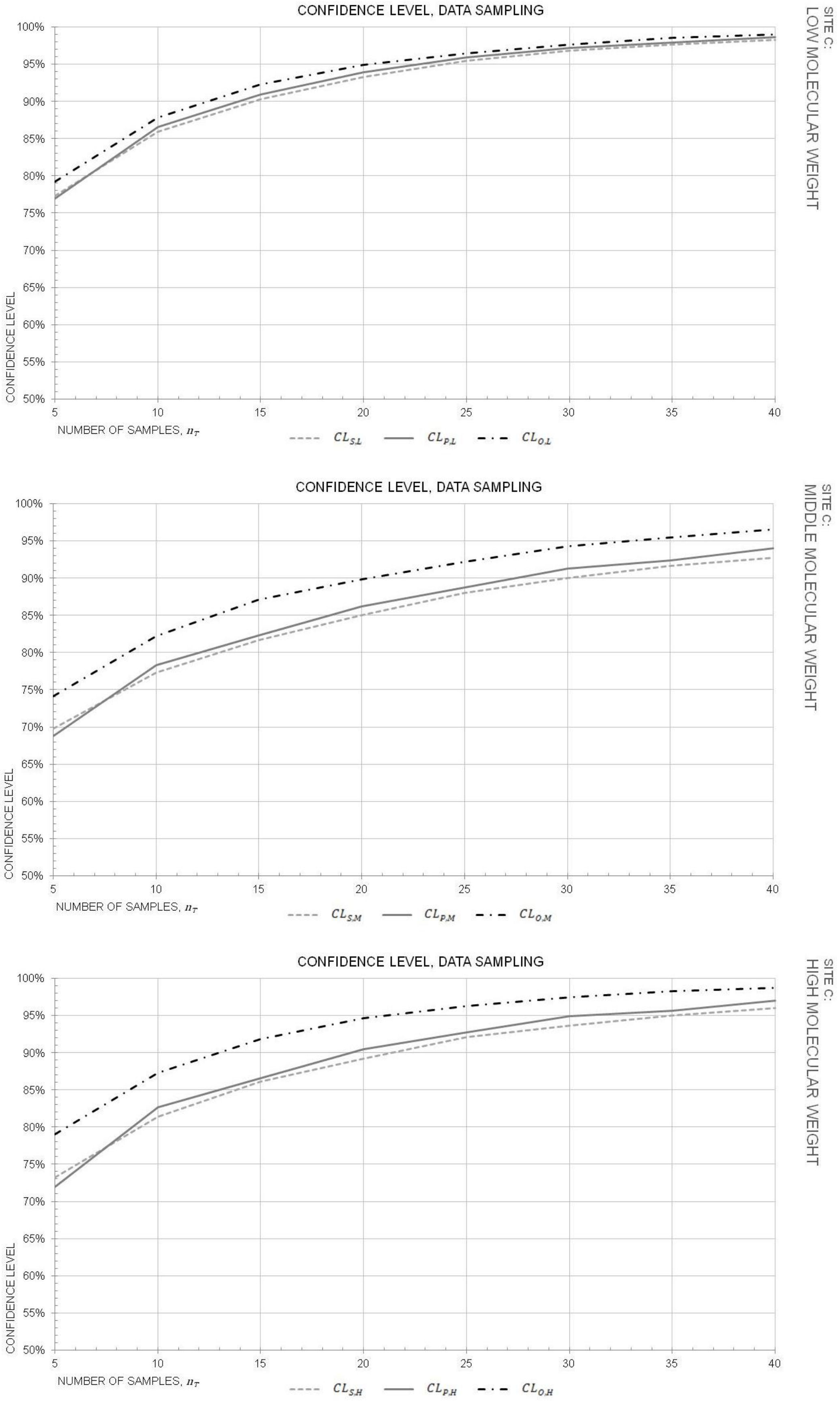


Figure 6.5: Data sampling confidence interval diagrams for Site C.

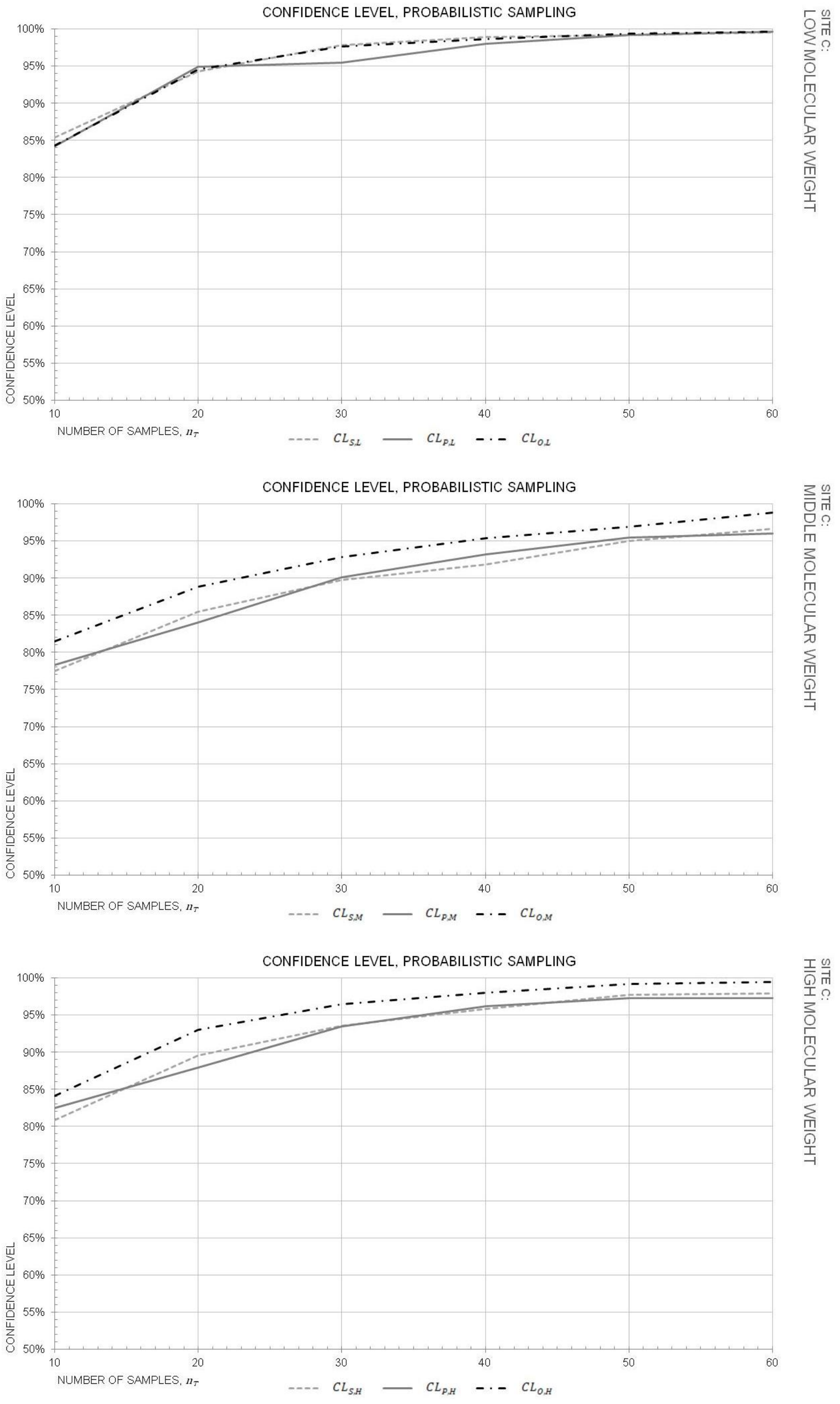


Figure 6.6: Probabilistic sampling confidence interval diagrams for Site C.

6.2.5 Result for Sampling Scheme Analysis, Site D

All soil classes are represented for Site D and contain at least two data values, as seen in Table 6.19. Class II contains 75% of the samples which together with the 13 % of the samples in class IV covers the major part of the data. The total mean values are significantly higher for the middle and high molecular weights than for the samples with low molecular weight. The means for Class II and IV are overall higher than the means for the other classes. The same trend applies to the coefficient of variation.

Table 6.19: Summary statistics for Site D.

STATISTICS	n_K	$?_?$	$?_?$	$?_?$	$C_{x,L}$	$C_{x,M}$	$C_{x,H}$
CLASS I	4	0,3	0,4	0,5	0,2	0,3	0,1
CLASS II	87	0,3	6,6	6,8	3,3	3,4	2,8
CLASS III	2	0,1	0,2	0,2	0,1	0,9	0,8
CLASS IV	15	0,8	6,9	6,0	1,4	1,8	1,4
CLASS V	8	0,2	2,0	2,4	0,5	1,0	1,0
ENTIRE AREA	*116*	*0,4*	*6,0*	*6,0*	*2,7*	*3,4*	*2,8*

The sample weight for the different sample schemes are presented in Table 6.20. For the PSRS all classes are represented with Class II as the most sampled. Since the number of samples and the coefficient of variation determine the sampling weight for the OSRS the high values for these parameters for Class II and IV results in that the classes together amount to approximately 99 % of the samples.

Table 6.20: Sampling weights for Site D.

STATISTICS	n_K	W_P	$W_{O,L}$	$W_{O,M}$	$W_{O,H}$
CLASS I	4	3%	0,2%	0,0%	0,0%
CLASS II	87	75%	83,9%	90,8%	92,1%
CLASS III	2	2%	0,0%	0,0%	0,0%
CLASS IV	15	13%	15,0%	8,4%	6,9%
CLASS V	8	7%	0,9%	0,7%	1,0%
ENTIRE AREA	*116*	*100%*	*100%*	*100%*	*100%*

As seen in Table 6.21 the theoretical standard error follows the same trend for low, middle and high molecular weights samples. The OSRS has an approximately 7 % more reliable result than both the PSRS and SRS, which both has the same standard error.

Table 6.21: Comparison of standard errors for the different sampling schemes for Site D.

MOLECULAR WEIGTH/ QUOTIENT	LMW	MMW	HMW
$\square_{\square,\square}/\square_{\square,\square}$	1,00	1,00	1,00
$\square_{\square,\square}/\square_{\square,\square}$	1,06	1,07	1,08

The confidence interval diagrams from the data sampling, seen in Figure 6.7, shows a slightly different result compared to the standard error. For the low molecular weight all three sample schemes reaches a confidence level of 100 % for all sample sizes. For middle and high molecular weight PSRS is the most reliable for the smallest sample size, but as with the standard error OSRS gives the most reliable result for most sample sizes, see Table 6.22 for a comparison of the levels.

Table 6.22: Comparison of confidence levels from the data simulation for Site D.

		LMW			MMW			HMW		
		SRS	PSRS	OSRS	SRS	PSRS	OSRS	SRS	PSRS	OSRS
LMW	SRS	NA	1,000016	1,00001						
	PSRS		NA	0,999994						
	OSRS			NA						
MMW	SRS				NA	0,99667	0,990401			
	PSRS					NA	0,993709			
	OSRS						NA			
HMW	SRS							NA	0,997858	0,992587
	PSRS								NA	0,994718
	OSRS									NA

For the probabilistic sample simulation the confidence interval for the low molecular weight shows the same result as for the data sampling. The diagrams for middle and high molecular weight follow a different trend however. For middle molecular weight the OSRS gives the most reliable result for all different sample sizes, while the other two sample schemes have a significantly lower confidence level. For the high molecular weight the sample schemes are more similar but even here OSRS gives the most reliable result except for small sample sizes, about 10 %, where SRS is the most reliable.

The result from the confidence level analysis indicates that fewer samples could have been taken with a retained confidence. Both sampling simulations, neglecting low molecular weight, has a confidence above 95 % for both middle and low molecular weights at about 50% of the soil samples for all sampling schemes, and requires even fewer using OSRS.

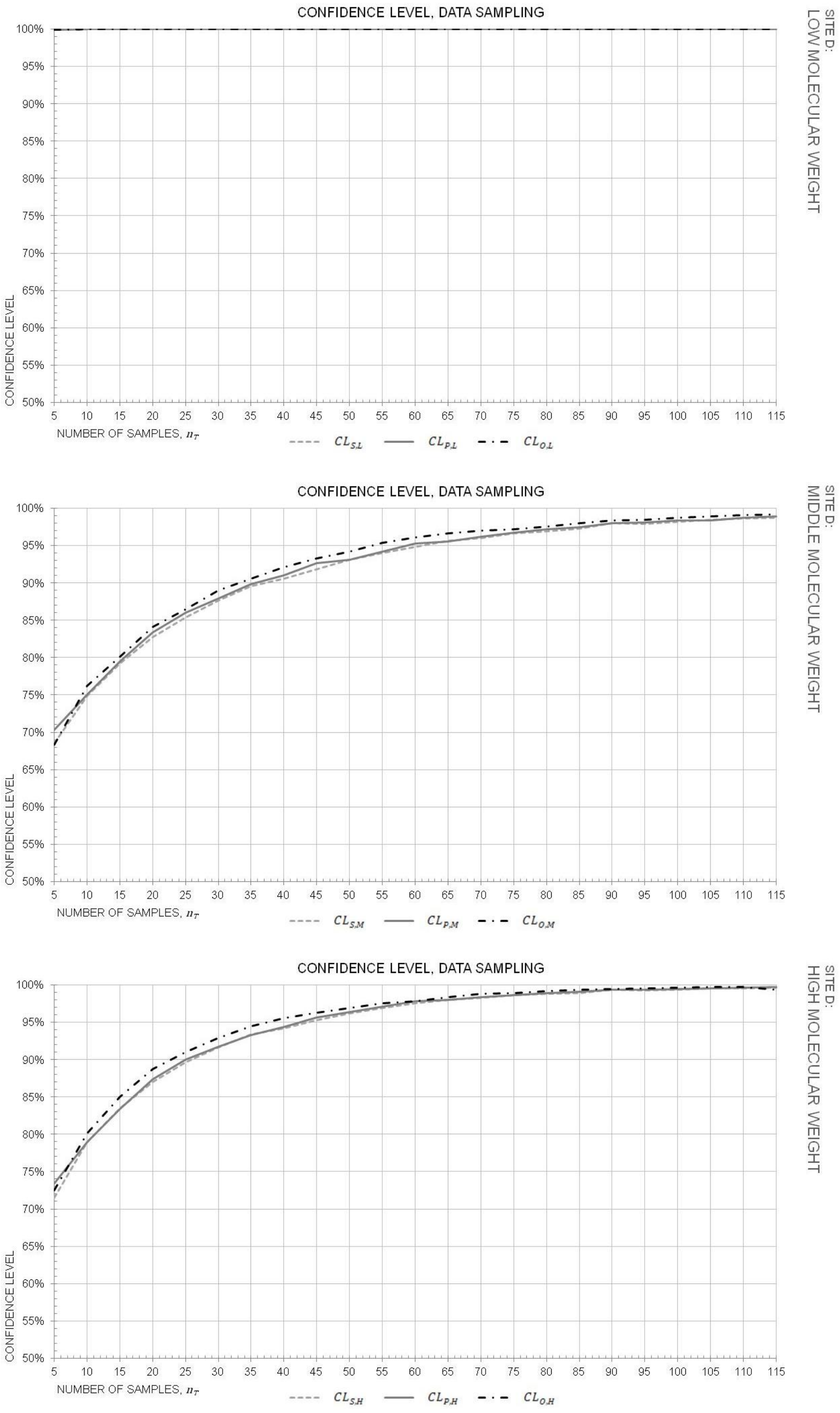


Figure 6.7: Data sampling confidence interval diagrams for Site D.

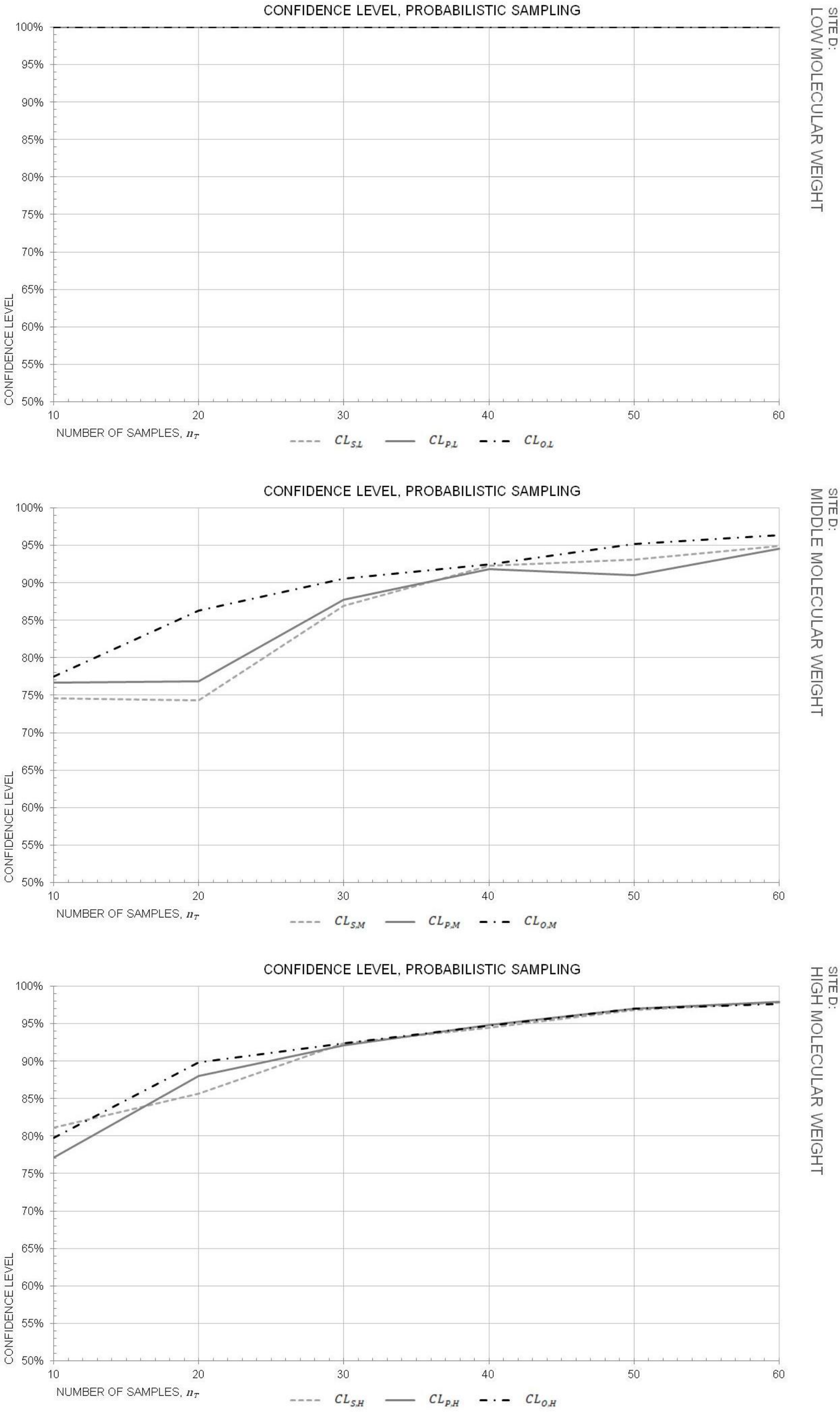


Figure 6.8: Probabilistic sampling confidence interval diagrams for Site D.

6.2.6 Result for Sampling Scheme Analysis, Site E

As seen in the statistical summary in **Fel! Ogiltig självreferens i bokmärke.** only Class I – Class III are represented for the site where class II consist of the major part of the samples. The mean value for Class I is below the guidelines for all different molecular weights and the coefficient of variation is low for the class as well. For Class II the mean value is below the guidelines for sensitive land use for the low molecular weight but not for middle and high molecular weights. However, the mean values are still below the guidelines for less sensitive land use. The coefficient of variation is about 5 times higher for Class II compared to Class I. For Class III the mean values are below the guidelines for less sensitive land use for all molecular weights. The coefficient of variation for Class III is low for low and middle molecular weights but significantly higher for high molecular weight.

Table 6.23: Summary statistics for Site E.

STATISTICS	n_K	$?_?$	$?_?$	$?_?$	$C_{x,L}$	$C_{x,M}$	$C_{x,H}$
CLASS I	3	0,2	0,2	0,0	0,5	0,5	0,0
CLASS II	48	2,0	6,6	4,9	2,0	2,6	2,6
CLASS III	2	2,5	4,0	2,0	0,5	0,1	8,9
CLASS IV	0	-	-	-	-	-	-
CLASS V	0	-	-	-	-	-	-
ENTIRE AREA	*53*	*2,0*	*6,1*	*4,5*	*2,0*	*2,7*	*2,7*

The sampling weights are presented in Table 6.24. 91 % of the samples are allocated to Class II for the proportional sampling scheme, and the sample size must reach at least 20 data values for all three classes to be represented. For the optimal sampling scheme Class II is the most sampled one with at least 98 % of the strata. Because of the low mean and coefficient of variation for class I the group is not included in the OSRS. Class III is sampled for both low and high molecular weights but only at a large sample size.

Table 6.24: Sampling weights for Site E.

STATISTICS	n_K	W_P	$W_{O,L}$	$W_{O,M}$	$W_{O,H}$
CLASS I	3	6%	0%	0%	0%
CLASS II	48	91%	98%	100%	99%
CLASS III	2	4%	2%	0%	1%
CLASS IV	0	0%	0%	0%	0%
CLASS V	0	0%	0%	0%	0%
ENTIRE AREA	*53*	*100%*	*100%*	*100%*	*100%*

The theoretical standard error, seen inTable 6.25, follows the same trend for all the groups of PAHs where OSRS is the most reliable scheme with a 5 % lower standard error than SRS and PSRS. The PSRS is only slightly better than the SRS.

Table 6.25: Comparison of standard errors for the different sampling schemes for Site E.

MOLECULAR WEIGTH/ QUOTIENT	*LMW*	*MMW*	*HMW*
[illegible]	1,01	1,01	1,01
[illegible]	1,05	1,06	1,06

The result from the data sampling simulation, seen in Figure 6.9, is different between the PAH groups. For low molecular weight OSRS is the most reliable one. For middle molecular weight OSRS and SRS are equally reliable while PSRS is continuously the least reliable sampling scheme, see Table 6.26. For high molecular weight SRS is continuously the most reliable and PSRS the least reliable alternative.

Table 6.26: Comparison of confidence levels from the data simulation for Site E.

		LMW			MMW			HMW		
		SRS	PSRS	OSRS	SRS	PSRS	OSRS	SRS	PSRS	OSRS
LMW	SRS	NA	0,996547	0,992303						
	PSRS		NA	0,995741						
	OSRS			NA						
MMW	SRS				NA	1,012958	1,001374			
	PSRS					NA	0,988565			
	OSRS						NA			
HMW	SRS							NA	1,034998	1,010411
	PSRS								NA	0,976245
	OSRS									NA

Performing a probabilistic sampling simulation the result seen in Figure 6.10 once again differs between the PAH groups. For low molecular weight PSRS and OSRS are equal and the most reliable sampling schemes. For middle molecular weight PAHs OSRS is continuously the most reliable alternative, while for high molecular weight it is the most reliable for sample sizes exceeding 16 samples.

The confidence levels for Site E are significantly higher for the probabilistic sampling than for the data sampling. For both middle and high molecular weights the confidence level barely exceeds 95 %, but for probabilistic sampling this occurs at approximately 40 samples. Except for the high molecular weight data sampling, where the PSRS is approximately 5 % below the other two, there is no significant difference between the sampling schemes in when the confidence level exceeds the confidence limits.

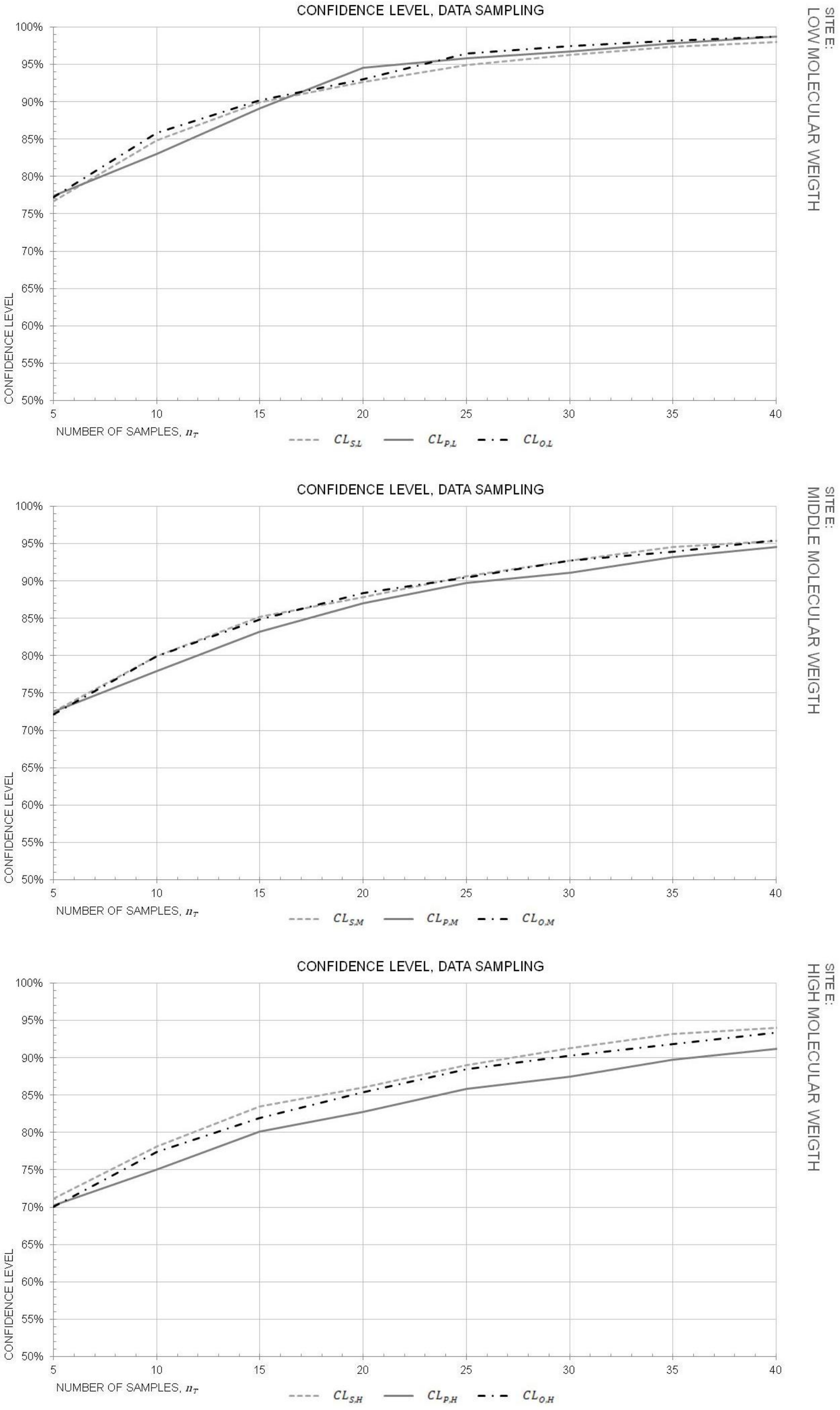


Figure 6.9: Data sampling confidence interval diagrams for Site E.

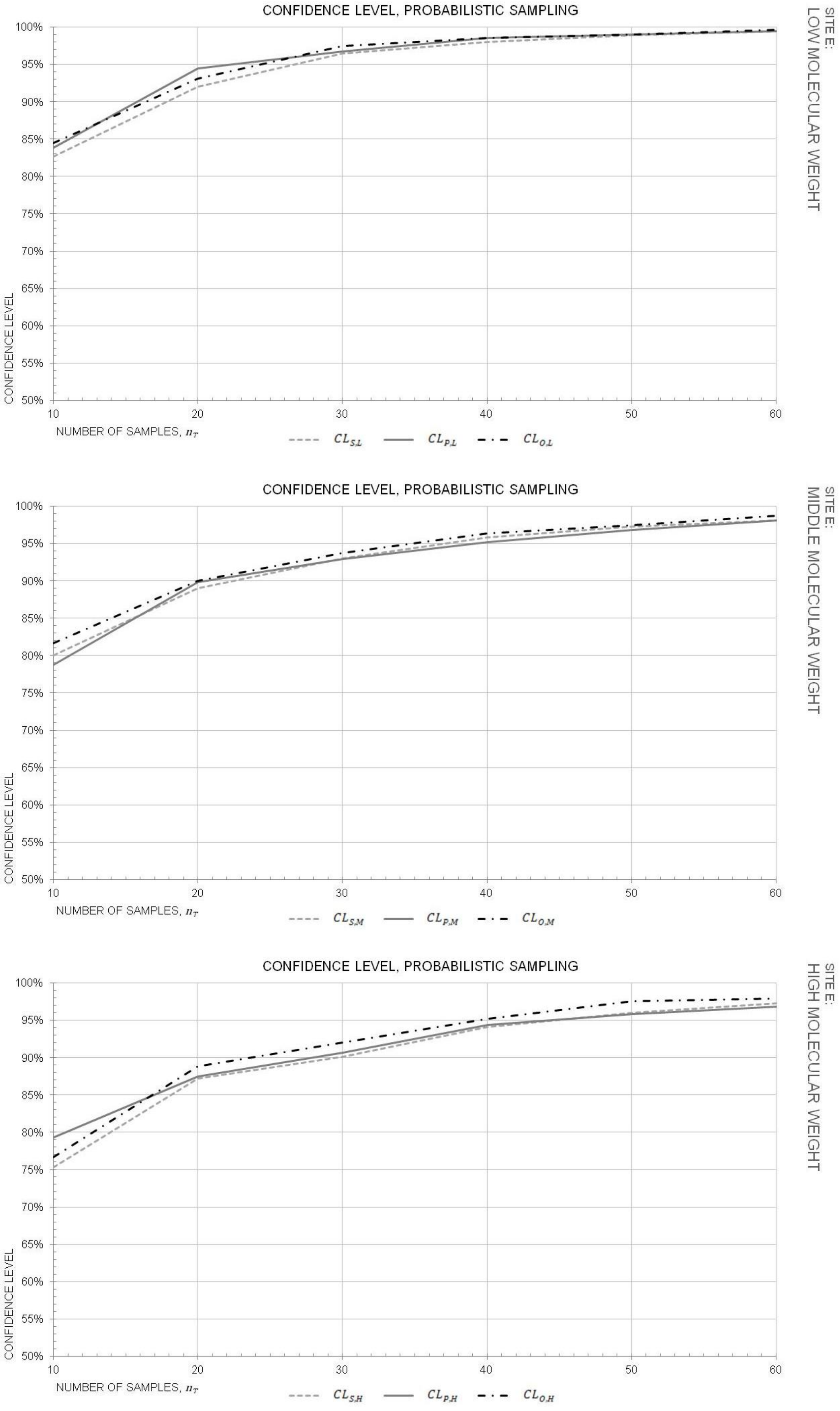


Figure 6.10: Probabilistic sampling confidence interval diagrams for Site E.

6.3 Results, TOC Analysis

Site A was the only area with a large number of data values for total organic carbon, TOC where almost all samples analysed for PAH also has data on TOC. For Site B 17 out of 47 samples and for Site C only 4 out of 40 samples contain information about TOC content. For site D and site E no information about TOC content is available. The results from the correlation analysis between TOC and contamination level are presented in Appendix 10. For Site A and B the results are fluctuating and no trend can be seen. For Site C soil with a higher TOC content exhibits a higher concentration of PAH.

Site A is the only area with sufficient data to calculate the theoretical standard error and the quotient between the sample schemes. The quotients for all PAH groups are presented in Table 6.27. The standard error is 8% lower for OSRS on low molecular weight, but for the other two the difference between SRS, PSRS and OSRS is marginal.

Table 6.27: Comparison of standard errors for the different sampling scheme using TOC stratification on Site A.

MOLECULAR WEIGTH/ QUOTIENT	LMW	MMW	HMW
[illegible]	1,00	1,00	1,00
[illegible]	1,08	1,02	1,01

7 Discussion

This chapter discusses the evaluation and analysis of the results in chapter 6. As with the results, the discussion is performed for the three different analysis steps. However, for the soil determination and TOC analysis there are no site-specific sections as only a general evaluation has been performed.

7.1 Soil Type Determination

As was stated in section 6.1, the results from the sieving indicated that approximately 80 % of the samples were placed in the correct soil class based on the field survey. The most common error was an overestimation of the content of fine sediments; all but two of the errors were of this type.

The result of the overestimation of the fine sediments will be an overrepresentation of Soil Class III, IV and V. As both the proportional and the optimal stratified random sampling are based on the proportion of the class to some extent, this will mean that too many samples will be taken in these classes. The statistical calculations will also be affected, since data values will end up in the wrong strata.

What should be remembered when evaluating the results from the sieving is that no wet sieving could be performed at the laboratory, resulting in an uncertainty regarding the amount of fine sediments. The results from the hydrometer analysis are used to compensate for this and the results are considered acceptable but they are not a completely accurate.

For a large number of the soil samples, the name was not correct. For this thesis, this is not considered as a problem if it is only a question of the grain fractions being in the wrong order. The result of the statistical analysis is not affected if, for example, gravelly sand is classified as sandy gravel or silty clay as clay as the soil class will be the same.

Even if the overall error in the field determination is considered low enough to continue with the analysis, they will contribute to the overall uncertainty in the sampling procedure. To connect to the previously discussed approach to Sampling Uncertainty, the error originating in faulty determination of the soil class would be considered to be a sampling uncertainty error as it would affect the sub-sampling performed to select which samples to submit for chemical analysis (Back, 2003, pp. 56-64).

The additional analysis made for Site E regarding fractions larger than 20 mm indicate that there is a possibility of neglecting a considerable proportion of the material in the soil volume during drilling as large fractions will not be extracted. For the analyzed site this does not affect the classification, but at sites with more fine materials it could result in errors in the soil classification. Once again, referring to the conceptual description of sampling uncertainty in Figure 2.1 - if the large fractions are ignored it is an example of an Increment Extraction Uncertainty.

In conclusion, the result from the soil determination analysis indicates that the field classification is sufficiently accurate for further analysis. The classification is not intended to function as a basis for more advanced calculations, for example geotechnical, and the precision indicated from the sieving is considered to be acceptable for the purposes in this thesis.

7.2 Sampling Scheme Analysis

The results from the sampling scheme analysis will be evaluated first for each site and then all the results will be evaluated to find similarities between the sites.

7.2.1 Sampling Scheme Analysis of Site A

The results of the statistical analysis of the data for Site A can be interpreted to indicate a difference in behavior between PAHs with low molecular weight compared to middle and high. While the theoretical analysis implies that the OSRS would be the most suitable sampling scheme for all molecular weights, the sampling simulations shows something else. While the OSRS has a better confidence level for low molecular weight for both sampling simulations, the PSRS and SRS is better for middle and high. Considering the fact that there is a notable difference in contamination level between low molecular weight PAHs and the other two groups, with a lower mean but higher coefficient of variation, could be an indication of different transportation history for the three groups.

As described in section 2.3 the three molecules included in the low molecule weight group have a significantly higher solubility with water and volatility than the heavier molecules. Considering the proximity to water of Site A, high groundwater flow and the lower mean concentration for low molecular weight, the results could indicate that the lighter molecules to a higher extent have been affected by transportation mechanisms and that the concentrations therefore are closer related to the soil type then for the heavier molecules where the spatial location of the samples in relation to the peak points in the area could be more important factors to allocate the sample in the most optimal way.

Finally, one important result to remember from the basic statistical analysis of the site is that Soil Class II had both the highest coefficient of variation and the highest mean concentration, which is surprising considering that high concentrations are commonly found in fine sediments.

7.2.2 Analysis of Site B

The theoretical results for Site B indicate that both the stratified random samplings would be suitable tools when performing a soil survey in the area. When this is applied with sampling simulation analysis, the results are quite different however and a SRS scheme seems more suitable for all sample sizes.

There could be several reasons for the difference between theoretical and practical results. These can be connected to both the conditions of the site and statistical factors. For one thing, the site is polluted by a point source, which likely leads to a stronger correlation between contamination level and distance from the source rather than from soil classes.

That there is no correlation between soil type and contamination level is supported by the fact that the results from the probabilistic sampling indicated an entirely different trend then both the data sampling and the theoretical analysis.

The soil samples in the area are unevenly distributed among the classes, and almost all the samples are allocated to the soil class with the highest mean and variance. In addition, one soil class contains only one data value that has a high concentration compared to the mean. Since this class has no variance, it is omitted during sampling with OSRS and considering that this omitted soil sample has a very high data value, it can affect the results.

7.2.3 Sampling Scheme Analysis of Site C

The results for Site C all indicate that optimal stratified random sampling is the most suitable sampling scheme for the site. However, the theoretical standard error indicated that the OSRS would have the largest difference in reliability compared to other sampling schemes for the low molecular weight. This is not reflected in the results from the sampling simulation, as the difference in confidence level between OSRS and the two other sampling schemes are larger for middle and high molecular weight than for low.

An interesting observation for the area is that the theoretical results for optimal sampling are best for the molecular weight that has a high variance for one soil class and equal for the rest. For high molecular weight, where the theoretical results are significantly worse the variation in variance is higher and this indicates that the method is most effective when there is one class with a high variance and an equal in the remaining. This is in line with the result from the sensitivity analysis.

7.2.4 Sampling Scheme Analysis of Site D

Even if all different soil classes are represented at the site the allocation of the data values differs between the classes where only two and four data values are within Class III and Class I respectively. As mentioned in the result most of the samples are allocated to Class II, especially for the OSRS.

According to the results from the theoretical standard error OSRS would give a slightly more accurate result than both PSRS and SRS. The difference is very low however, and this is reflected in the results from the data sampling simulation where the three different sampling schemes yield an almost equal certainty but with OSRS as a slightly more reliable alternative.

The result for the probabilistic data simulation gives a different result for middle molecular weight however as the optimal stratified random sampling is the most reliable by a larger marginal for some sample sizes, but these larger differences might be calculation anomalies. The conclusion from the two different sampling simulations is that there is a correlation between soil class and contamination level.

The cause of the 100% confidence level for low molecular weight for all sample sizes and sampling schemes is unknown, but might be because of the low data values and that many of them are equal.

7.2.1 Sampling Scheme Analysis of Site E

The data for Site E is not statistically ideal for the stratification used in the thesis since almost all of the soil samples are concentrated to one class. This means that no matter which sampling scheme is used, a majority of the samples will be taken from this class and the result will be almost the same for all schemes. This is reflected in the results from both the standard error and sampling simulation analyses, as all the sampling schemes yields similar reliability. Which one is the most reliable differs between the analyses, and there seem to be no clear connection between soil type and contamination level for the site.

The high concentration of soil samples in one class was also noted for Site B, and Site E also share other properties with that site. In addition to the high concentration of samples in one class, the area has been contaminated by industrial activities and the area should be considered a point source. That the contamination level does not correlate with the soil type is not surprising as the spatial location should be more important.

7.2.2 Overall discussion of the Sampling Scheme Analysis

There is no pattern common to all sites in the basic statistical observations. The highest mean value can be found in Class II for three sites and Class I and III for one site each. The coefficient of variation is the highest for Class II for four sites which could be a pattern, especially since it is the second highest for the fifth site. Since it is not conclusive and all sites do not have samples in all classes, nothing can be said for certain regarding the basic statistics.

The general result from the calculated standard error is that the OSRS is always a more reliable alternative than both SRS and PSRS. How much better it is varies, but it is on average 16% more reliable than the SRS and 14% more reliable then the PSRS. For all the sites where some kind of relationship between soil type and contamination level can be seen, the standard error for SRS and PSRS is approximately the same.

The results from the simulated confidence levels indicate that fewer samples could have been taken at the sites with a retained high confidence. Several simulations reach a 100 % confidence well before the sampling size of the survey, and since a lower confidence usually can be accepted several simulations indicate that the sample size could have been smaller. For the sites where OSRS seems to be effective, C and D, this scheme would have required approximately 10 fewer samples.

What can be seen in the data is that there seem to be a correlation between soil type and contamination level for Sites C and D. These sites have similar contamination history; with the pollution originating in already contaminated materials that have been used to fill up the site. Compared to Site B and E, where the pollutions originate from point source releases there seems to be no correlation and the assumption of a correlation between soil type and contamination level is faulty. For Site A, which can be considered to be something in between point and diffuse, a connection can be observed for low molecular weight but not for the other two.

For the sites where a correlation between contamination level and soil class can be seen the OSRS scheme is more reliable according to the standard error than what can be seen in the sampling simulations. For example for middle molecular weight for Site C the OSRS has a standard error that is 31 % lower than the other two sampling schemes but the difference in the data sampling simulation is only approximately 5 %.

Relating to the approach to sampling uncertainty explained in section 2.1. If the stratification according to soil class is successful, it will help to reduce the Short-range Variability as it is related to variations due to geological properties such as grain size. The method cannot reduce the variability that originates in spatial variations in the contamination level, which is exemplified by the fact that the method does not work for the sites polluted by point sources.

7.3 TOC Analysis

As seen in the diagrams in Appendix 10 it appears as if there is no correlation between the two parameters, TOC and contamination level, for Site A or B. For Site C the parameters seem to be correlated but the number of data values is insufficient to make any conclusions and it could be a coincidence.

The quotients for the standard errors are lower for TOC compared to the earlier calculated quotients for soil classes, which indicates that there is a higher correlation between soil classes and contamination level than between TOC and contamination level.

The amount of available data for the TOC is however insufficient to make any certain observations. TOC data is only available for three of the sites, and with few data values for one of these and a very narrow interval for one more. The site where a satisfying amount of data is available is extreme in its contamination history, and for a better analysis more sites with other conditions would have been required.

8 Conclusions

The general conclusion from the analysis performed is that a stratified sampling based on soil classes is a well-suited tool under the conditions that will be described below. The sieving confirmed that the quality of the field determination, although it can be improved, is accurate enough. In addition, the TOC analysis indicates that the soil type seems to be more correlated to the contamination levels than the TOC content.

The conditions under which the method works are in line with what could be expected beforehand. The stratification method appears to be effective in areas where the contaminations originate in the filling materials and not from a point source at the surface. For Site A, where the contamination originates in the filling materials but the method is less effective, the interpretation is that the area is too extreme for the method to work. It should be considered an industrial landfill, with high concentrations and several hot spots, rather than filling materials that were used to raise the ground level like at Site C and D.

As the results are too inconclusive to give a general guideline, the method will work best if there is some knowledge of the contaminations at the site before the sample scheme is designed. This means that the optimal stratified random sampling is not suited for a preliminary assessment, but can reduce expenses and improve the reliability of the results if applied in later stages, such as for an in-depth assessment.

A result that was unexpected was that there was no soil class that had the highest variation, or mean, for all sites. If there had been a clear trend, a better recommendation on sampling scheme design could have been made. Based on the current data, no soil class to sample more in general can be recommended. However, it seems as the variation is generally lower in the Class IV and V, where fine sediments dominates, and this could indicate that fewer samples can be taken here.

It is also noteworthy that the friction material often had the highest mean concentration values, which was unexpected considering previous research. The reason for this can be that even if the soil class is a friction soil, there could still be up to 15% clay and silt included in the sample. If the contamination process has resulted in a higher exposure to the friction soils, which will be the case if the contaminations originate at the ground surface from for example a factory and if the friction soil is the upper soil type, it is likely that the contaminates will bind to the fine fractions included in the friction soil. The fine fractions below would then have a lower contamination level, as they would be exposed to a lower contamination level.

To investigate if this was the case, an additional sieving and chemical analysis procedure that was not applied here could have been performed, where a number of randomly selected samples were sieved and each fraction analyzed individually. From this, the contamination level in each fraction of the soil sample would be known which would answer the question of where in the grain size distribution the contaminations are found.

The results for Site A indicate that there might be a possibility of using the stratification method for areas polluted by point sources as well as diffuse. For the middle and high molecular weight there appears to be no clear connection between contamination level and soil type. For the low molecular weight however, a correlation appears to exist. As described earlier, this could indicate that for an area that has been subjected to transport mechanisms and where the contaminant is mobile a connection between contamination and geological conditions is present. The likely conditions required for this to occur are that the area is left undisturbed for a longer period of time, with no ongoing contamination so that the peak concentrations have been given time to be diluted.

While most focus has been on the optimal sampling, which is also the method that had the best results, there are benefits in using the proportional stratification if no information on the site is available. It has a slightly higher confidence than SRS for Site C and D, and considering that there are few structured methods for selecting samples for chemical analysis it is also a sampling scheme that is easy to apply if a random sample is desired.

The sieving performed as a part of the thesis confirms a basic assumption upon which the analysis is based. The ocular determination performed in the field has a quality that is sufficient to base the stratification upon. There is a significant error in the field determination, especially regarding the name of the soil, but as long as most of the samples are in the correct class the stratification should result in a more reliable analysis. For a better result, the determination quality of the field determination should be improved, especially regarding the content of fine sediments.

The additional grain size distribution analysis performed for Site E, where data from the coarser sieving performed at the site during remediation was included, indicated a possible source of error in the field determination. During the exploratory drilling it is uncommon to have fractions above gravel in the extracted sample, but the data from Site E had volumes where up to 37% was above 40 mm. For the samples analyzed, this had no effect since they were all in Class II for the site but it indicates that a large part of the subsurface volume can be stones or boulders that at best will be registered as resistance during drilling and this can affect the classification. For a better geological representation it is recommended that the auger drilling is complemented with excavation pits, or in areas where this is not possible sounding-methods that detects boulders, which will show a more accurate soil stratigraphy. It should be noted that while common in Sweden, auger drilling is prohibited in many countries due to the risk of disturbing the samples (Swedish Geotechnical Soceity, 2004).

The short analysis of the relationship between TOC content and contamination level indicate that the initial assumption made in this thesis, that the relationship between soil type and contamination was the more interesting, was correct. While this result is unreliable due to a low amount of data, the result for Site A, where there is a satisfactory amount of data, indicates that while there could be a relationship, the stratification based on soil class yields a more reliable result.

Another aspect that weighs in favor for the soil type based stratification is the cost. The TOC analysis has to be performed at a laboratory, which takes time and can be expensive if it has to be performed on a large number of soil samples. The soil type classification on the other hand is already a part of the standardized field survey and adds neither additional expenses nor time consuming phases to the work process.

One of the things that have been studied in the thesis is how the method relates to the approach to total sampling uncertainty. Applying an optimal stratified sampling scheme will decrease the uncertainties related to the variability caused by the geology as it systemizes how the samples are allocated based on how the contaminants vary between the fractions. It adds other possible errors however, related to the expertise of the surveyor and how the samples are extracted.

Since the difference in reliability between the sampling schemes is relatively low the use of the stratification method can be discussed. For example the two additional sources of uncertainty with the method, erroneous field classification and the extraction uncertainty, could negate the increased reliability gained with the stratification.

Another problem that could arise using an optimal stratified random sampling scheme is that the sampling weights are based on the variance. If a soil class has a high mean but low variation, it will not be selected for sampling in the OSRS, which can lead to an unreliable sampling result. In this case, a proportional sampling will give a more reliable result. This is partially illustrated in the results for Site B, where one of the soil classes has only one sample with a very high concentration. As no standard deviation can be calculated for on one data value, the class is omitted in the OSRS. In this particular case however, this could be considered something positive as one soil sample out of 50 could be considered to be an anomaly rather than something to be included in the survey. Errors could arise for classes with more data values but low variance however, and it should be remembered when choosing sampling scheme.

During the research of background information in this subject, a major problem with the current sampling methodology was identified. The handbooks from both the Swedish and US Environmental Protection Agencies provides in-depth information and guidance on the design of the drilling pattern, but nothing on selecting soil samples for chemical analysis. This is likely a source of a statistical error, as the common method will be to select which samples to submit for chemical analysis based on the field observations and knowledge of the site history. While this will likely find most of the pollution, it is not a sound statistical method as it is not random, and therefore no statistics can be calculated on the data.

If only suspected polluted samples are selected, there will be a bias and the measured concentrations will only be the extreme values, the peak points, and not representative for the entire site. While it is important to identify and remove hazardous contamination levels methods like this can lead to an over-extensive remediation, which is not sustainable from an environmental or economic perspective. It should also be remembered that even if a stratified sampling method is used it is important that all the samples sent for chemical analysis are randomly selected disregarding any impressions from the field survey. If applied correctly however, the results from this thesis will provide a tool to perform more steps compared to the current handbooks of the soil survey in a statistically sound manner.

9 Further Studies and Other Applications

There are several ways the results from this thesis can be expanded and used both in other real life applications and as a starting point for further theoretical studies.

The analysis performed in this thesis could be improved in several ways. While an analysis regarding the correlation between TOC content and contamination level has been performed, the amount of data is insufficient to make any conclusions with certainty. A study where sites with better data for TOC, both in number of data values and covering a larger interval, should be performed to be able to say for certain that the soil fraction is more important than the organic content.

One of the assumptions that can be questioned is the way organic soils are handled. In the current classification, organic soils are classified as Class III if they are the primary soil type but if they are a secondary soil type they are not considered during classification. As the connection between organic content and contamination level is still not clear, the decision to include them in the analyzed data can be questioned.

Another way to improve the analysis would have been to use sites where the data has a higher degree of homogeneity between the sites. The available data for the reference sites differs greatly with different survey sizes between 40 and 120, TOC available for almost all samples on one site and no samples on two sites, no grain size distributions for Site D and only information on larger sieving fractions for Site E. The results would have been more compatible if the sites had the same amount and type of data.

It should also be compared to other methods of selecting samples for analysis, especially concerning how the results will differ if the samples are selected subjectively based on survey protocols compared to using a structured, statistical method. As no other methods to select samples for chemical analysis have been found during the work with this thesis, it appears to be an area where there is a lack of knowledge and the survey and remediation process could benefit from more research being focused at this.

While the focus in this study was to find ways to improve the selection of which soil samples to send for chemical analysis, at least the proportional stratified sampling method can of course be used in the design of the drilling pattern for the survey if the soil fractions are known. The knowledge will have to be detailed, and it will usually not be enough with the geological maps available as they are too inexact, and it demands that a site survey has been performed beforehand, for example a geotechnical survey.

A common remediation method is to divide an area into several smaller sections and then sample and remove section by section. Usually a grid is used, where each section has the same size, but the results from this thesis can be applied to design a more dynamic remediation system. If there is a knowledge on which soil fractions that have the highest variation and which have the least, the remediation can be adapted to this. The soil fraction with a high variation can be divided into smaller sections while the soil fractions where the variation is low does not demand as many sample points and can be larger. This could reduce the number of required samples but the larger sections will also reduce construction time.

The immediate application of the methodology is not only in the survey process. Sampling is also an important part of the remediation process, where soil is analysed continuously to determine how it should be disposed of. As a field survey commonly has been performed before the remediation, data on the variance of the different soil classes will be available and the methods can be used to optimize the sampling during remediation, which will expedite the work process and reduce costs.

10 Recommendations

The recommendation from this thesis is that designing a sampling scheme using the optimal stratified random sampling scheme is a method that is well suited if:

- sampling has been performed in the area before,
- previous sampling has been performed statically correct manner,
- the materials are non-landfill filling materials with diffuse contamination, and
- the field survey does not consist of only auger drilling for more reliable results.

A proposed methodology for performing an optimal stratified random sampling scheme is:

1. Classify data from a previous site assessment and calculate mean values and standard errors for the classes.
2. Classify the data from the second field survey and calculate the weights of each class using the statistical methods described in section 2.4.
3. Use the mean values and standard deviations from the first survey together with the weights from the second to calculate the sampling fractions.
4. Select the soil samples to send for analysis according to the sampling weights. The selection of samples must be random to avoid any bias.

If a stratified random sampling scheme is applied under the conditions above the gain will be a lower cost compared to a simple random sampling and a more representative value compared to a biased subjective sample selection. A proportional stratified sampling scheme can yield more reliable results than a simple random sampling scheme if applied, if no samples have been taken in the area beforehand.

The work with this thesis has also identified a number of possible weaknesses with the current survey procedures, which are:

- There is currently no standardized method for selecting samples for chemical analysis which will result in unreliable statistics and a likely bias.
- For all the field surveys performed in this thesis, the surveyors tended to overestimate the fine sediment content in the sample.

Finally, ways to expand the work performed for this thesis are:

- A more in-depth analysis of the relationship between contamination level and TOC content since previous research indicates that PAHs bind to organic material.
- A study of remediation applications. The knowledge on variation can be used to optimize cell-sizes and sampling frequency, which would lead to more efficient and sustainable remediation.

References

Ale kommun/Norconsult. (2009). *Surte 2:38 och del av 43:1, Ale Kommun: MKB för tillståndsansökan.* Göteborg: Norconsult.

Ale kommun/SWECO VIAK. (2007). *Tidermans Utfyllnadsområde: Huvudstudie.* Göteborg: SWECO VIAK.

Back, P.-E. (2003). *On Uncertainty and Data Worth in Decision Analysis for Contaminated Land.* Göteborg, Sweden: Department of GeoEngineering, Chalmers University of Technology.

Back, P.-E. (2006). *Value of Information Analysis for Site Investigations in Remediation Projects.* Göteborg, Sweden: Department of GeoEngineering, Chalmers University of Technology.

Bengtsfors kommun/Geo Innova AB. (2003a). *Mark- och grundvattenförhållanden vid EKA-området i Bengtsfors (EKA 2002:1).* Göteborg: Geo Innova AB.

Bengtsfors kommun/Geo Innova AB. (2003b). *EKA-projektet i Bengtsfors: Föroreningssituationen i mark och grundvatten (EKA 2000:2).* Göteborg: Geo Innova AB.

Bernes, C. (1998). *Organiska miljögifter.* Värnamo: Naturvårdsverket.

Brinkhoff, P. (2011). Remedial techniques for soil and groundwater. *Enviornmental Risk Assessment in Engineering.* Göteborg, Sweden: Chalmers Univeristy of Technology.

Burden, D. S., & Sims, J. L. (1999). *Fundamental of Soil Science as Applicable to Management of Hazardous Wastes.* Washington D.C.: US EPA.

Geological Survey of Sweden. (2011, 09 01). Groundwater map, detailed survey: Gothenburg. Uppsala.

Geological Survey of Sweden. (2011). *Map viewers - Quarternary deposits*. Retrieved December 8, 2011, from Geological Survey of Sweden: http://www.sgu.se/sgu/eng/produkter-tjanster/kartvisare-jord_e.html

GF Konsult. (2006). *Fördjupad förstudie på objektet Tidermans utfyllnadsområde, fastighet Surte 2:30, Ale kommun: MIFO fas 2.* Vänersborg: Västra Götaland County Administration.

Google. (2011, October 6). *Google Maps*. Retrieved October 6, 2011, from Google: http://maps.google.com

Göteborgs stad/Norconsult AB. (2011). *Ny Göta älvbro - Prövningsprocesser: Inventering - Föroreningsstatus i mark och sediment.* Göteborg: Norconsult AB.

Health and Safety Authority, Ireland. (2010). *EU Ban on chemicals in tyres*. Retrieved 12 13, 2011, from Health and Safety Authority: http://www.hsa.ie/eng/

Karlsson, R., & Hansbo, S. (1984). *Jordarternas indelning och benämning.* Linköping: Swedish Geotechnical Soceity.

Miljöbalken. (1999, june). Lag 1998:811 om införandet av Miljöbalken. Vadstena, Sweden: Förlagshusen.

Miljöförvaltningen. (2006). *Faktablad nr 135: Asfalt och tjärasfalt.* Göteborg: Göteborgs Stad.

Mölndals stad. (2010). *Program för detaljplanering av Trädgården 1:124 m.fl. (f.d. Hexion).* Mölndal: Mölndals stad.

NCC Boende AB/SWECO. (2009). *Fastigheten Trädgården 1:124, Kvarnbyparken, Mölndal: Fördjupad riskbedömning inklusive beräkning av platsspecifika riktvärden.* Göteborg: SWECO.

Norberg, T. (2011). Monte Carlo Handouts. *Environmental Risk Assessment in engineering.* Gothenburg: Chalmers university of Technology.

Norra Älvstranden Utveckling AB/SWECO. (2009). *Östra Kvillebäcken : Fördjupad riskbedömning inklusive beräkning av platsspecifika riktvärden.* Göteborg: SWECO.

Norrman, J. (2004). *On Bayesian Decision Analysis for Evaluating Alternative Actions at Contaminated Sites.* Gothenburg, Sweden: Department of GeoEngineering, Chalmers University of Technology.

Norrman, J., Back, P.-E., Engelke, F., Sego, L., & Wik, O. (2009). *Provtagningsstrategier för förorenad jord (Report 5888).* Stockholm, Sweden: Swedish EPA.

Norrman, Jenny; Purucker, Tom; Back, Pär-Erik; Engelke, Fredric; Stewart, Robert. (2009). *Metodik för statistisk utvärdering av förorenad jord.* Stockholm, Sweden: Swedish EPA.

Rice, J. A. (2007). *Mathematical Statistics and Data Analysis* (Third Edition ed.). Belmont, CA: Brooks/Cole.

Sterner, O. (2003). *Förgiftningar och miljöhot.* Lund, Sweden: Studentlitteratur.

Stevens, R. (2010). *Environmental Geology.* Göteborg: Department of Earht Sciences, University of Gothenburg.

Swedish EPA. (1997). *Generella riktvärden för förorenad mark (Report 4638).* Stockholm: Swedish Environmental Protection Agency.

Swedish EPA. (2000). *Method for inventories of contaminated sites (Report 5053).* Stockholm: Swedish Environmental Protection Agency.

Swedish EPA. (2007). *Oavsiktligt bildade ämnens hälso- och miljöegenksaper - Report 5736.* Stockholm, Sweden: Swedish EPA.

Swedish EPA. (2011, 04 08). Frågor och svar om de generella riktvärdena. Stockholm.

Swedish Geotechical Society. (1996). *Geoteknisk Fälthandbok - Allmänna råd och beskrivningar (Report 1:96).* Linköping, Sweden: Swedish Geotechical Society.

Swedish Geotechnical Soceity. (2004). *Fälthandbok - Miljötekniska markundersökningar (SGF Report 1:2004).* Linköping, Sweden: Statens geotekniska institut.

US EPA. (2002). *Guidance on Choosing a Sampling Design for Environmental Data Collection (EPA QA/G-5S).* Washington D.C., USA: US EPA.

Wikipedia. (2011, 08 28). Polycyclic aromatic hydrocarbon. Wikimedia Foundation, Inc.

PAH Conversion

For the areas where the data is divided into the old classification with the carcinogenic and non-carcinogenic PAH groups the values have to be converted into the new classification with low, middle and high molecular weights, accordingly to Figure A below. The values for each group are calculated as the sum of the concentrations for the PAHs in the group[1,2]. If a value is below the accepted level of certainty at the laboratory, the given value for the molecule is divided by 2 to get the mean.

While the laboratory presents sums where one or more of the value was below the accepted level of certainty with a less then sign (<), this is not done in the collected data in this thesis to enable calculations.

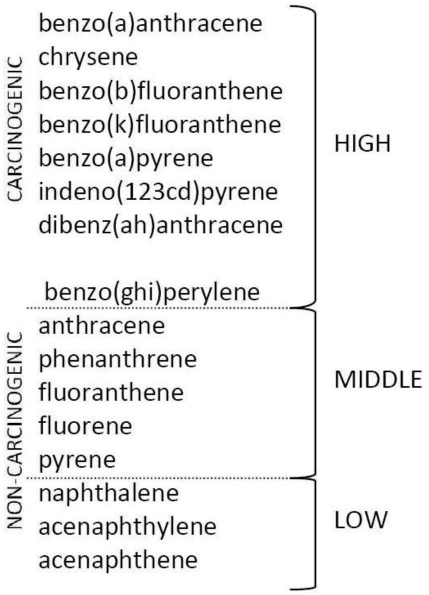


Figure A: The comparison between the former used PAH groups to the left and the present used PAH groups to the right.

---

[1] Swedish EPA 1997 (Generella riktvärden för förorenad mark, rapport 4638)

[2] Swedish EPA 2011 (Frågor och svar om de generella riktvärdena för förorenad mark)

Soil Classification

The abbreviations for the soil fractions follow the Swedish Geotechnical Surveys denotation system. In this system, the fraction names are abbreviated with the letter in the parenthesis in Table A, usually the first two letters. For composite soils, the primary soil is written last with a capital first letter and other components are then listed in order of magnitude, with the one with the lowest amount first. For example, silty clay is written siCl, and clay with silt and sand where there is more sand then silt is written sisaCl. If there is a very low content of a fraction, it is written within a parenthesis, e.g. (si)saCl.[1]

The soil texture triangle, seen in Figure A, is used to determine the correct name for the sieved soil samples.

Table A: The fraction names with associated grain sizes.

NAME (ABBREVIATION)	GRAIN SIZE [mm]
BOULDERS (Bo)	>600
STONE (St)	600-60
GRAVEL (Gr)	60-2
SAND (Sa)	2-0,063
SILT (Si)	0,063-0,002
CLAY (Cl)	<0,002
ORGANIC SOIL (Or)	-
PEAT (P)	-
MUD (Mu)	-

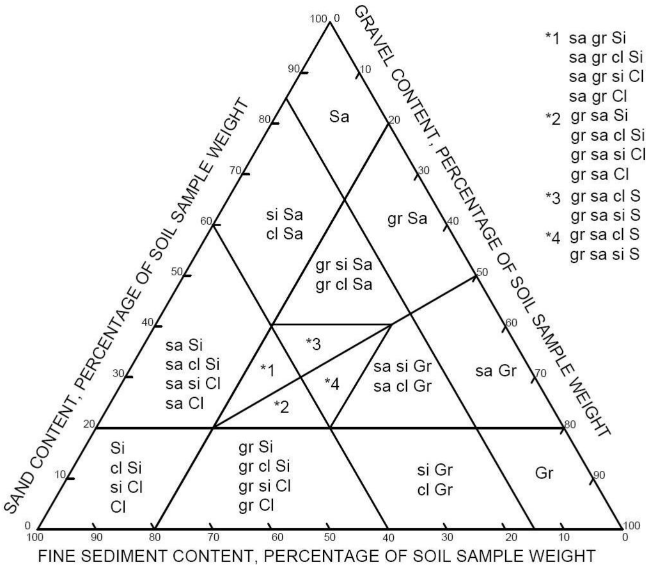


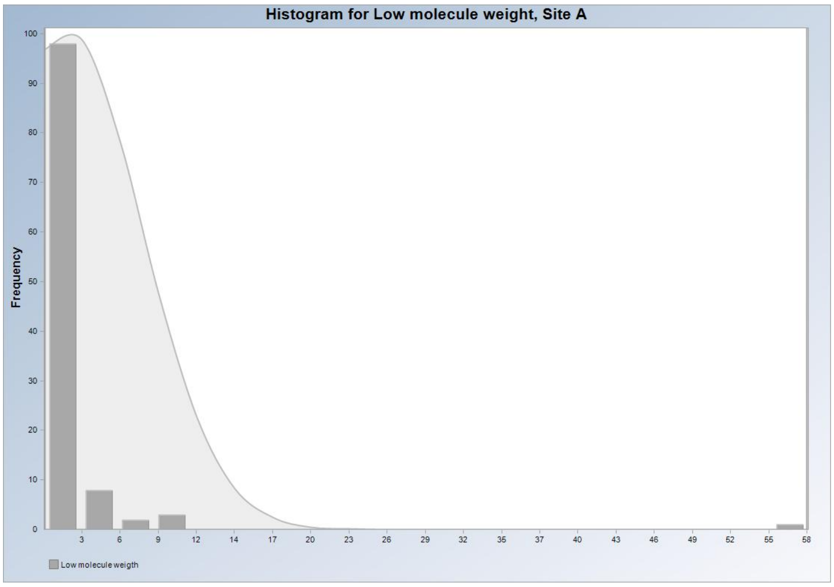
Figure A: The soil texture triangle.

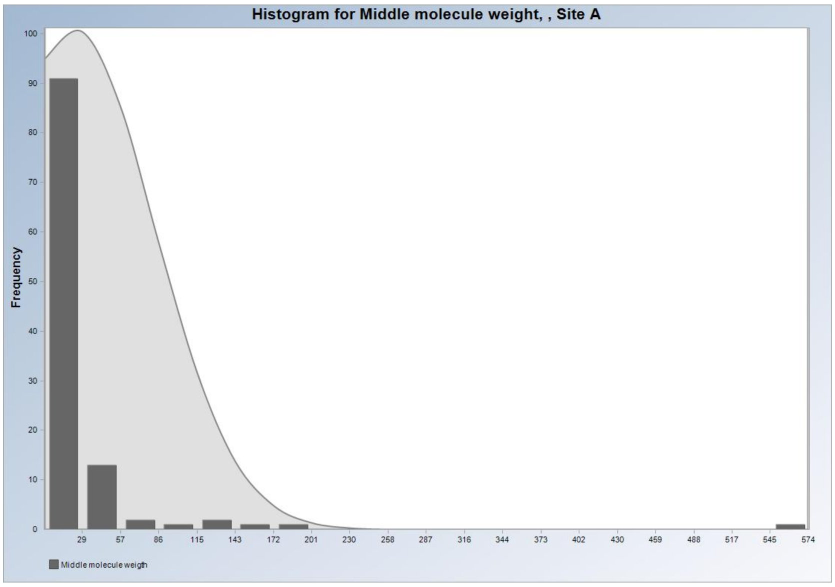
---

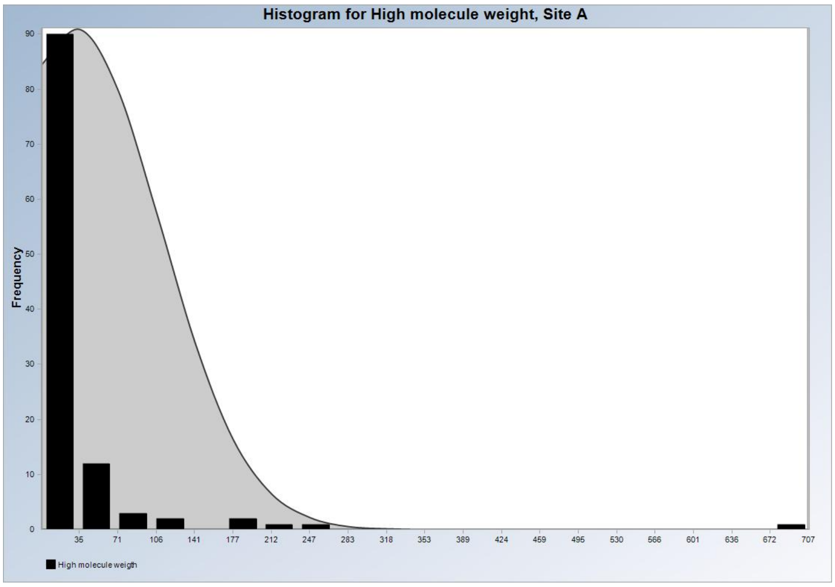
[1] Swedish Geotechnical Soceity, 2001, *Beteckningssystem för geotekniska utredningar.*

Appendix 3: Histograms

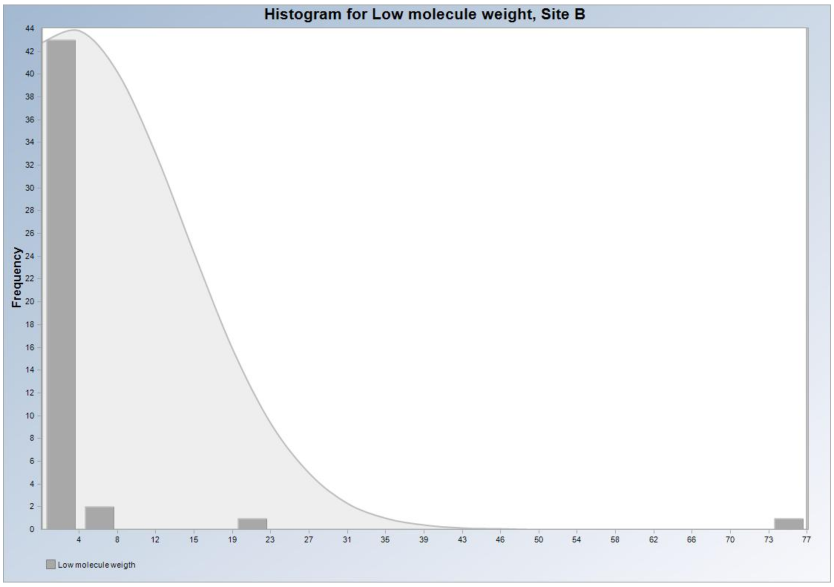
Histograms

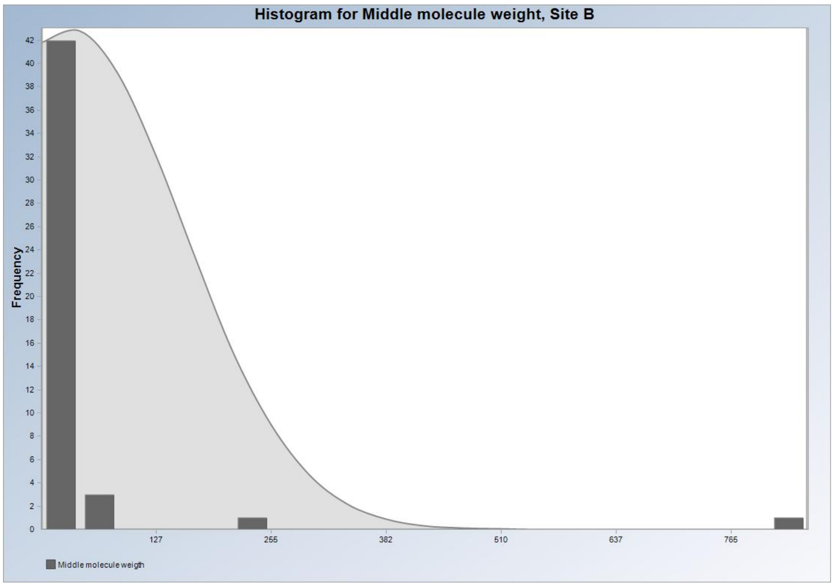


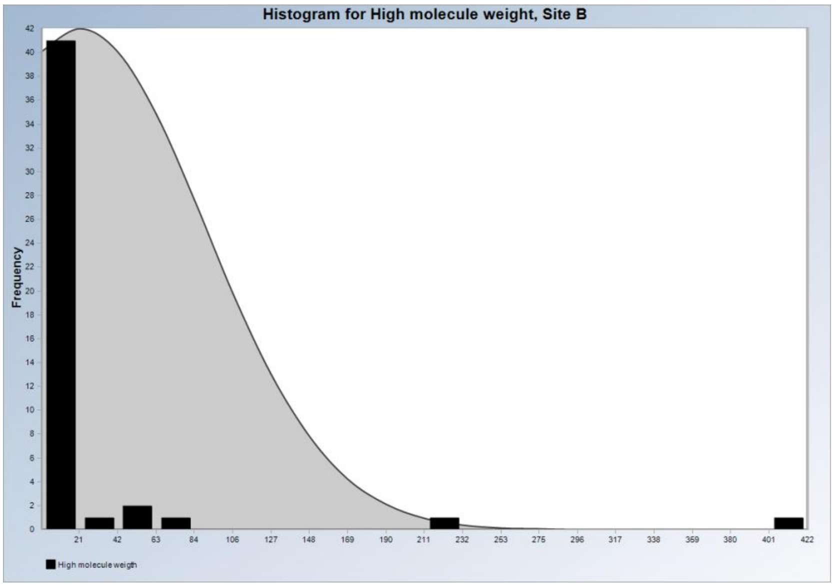




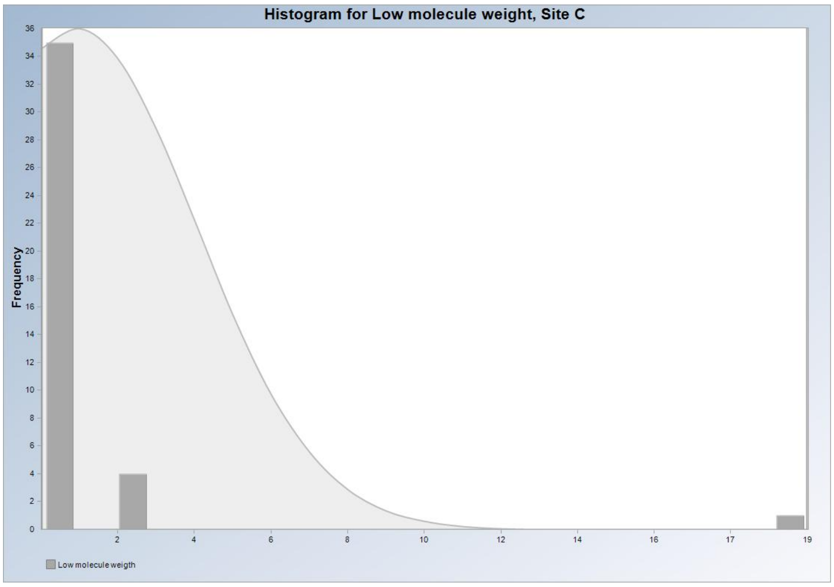
Appendix 3: Histograms

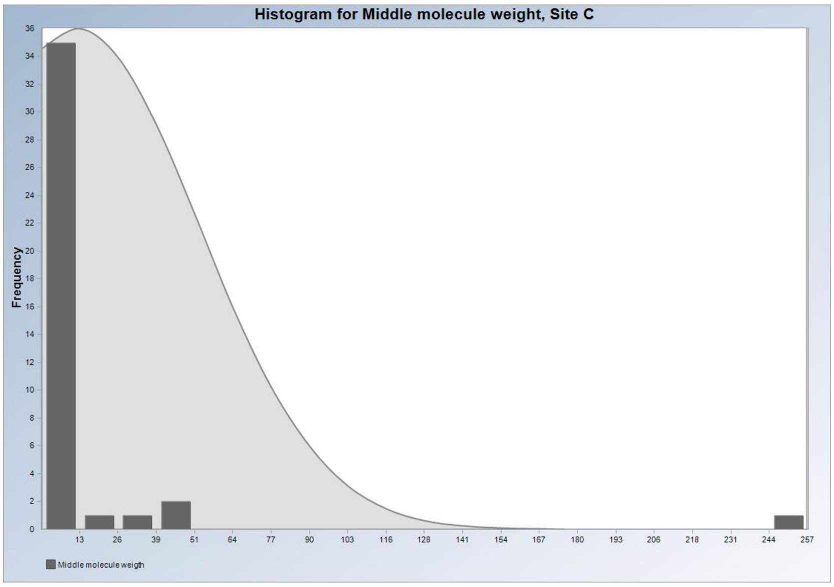


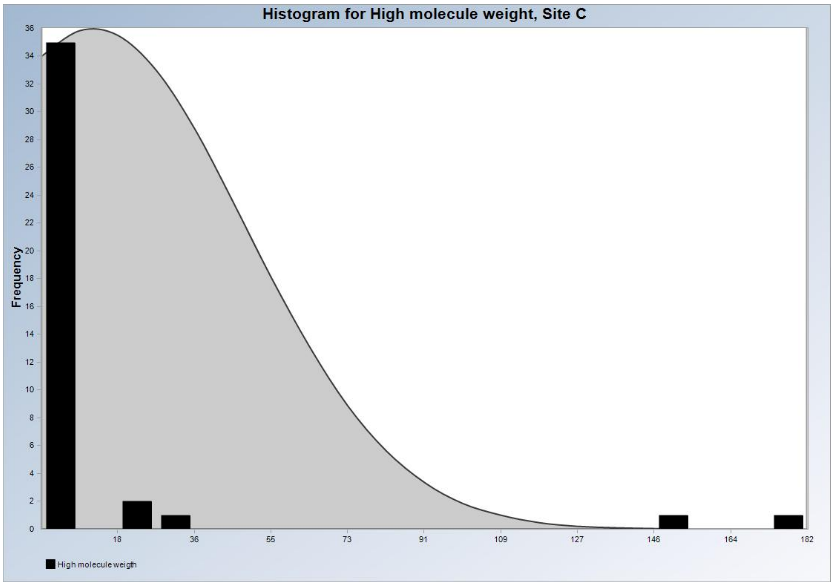




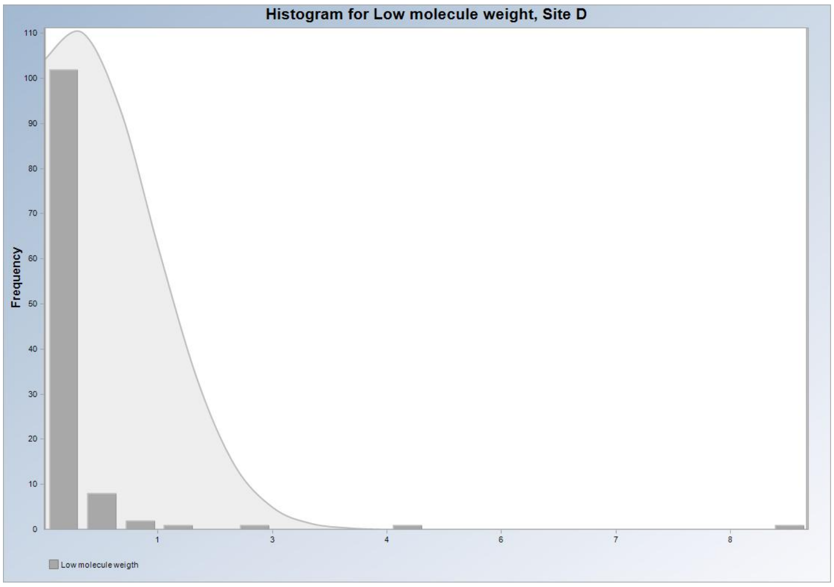
Appendix 3: Histograms

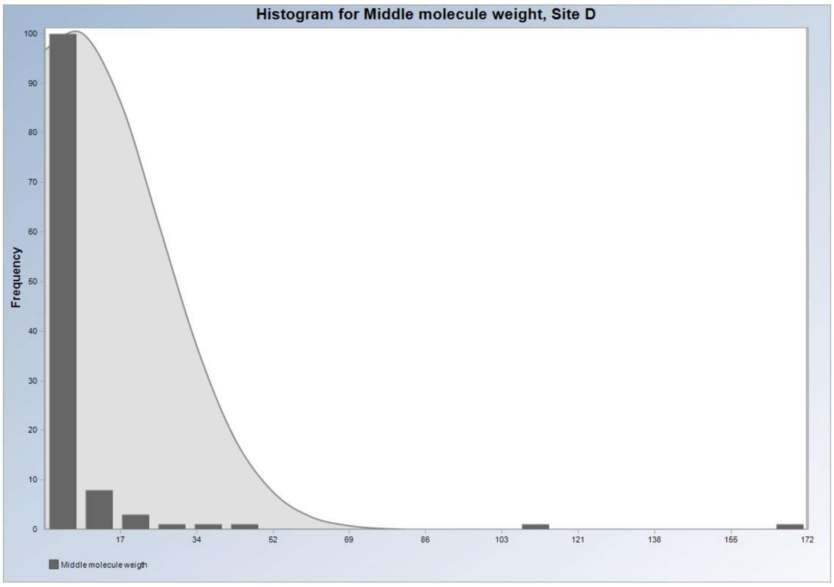

Histogram for Low molecule weight, Site C

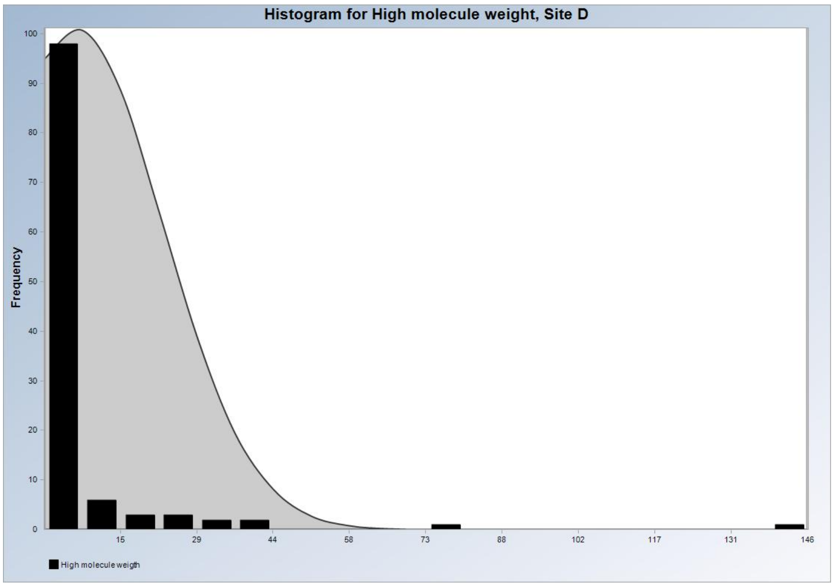

Histogram for Middle molecule weight, Site C


Histogram for High molecule weight, Site C

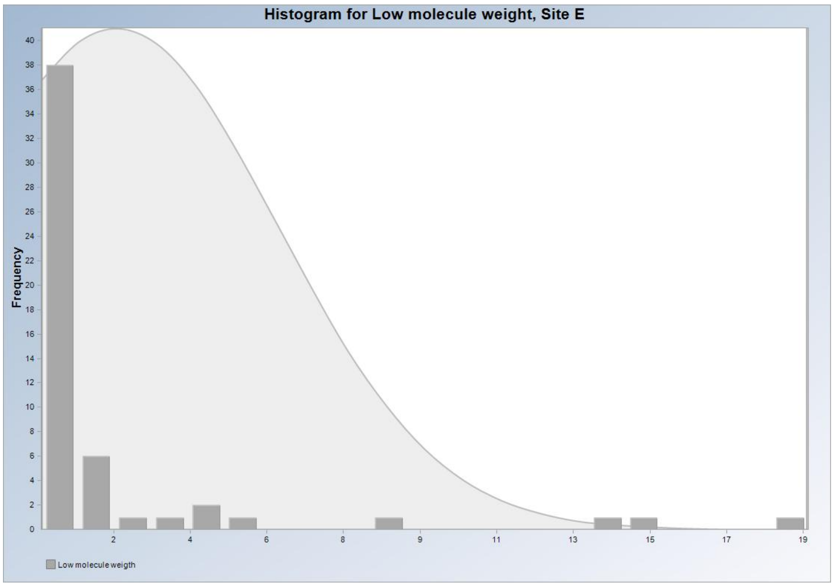
Appendix 3: Histograms

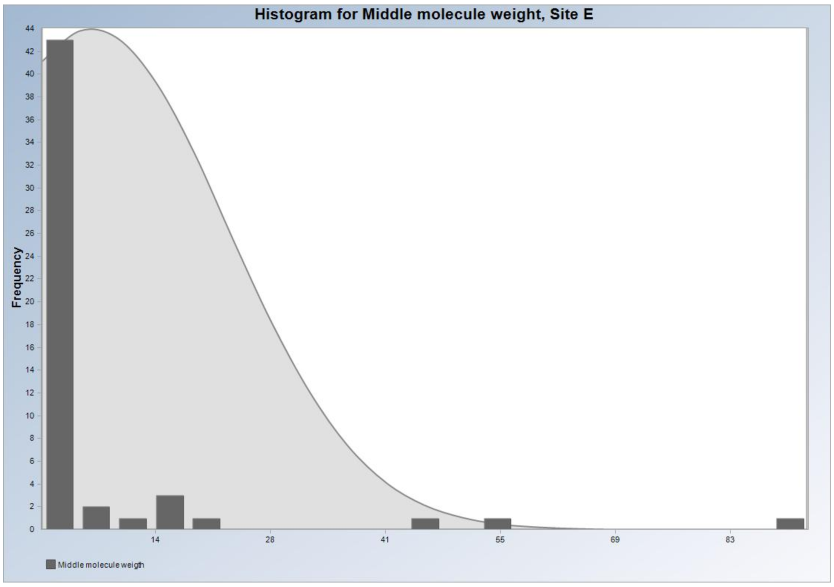


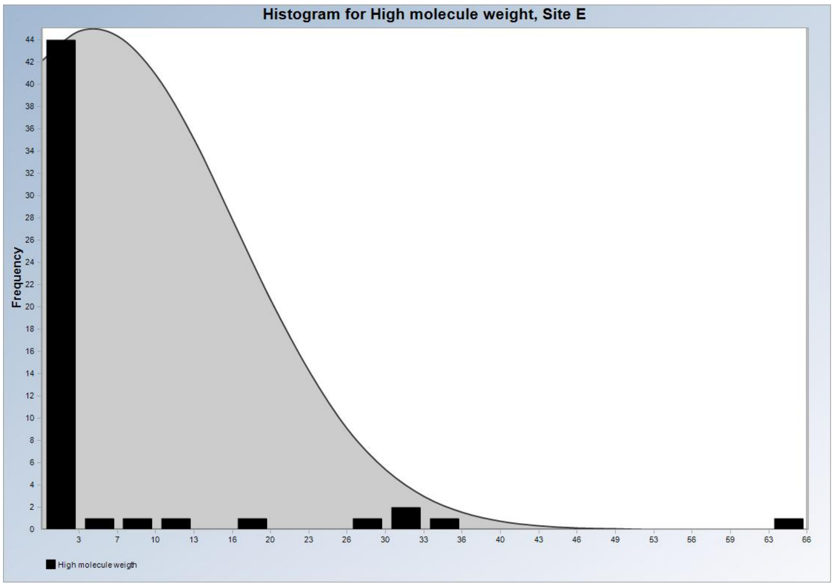




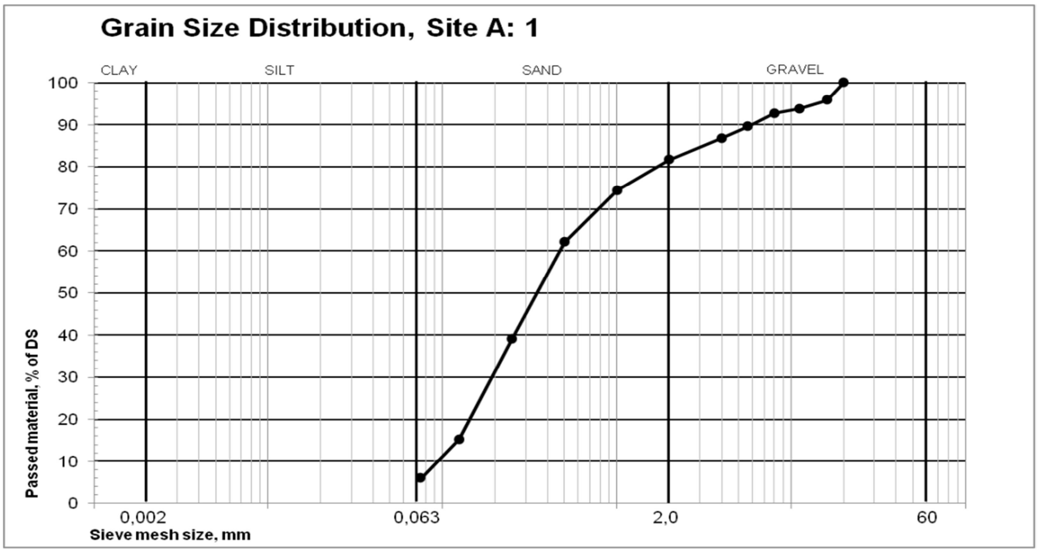
Appendix 3: Histograms

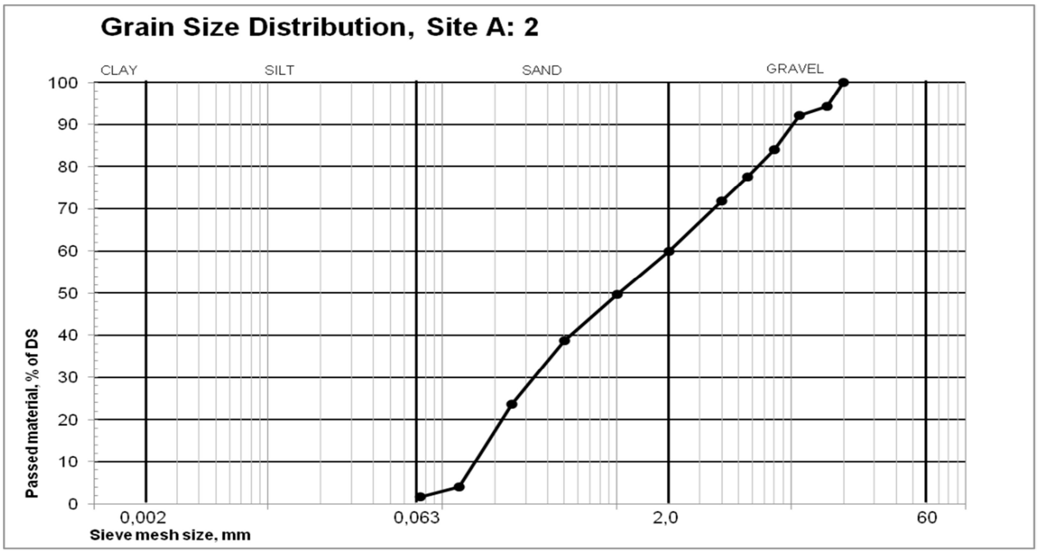

Histogram for Low molecule weight, Site E

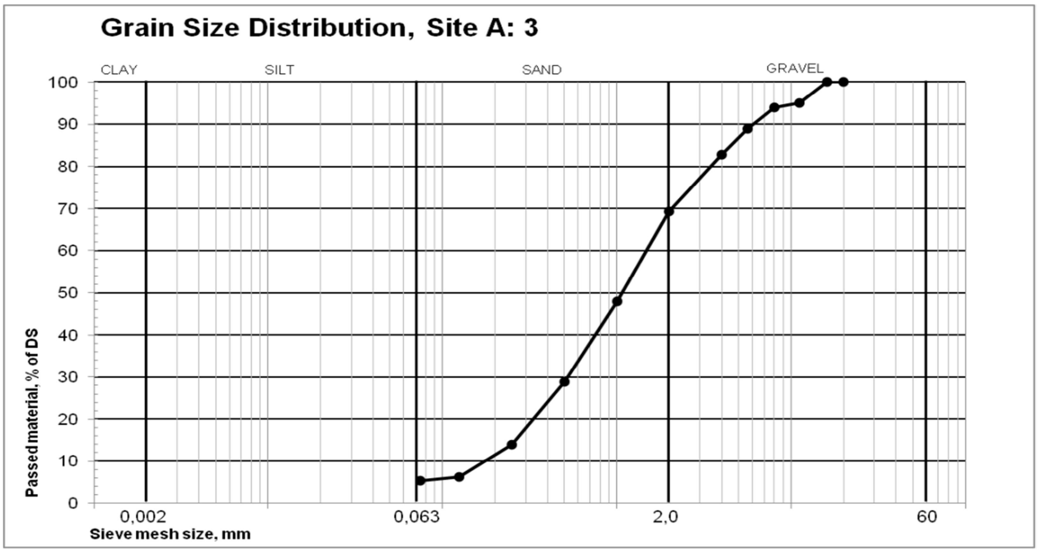

Histogram for Middle molecule weight, Site E


Histogram for High molecule weight, Site E

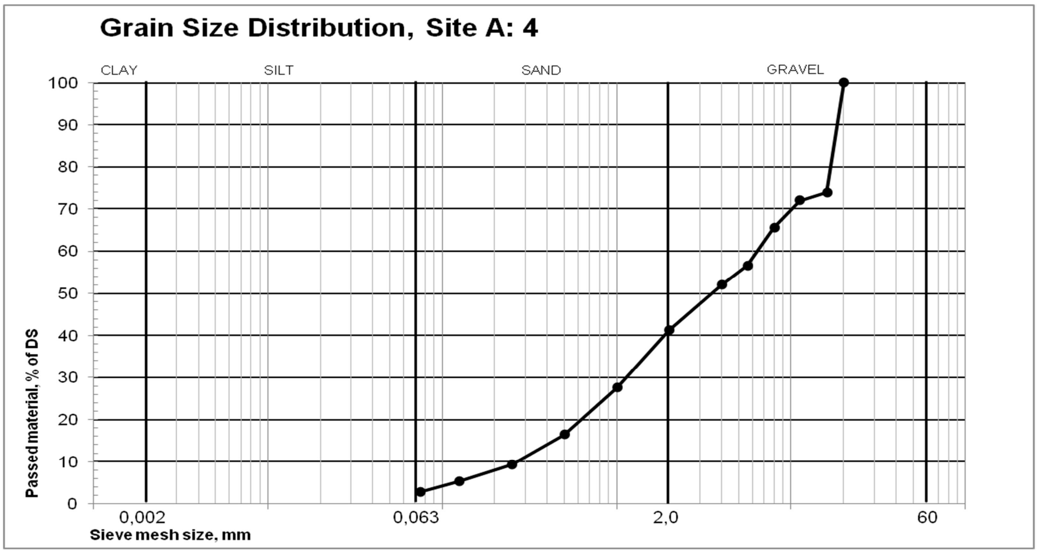
Grain Size Distributions

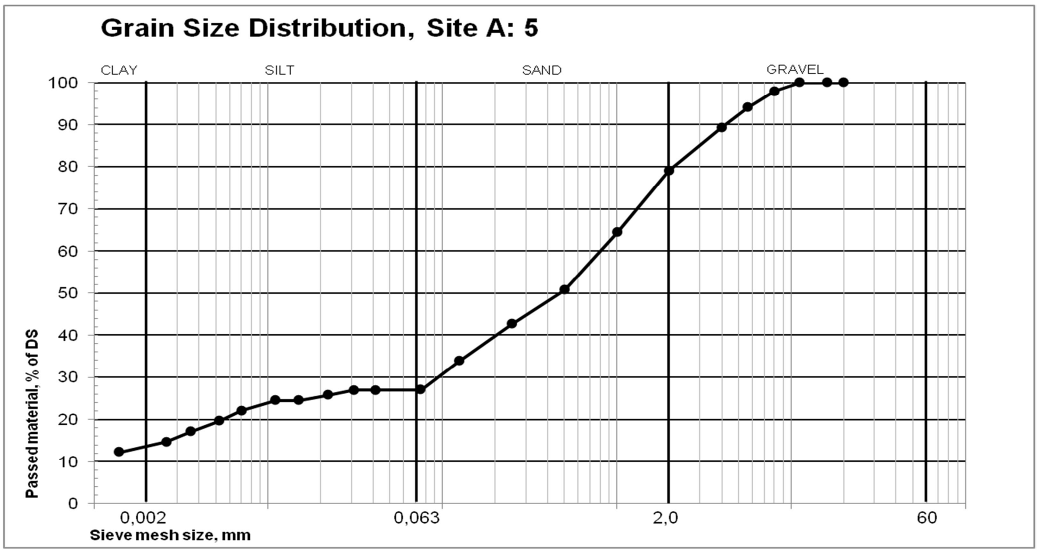


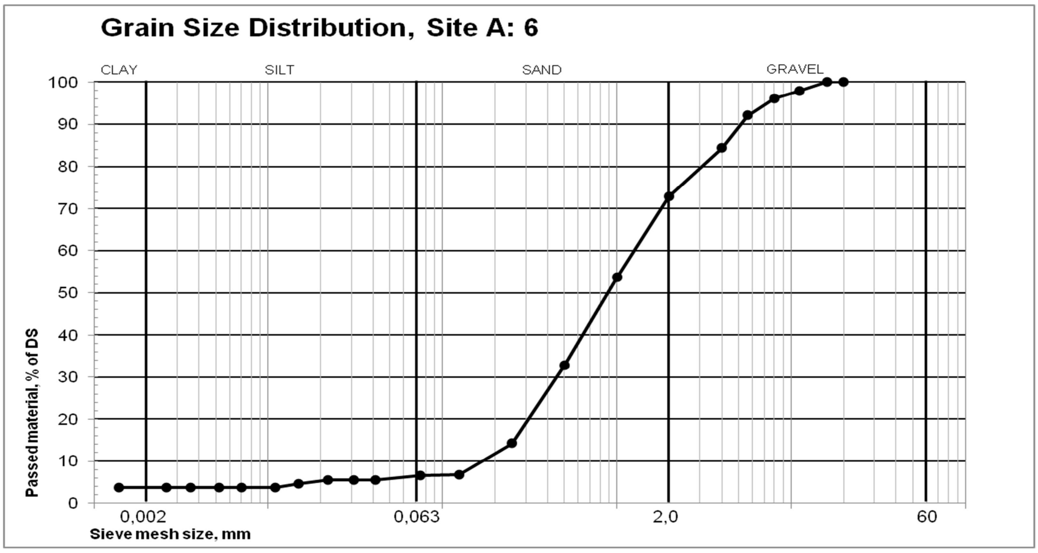




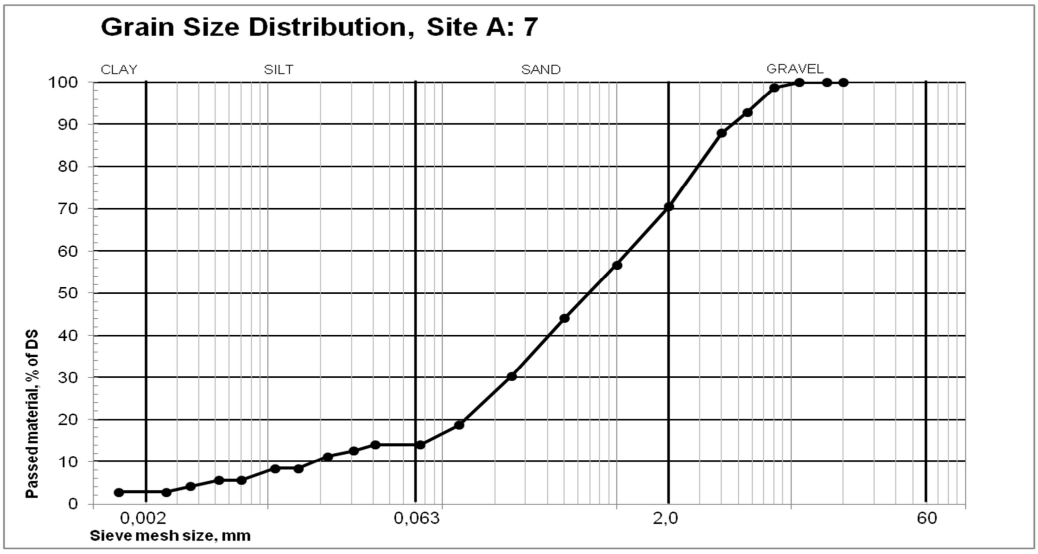
Appendix 4: Grain Size Distributions

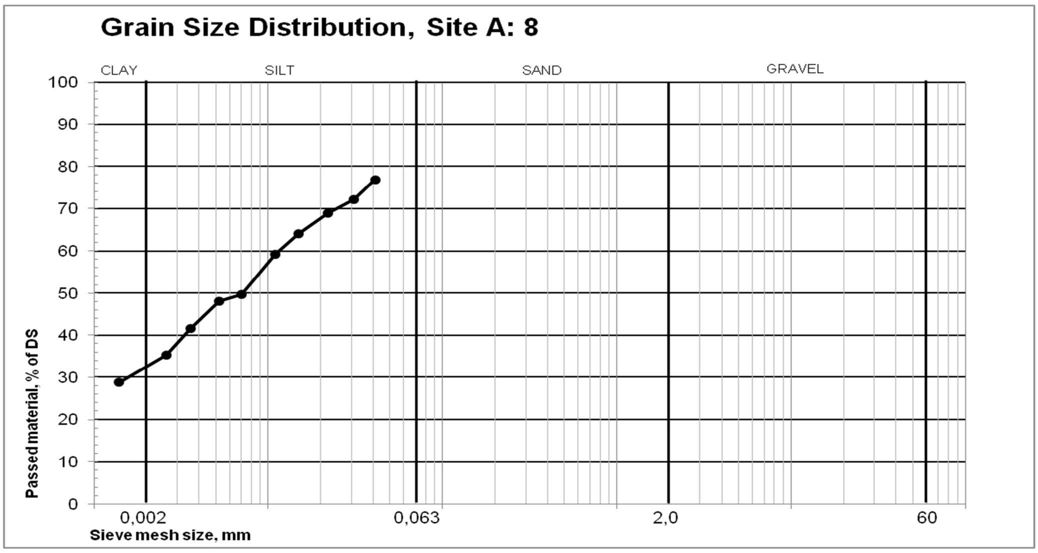


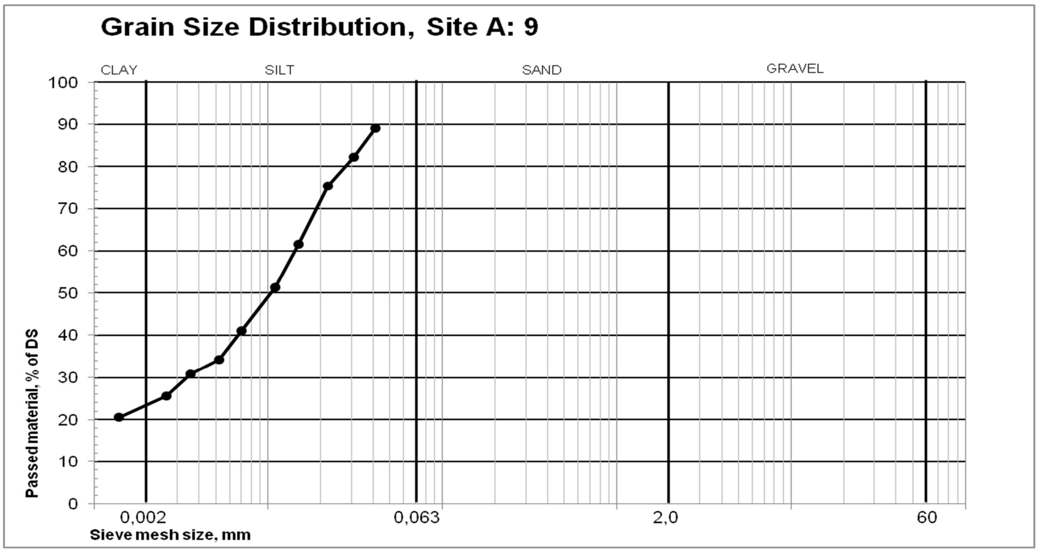




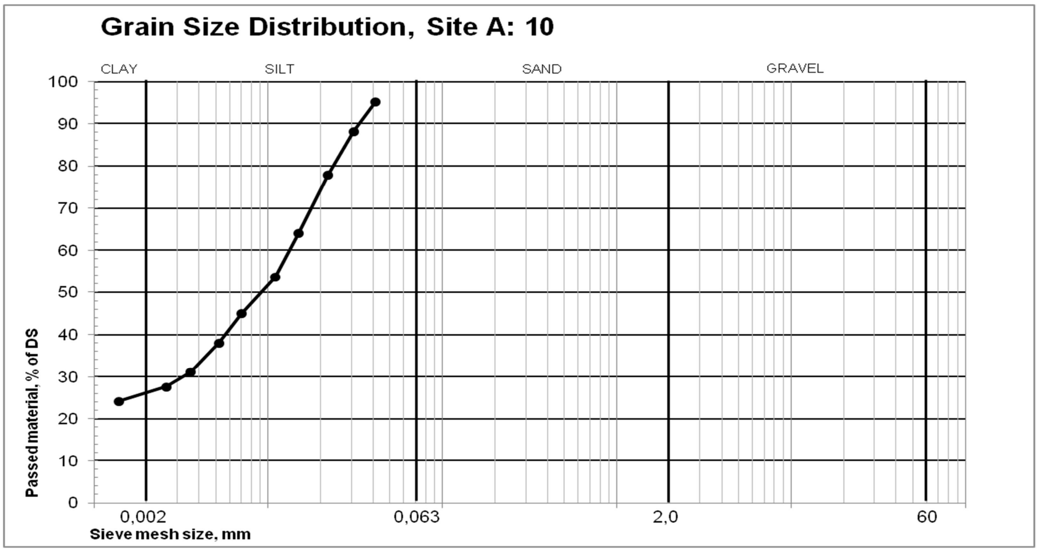
Appendix 4: Grain Size Distributions



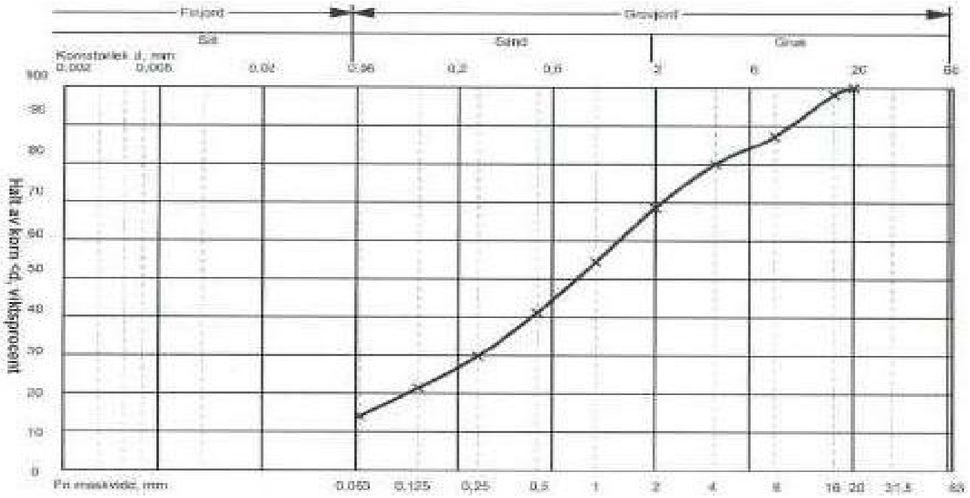




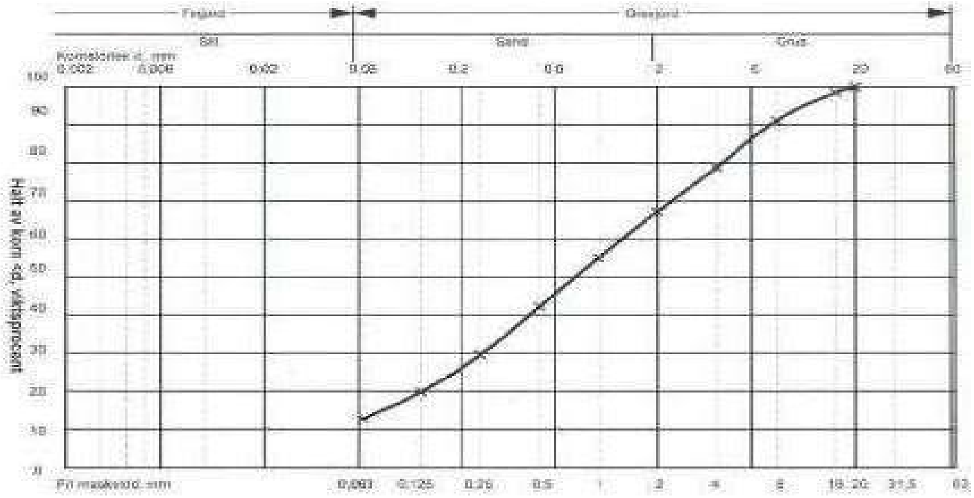
Appendix 4: Grain Size Distributions



Grain size distribution, Site B: 1[1]



Grain size distribution, Site B: *2*

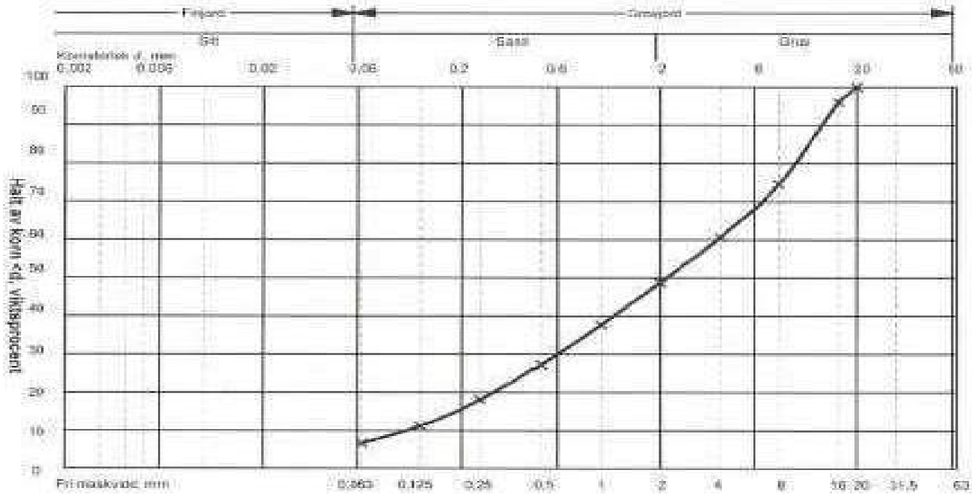


---

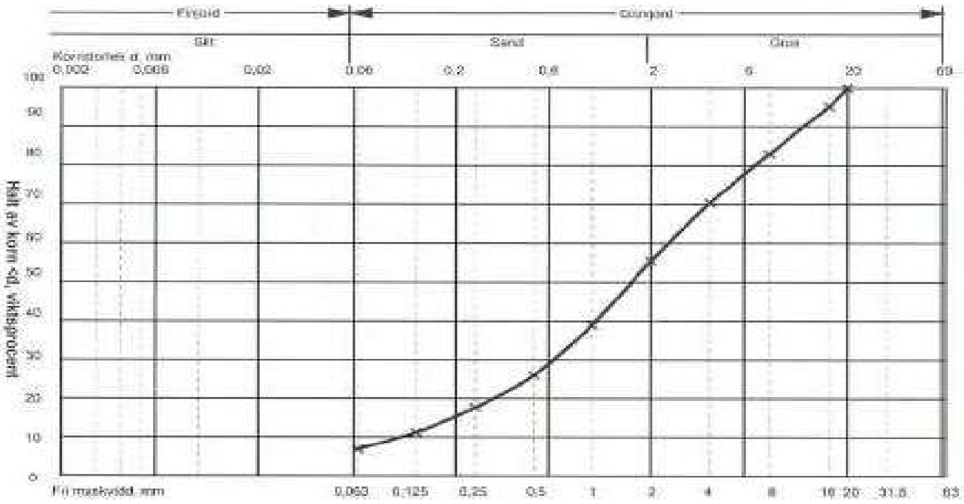
[1] All grain size distributions for Site B are from: Bengtsfors kommun/Geo Innova AB, 2003c, *Sammanställning av resultat från fältundersökningar och laboratorieanalyser (EKA 2002:7)*

Appendix 4: Grain Size Distributions

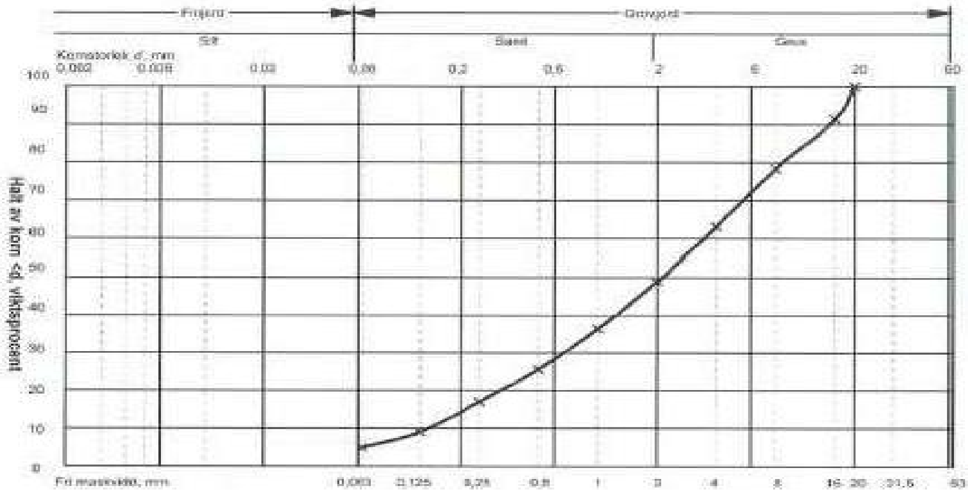
Grain size distribution, Site B: *3*



Grain size distribution, Site B: *4*

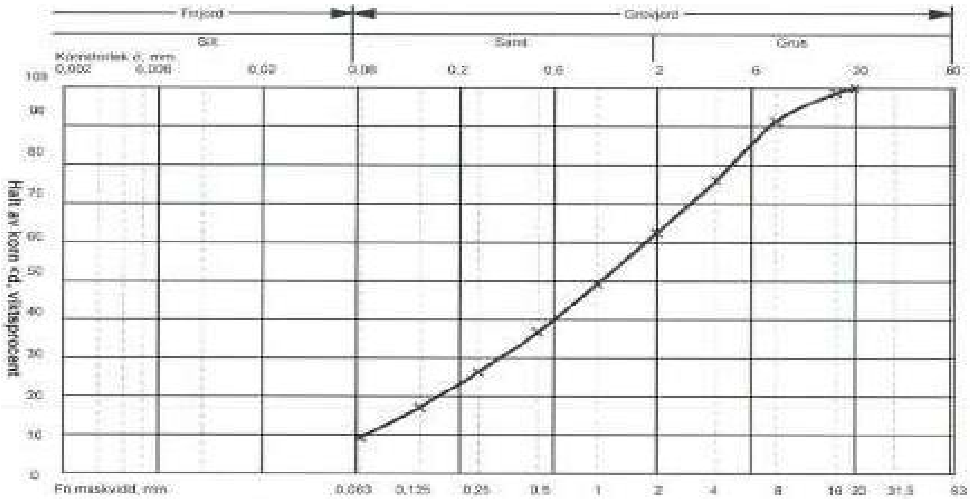


Grain size distribution, Site B: 5

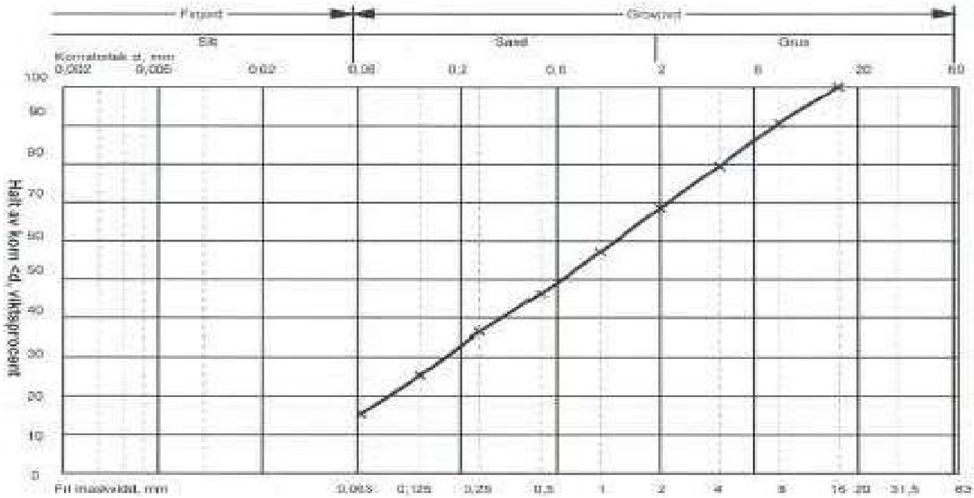


Appendix 4: Grain Size Distributions

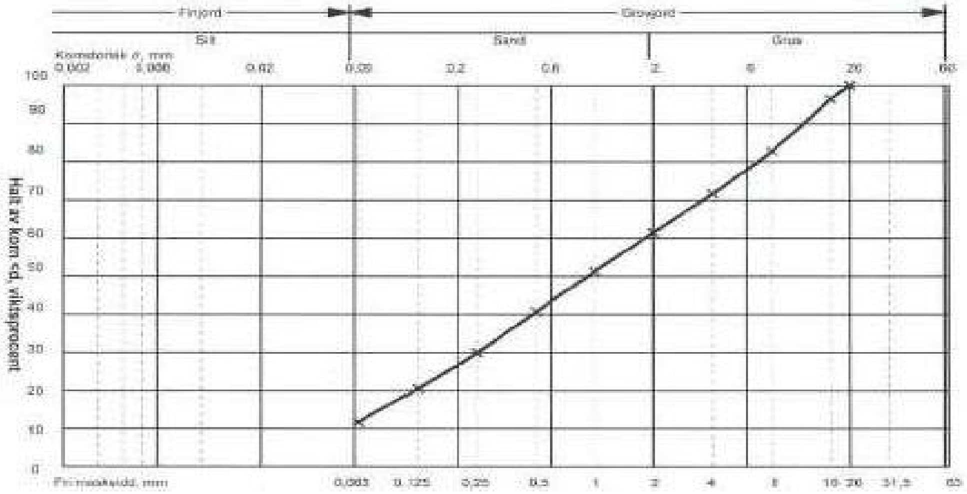
Grain size distribution, Site B: *6*



Grain size distribution, Site B: *7*

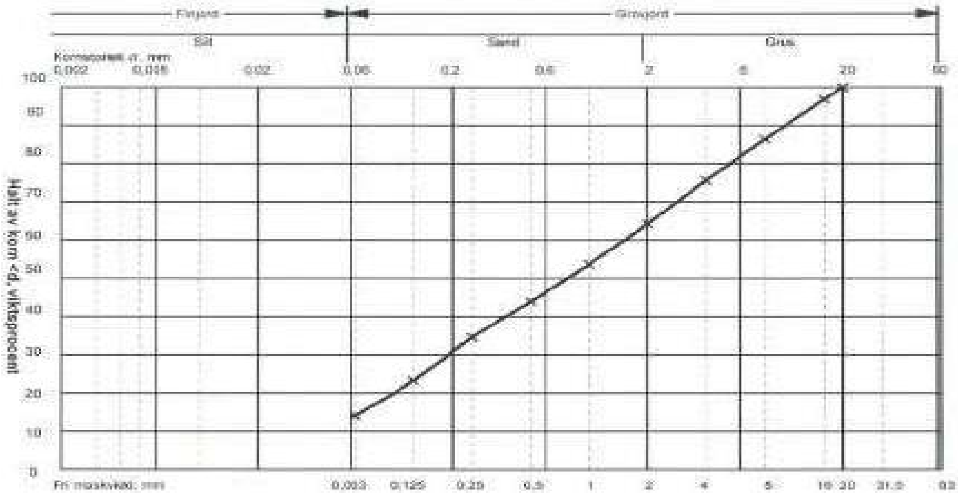


Grain size distribution, Site B: 8

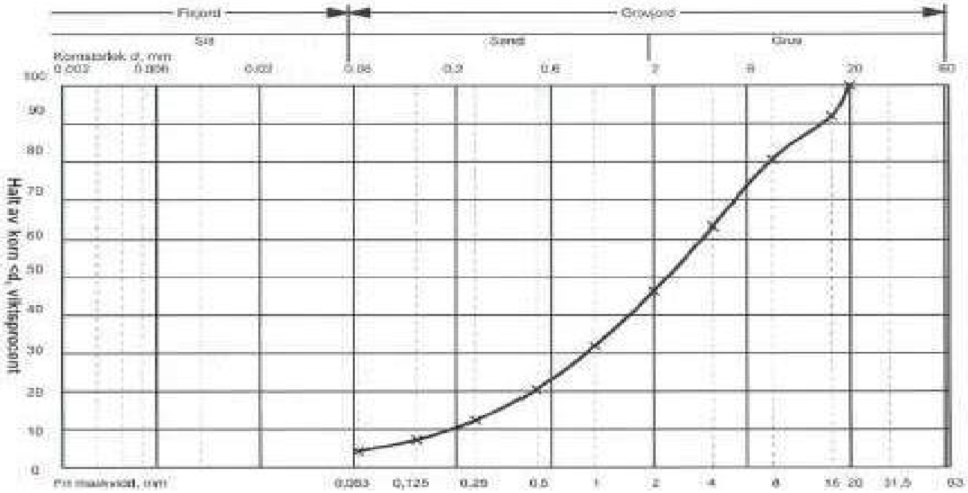


Appendix 4: Grain Size Distributions

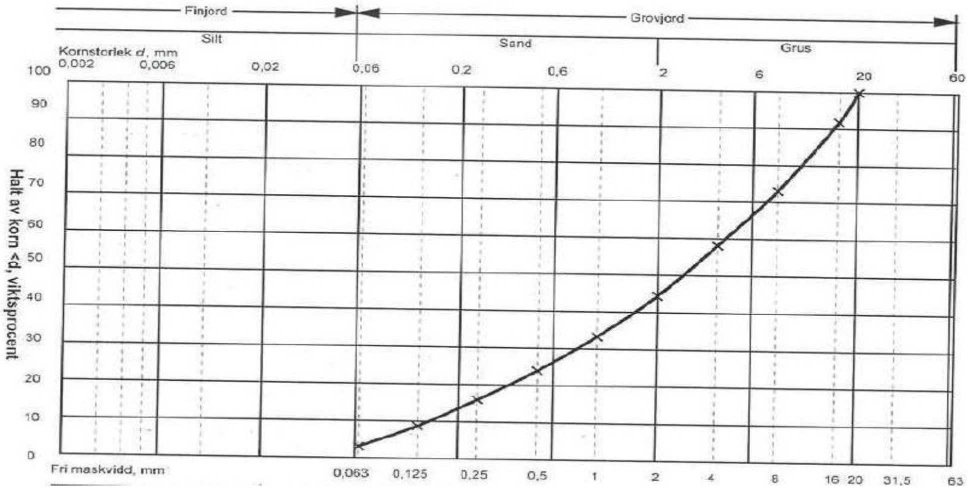
Grain size distribution, Site B: *9*



Grain size distribution, Site B: 10

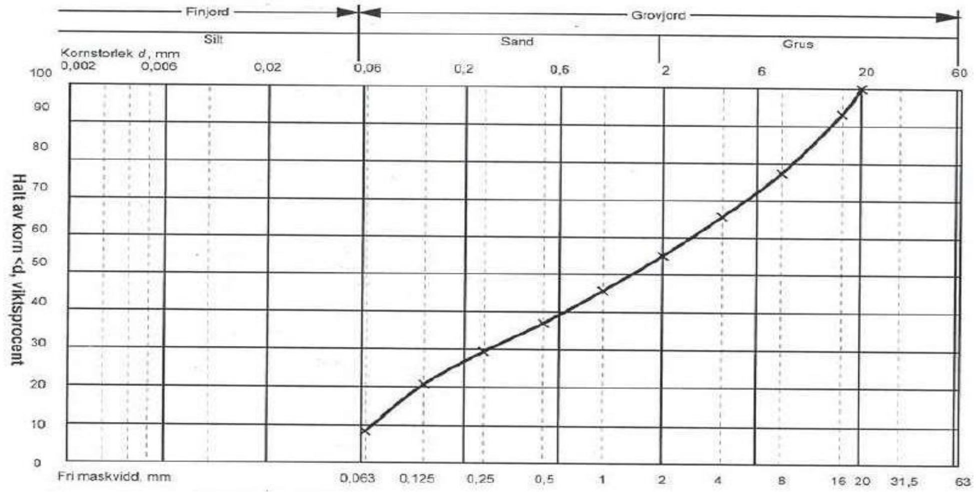


Grain size distribution, Site B: 11

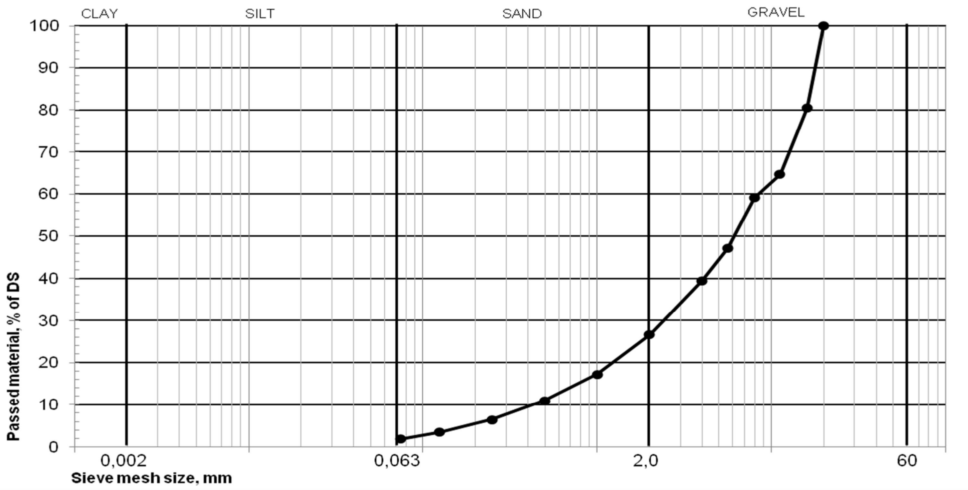


Appendix 4: Grain Size Distributions

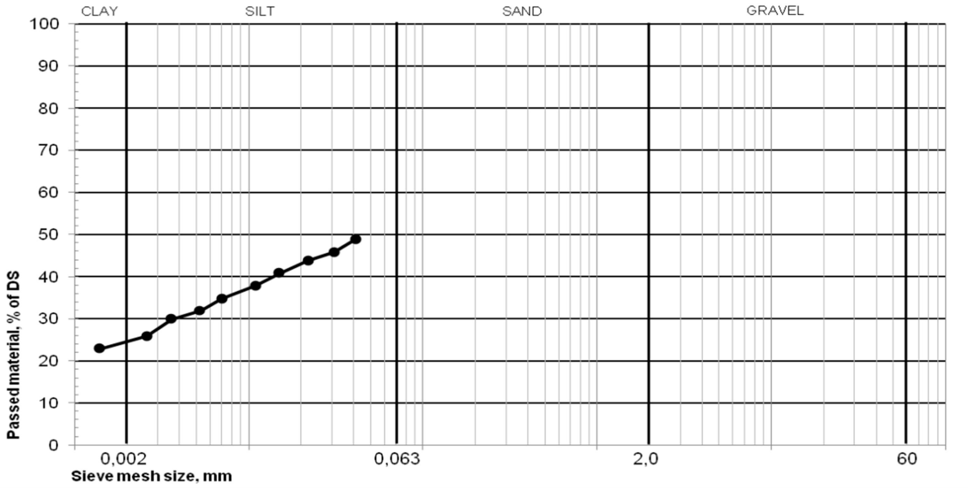
Grain size distribution, Site B: 12



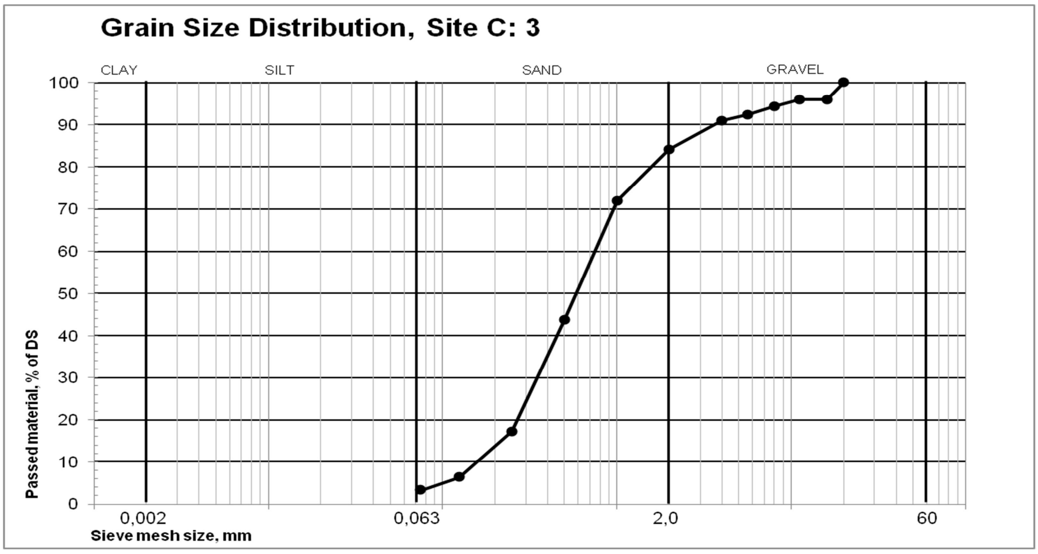
Grain Size Distribution, Site C: 1

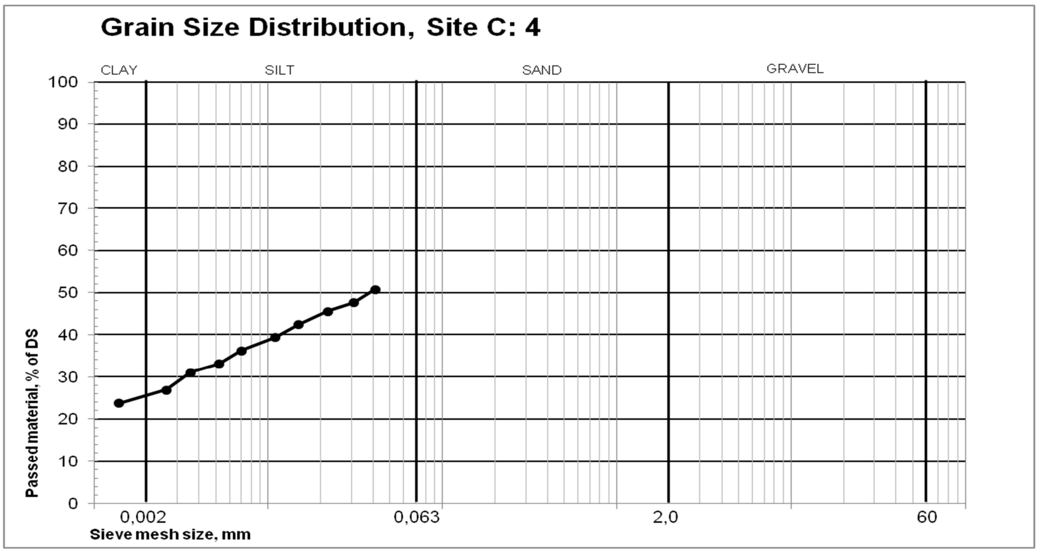


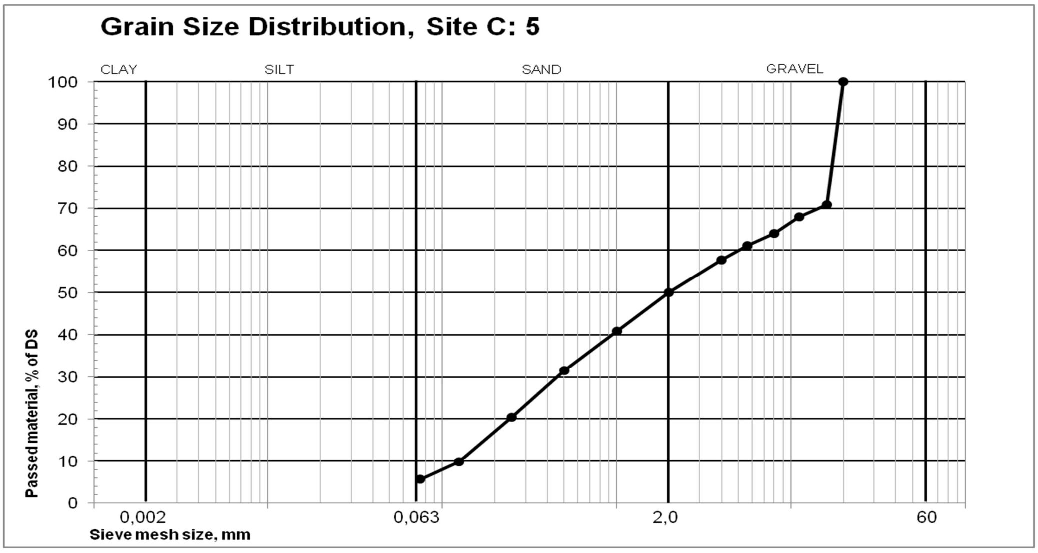
Grain Size Distribution, Site C: 2



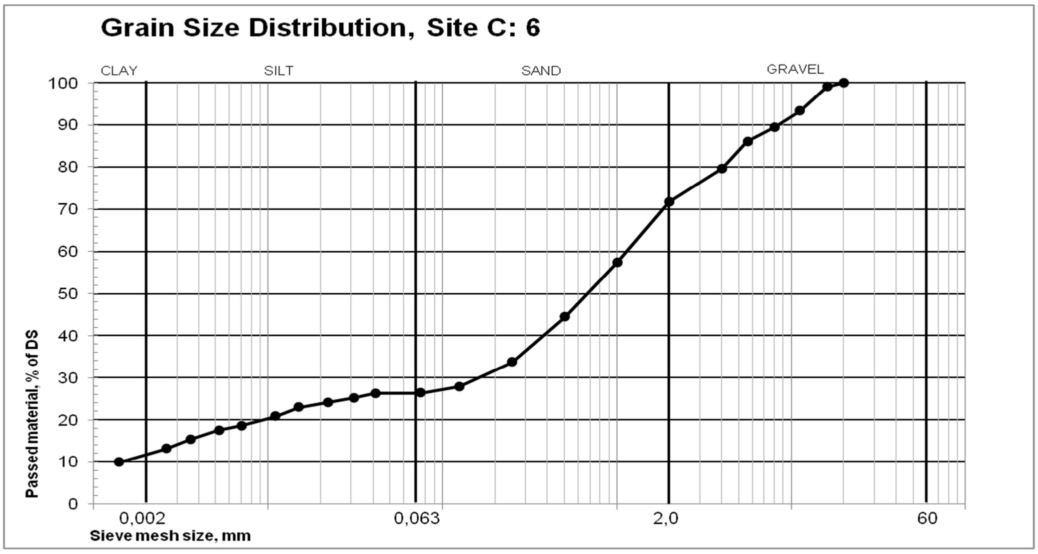
Appendix 4: Grain Size Distributions

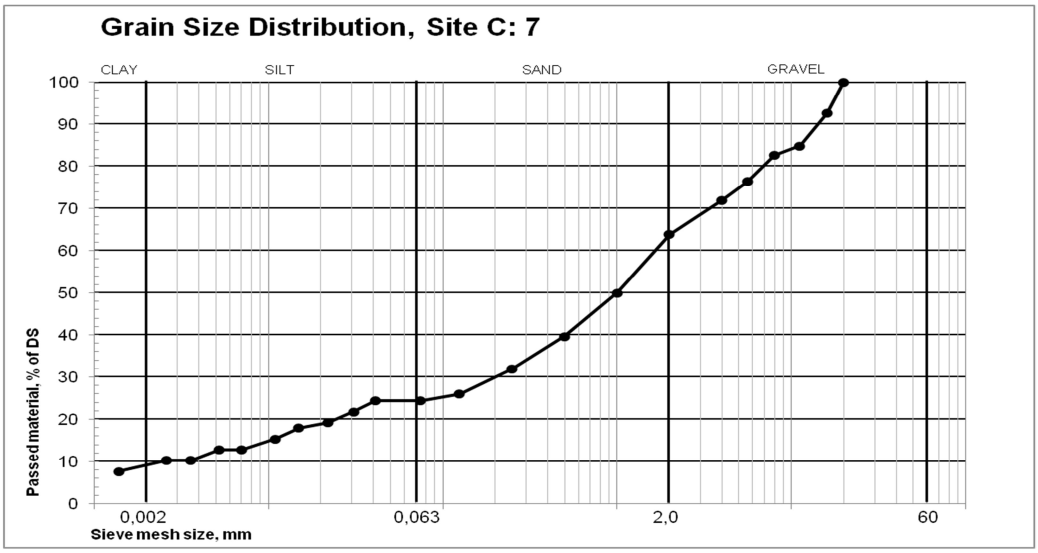


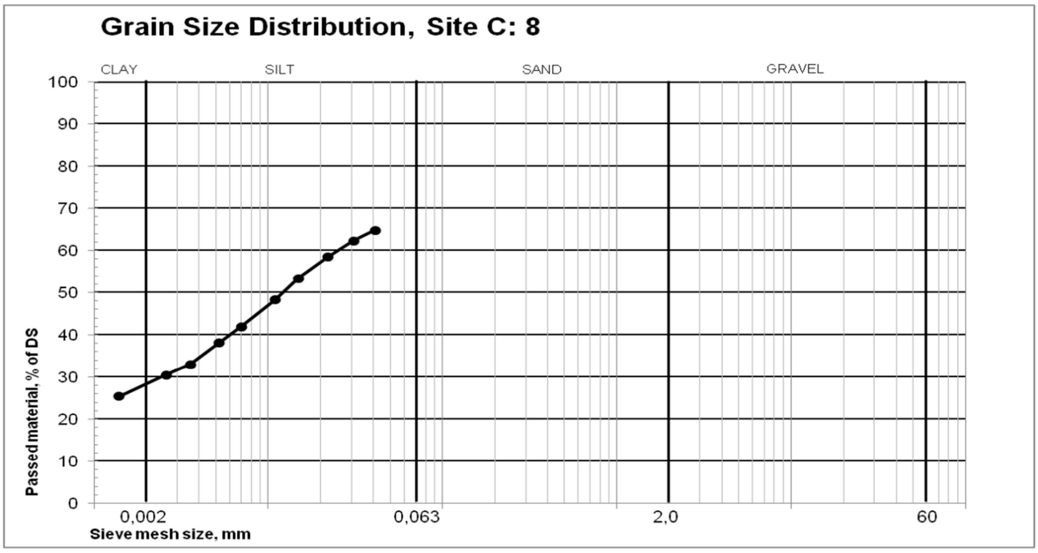




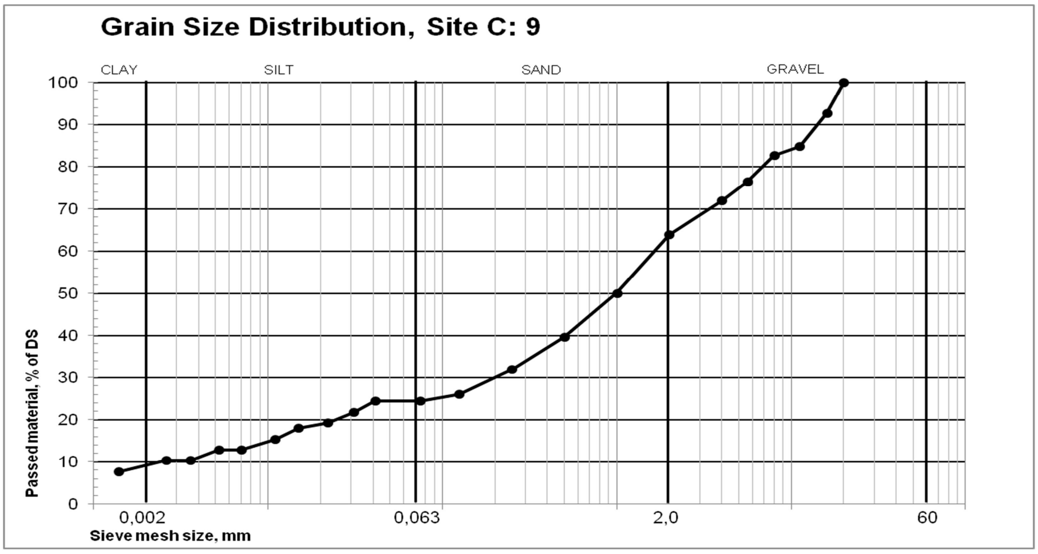
Appendix 4: Grain Size Distributions

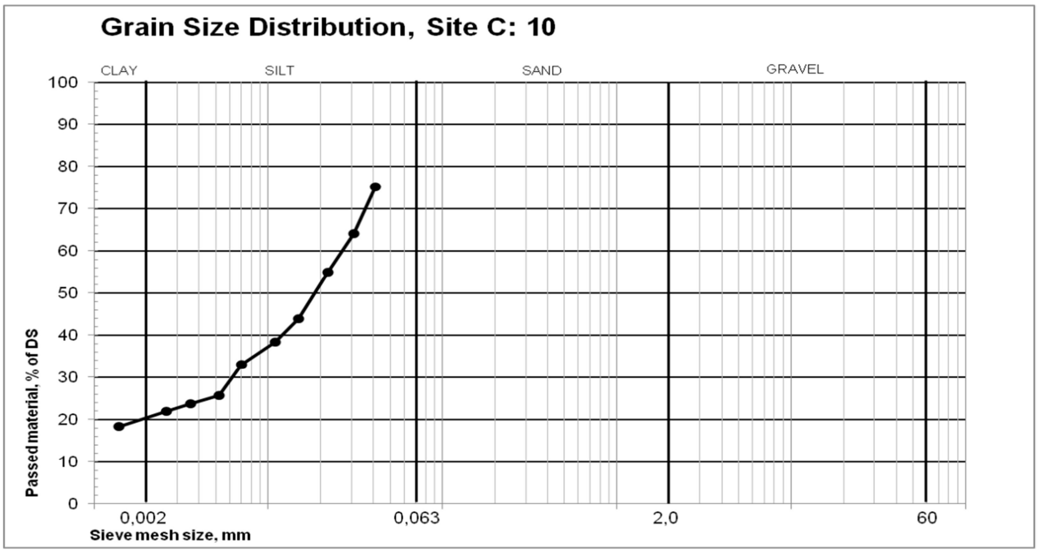


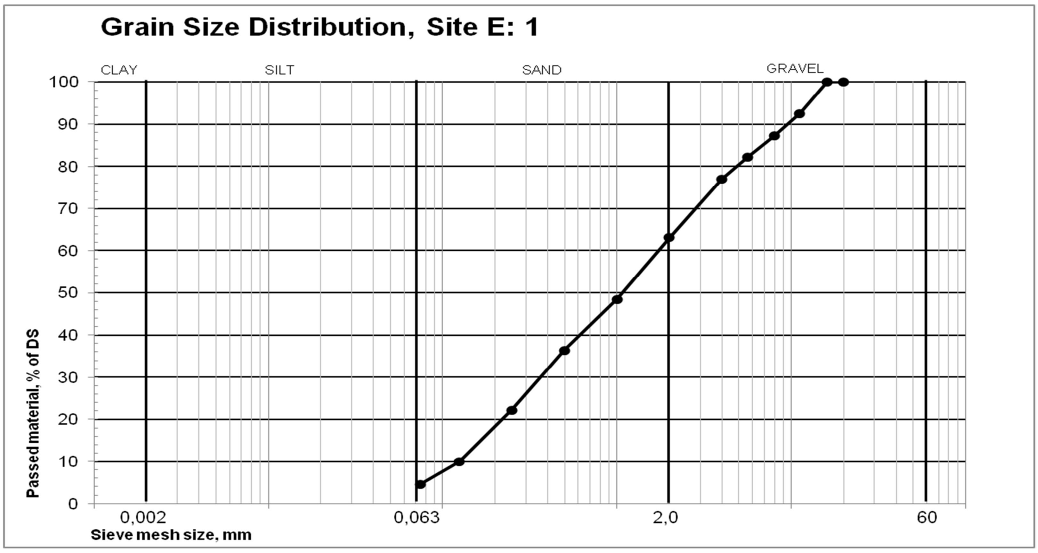




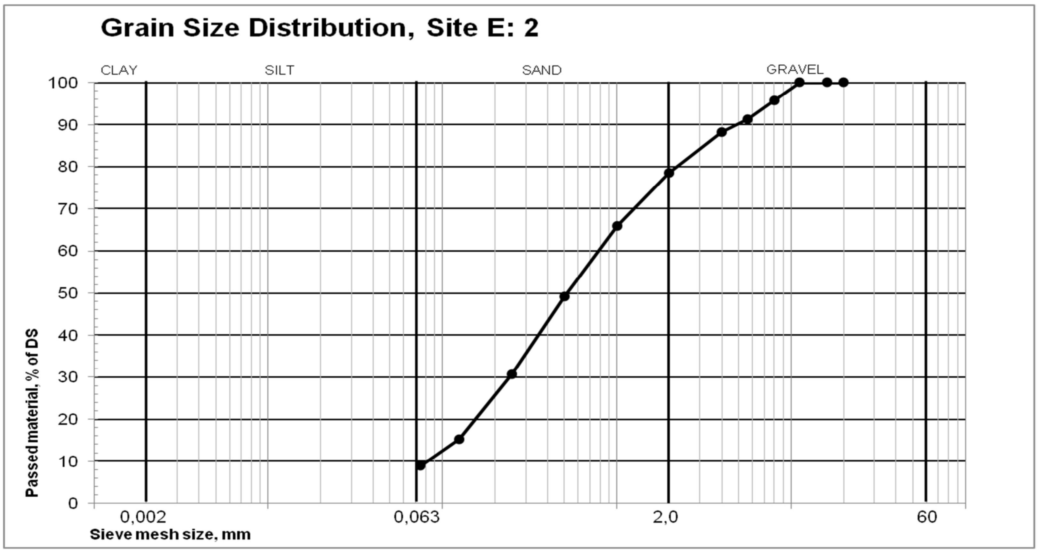
Appendix 4: Grain Size Distributions

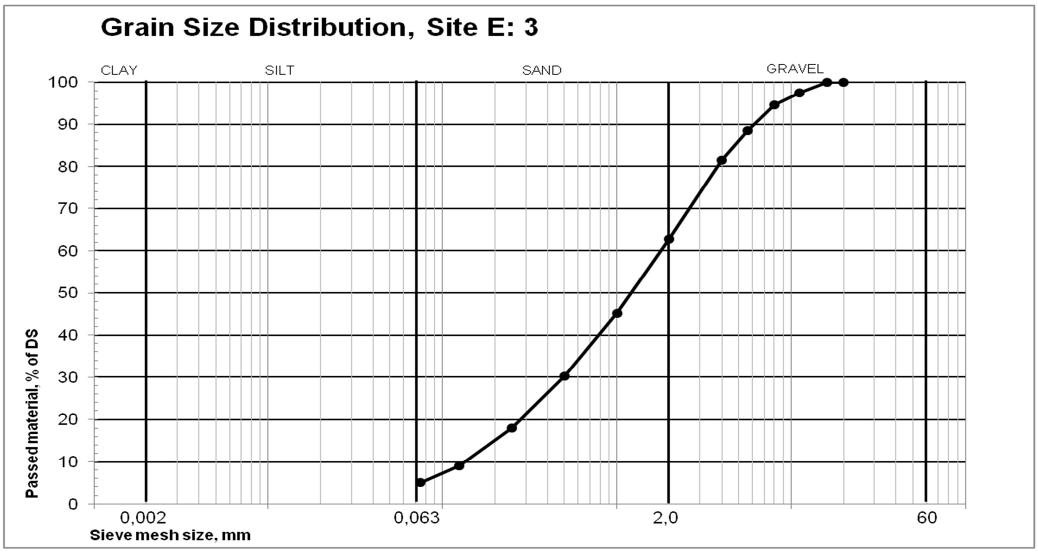


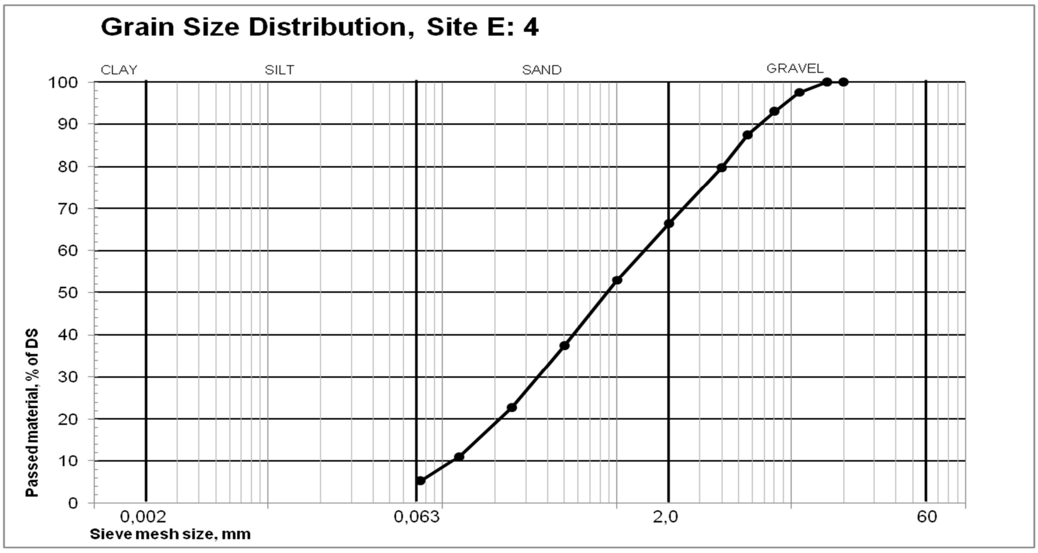




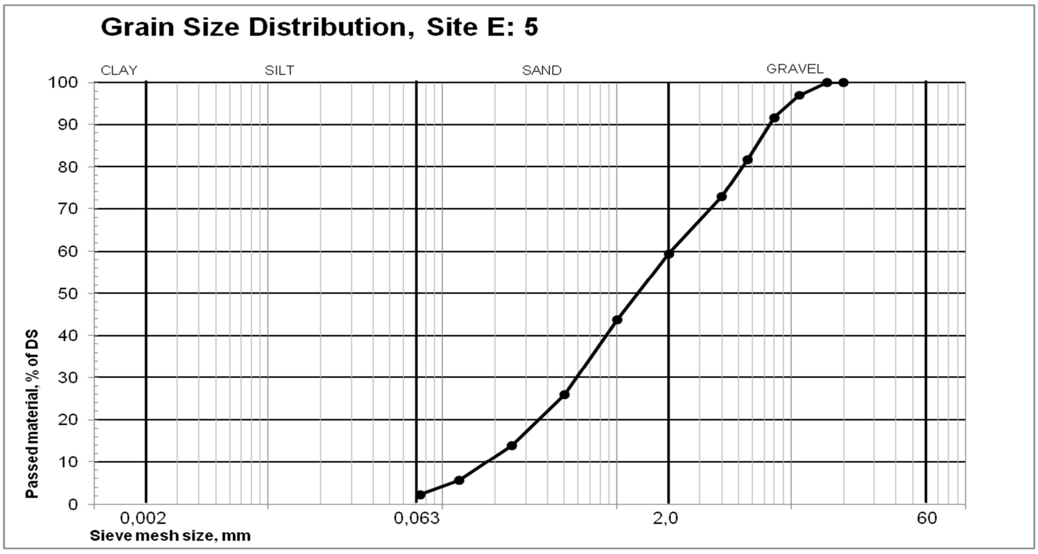
Appendix 4: Grain Size Distributions

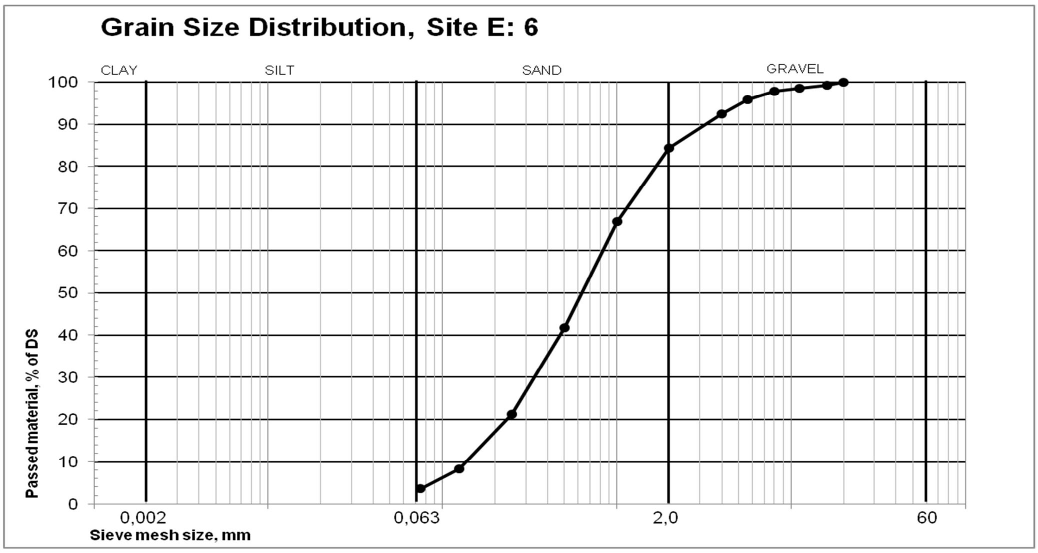


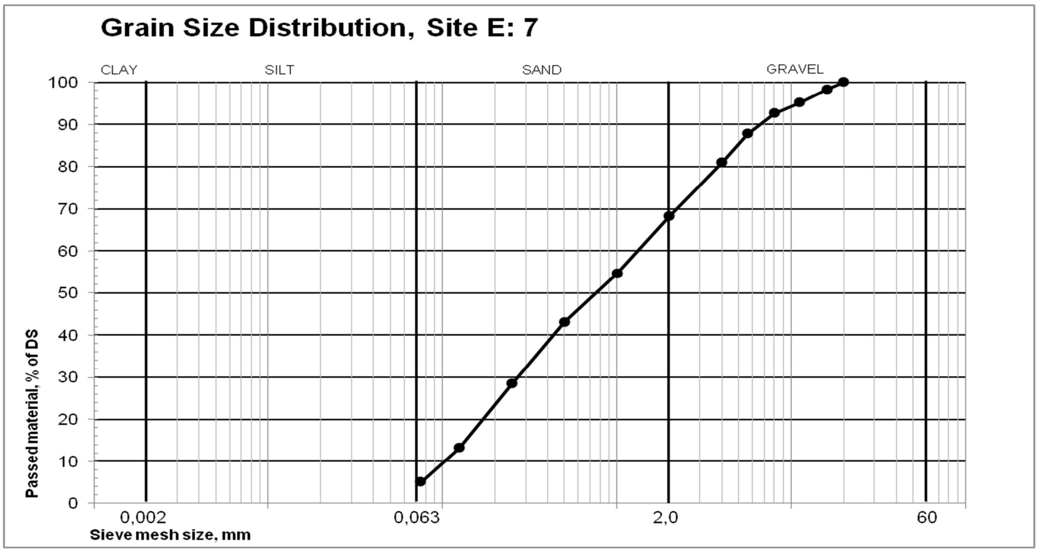




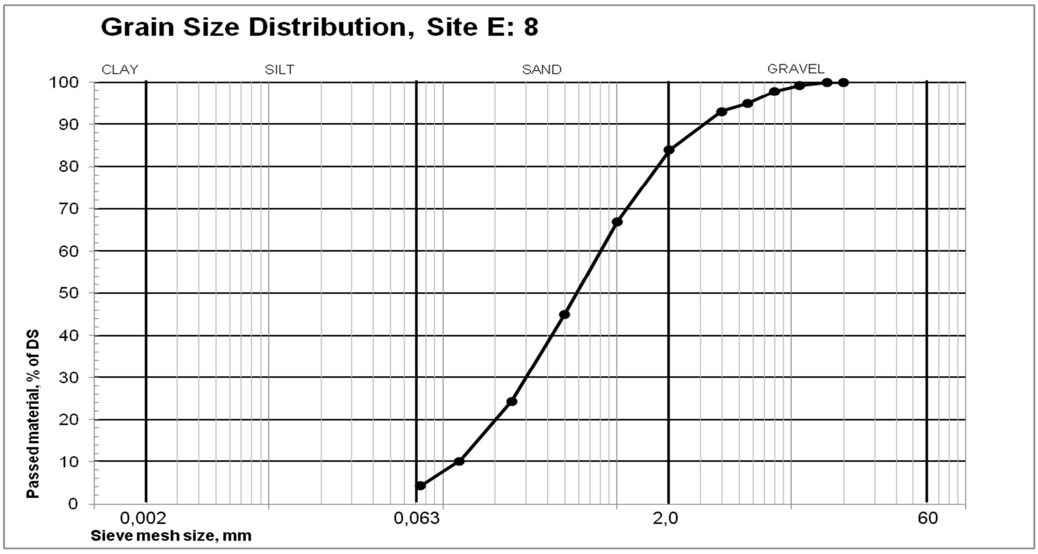
Appendix 4: Grain Size Distributions

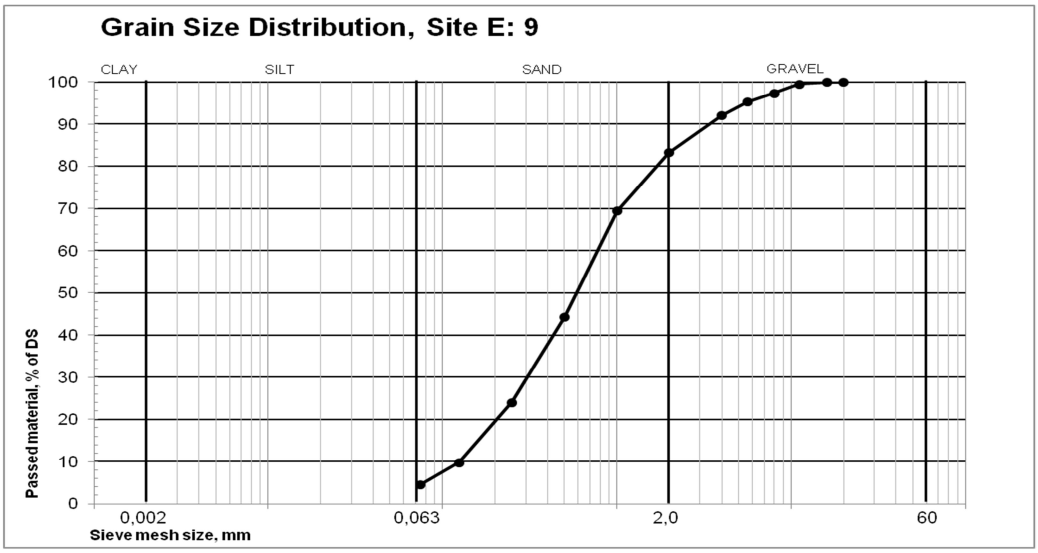

Grain Size Distribution, Site E: 5

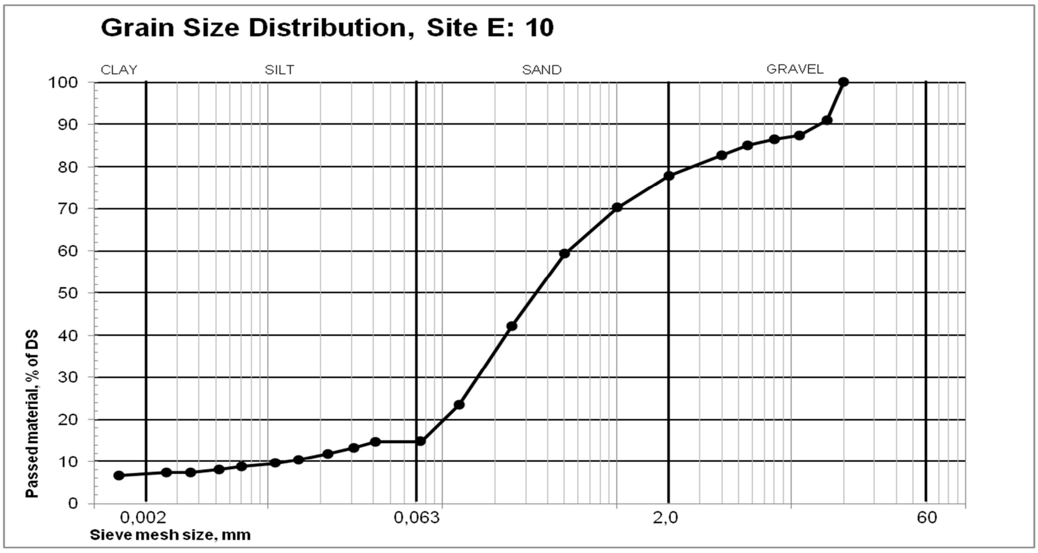

Grain Size Distribution, Site E: 6


Grain Size Distribution, Site E: 7

Appendix 4: Grain Size Distributions







In-situ Sieving, Site E

The weights and percentages for the three different sieving fractions at Site E are presented in Table A. The sieving fractions are[1]:

- Fine fraction, 0 – 10 mm
- Intermediate fraction, 10 – 40 mm
- Coarse fraction, Above 40 mm

Table A: Weights and percentages from in-situ sieving at Site E.[2]

	Weight (Mg)			**Percent**		
Soil Sample	*Fine*	*Intermediate*	*Coarse*	*Fine*	*Intermediate*	*Coarse*
Site E: 1	126,5	0	47	73 %	0 %	27 %
Site E: 3	110	13	41	67 %	8 %	25 %
Site E: 4	110	13	41	67 %	8 %	25 %
Site E: 5	43	0	0	100 %	0 %	0 %
Site E: 6	153	33	58	63 %	14 %	24 %
Site E: 7	76	0	44	63 %	0 %	37 %
Site E: 8	99	22	42	61 %	13 %	26 %
Site E: 9	81	24	50	52 %	15 %	32 %
Site E: 10	110	0	32	77 %	0 %	23 %
Mean	*100,94*	*11,67*	*39,44*	*69 %*	*6 %*	*24 %*

---

[1] Johannes Senning, NCC Teknik, telephone conversation 2011-12-02

[2] NCC Construction, Daily reports for Hexion - September and August 2011, Project 4532162

Sensitivity Analysis

The calculation underlying the numerical sensitivity analysis is presented in Table A.

Table A: The numerical sensitivity analysis for the stratified random sampling schemes.

Trial	n	w1	w2	w3	s1	s2	s3	sxp	sxo	sxp/sxo
n, 1	10	0,33	0,33	0,34	1	5	10	4,26	2,89	1,47
n, 2	20	0,33	0,33	0,34	1	5	10	2,13	1,45	1,47
n, 3	30	0,33	0,33	0,34	1	5	10	1,42	0,96	1,47
n, 4	40	0,33	0,33	0,34	1	5	10	1,06	0,72	1,47
n, 5	50	0,33	0,33	0,34	1	5	10	0,85	0,58	1,47
s, 1	20	0,33	0,33	0,34	1	1	10	1,73	0,82	2,10
s, 2	20	0,33	0,33	0,34	1	10	100	171,67	70,80	2,42
s, 3	20	0,33	0,33	0,34	100	100	1	330,02	220,05	1,50
s, 4	20	0,33	0,33	0,34	5	10	10	3,72	3,49	1,08
s, 5	20	0,33	0,33	0,34	1	1	100	170,03	60,07	2,83
w, 1	20	0,80	0,1	0,1	1	5	10	0,67	0,26	2,51
w, 2	20	0,60	0,2	0,2	1	5	10	1,28	0,65	1,98
w, 3	20	0,50	0,3	0,2	1	5	10	1,40	0,80	1,75
w, 4	20	0,40	0,4	0,2	1	5	10	1,52	0,97	1,57
w, 5	20	0,20	0,6	0,2	1	5	10	1,76	1,35	1,30
s, w, 1	20	0,80	0,1	0,1	1	1	10	0,55	0,18	3,02
s, w, 2	20	0,60	0,2	0,2	1	10	100	101,03	25,54	3,96
s, w, 3	20	0,50	0,3	0,2	100	100	1	400,01	321,60	1,24
s, w, 4	20	0,40	0,4	0,2	5	10	10	3,50	3,20	1,09
s, w, 5	20	0,20	0,6	0,2	1	1	100	100,04	21,63	4,62
s, w, 6	20	0,80	0,1	0,1	1	1	10	0,55	0,18	3,02
s, w, 7	20	0,60	0,2	0,2	10	1	100	103,04	34,32	3,00
s, w, 8	20	0,50	0,3	0,2	100	100	1	400,00	321,60	1,24
s, w, 9	20	0,40	0,4	0,2	10	5	10	3,50	3,20	1,09
s, w, 10	20	0,20	0,6	0,2	1	1	100	100,00	21,63	4,62

Basic Statistical Parameters

Table A: Number of soil samples, mean values and standard deviations for Site A.

STATISTICS	n	x_L	x_M	x_H	s_L	s_M	s_H
CLASS 1	39	1,14	17,19	27,03	1,65	28,99	39,95
CLASS 2	44	2,20	27,67	38,28	8,72	91,24	114,39
CLASS 3	20	2,14	21,16	28,84	3,02	30,90	55,05
CLASS 4	10	0,48	6,26	7,68	0,41	7,14	9,16
CLASS 5	7	0,49	10,49	13,49	1,06	23,78	31,76
ENTIRE AREA	120	1,60	20,39	29,05	5,51	59,22	76,58

Table B: Coefficient of variation and variance for Site A.

STATISTICS	C_L	C_M	C_H	s^2_L	s^2_M	s^2_H
CLASS 1	1,45	1,69	1,48	2,73	840,47	1595,65
CLASS 2	3,96	3,30	2,99	76,08	8325,48	13084,35
CLASS 3	1,41	1,46	1,91	9,10	955,00	3030,13
CLASS 4	0,86	1,14	1,19	0,17	50,93	83,96
CLASS 5	2,19	2,27	2,35	1,13	565,33	1008,74
ENTIRE AREA	3,43	2,90	2,64	30,32	3507,20	5863,97

Table C: Number of soil samples, mean values and standard deviations for Site B.

STATISTICS	n	x_L	x_M	x_H	s_L	s_M	s_H
CLASS 1	1	20,80	216,70	219,30	-	-	-
CLASS 2	40	2,66	28,80	16,55	12,22	134,25	67,49
CLASS 3	6	0,80	8,26	8,47	1,19	11,34	10,57
CLASS 4	0	-	-	-	-	-	-
CLASS 5	0	-		-	-	-	-
ENTIRE AREA	47	2,81	30,18	21,23	11,59	126,94	68,96

Table D: Coefficient of variation and variance for Site B.

STATISTICS	C_L	C_M	C_H	s^2_L	s^2_M	s^2_H
CLASS 1						
CLASS 2	4,58	4,66	4,08	149,23	18022,52	4554,23
CLASS 3	1,49	1,37	1,25	1,41	128,52	111,72
CLASS 4						
CLASS 5						
ENTIRE AREA	4,12	4,21	3,25	134,26	16114,55	4755,44

Appendix 7: Basic Statistical Parameters

Table E: Number of soil samples, mean values and standard deviations for Site C.

STATISTICS	**n**	**x_L**	**x_M**	**x_H**	**s_L**	**s_M**	**s_H**
CLASS 1	0	0	0	0	0	0	0
CLASS 2	8	0,22	1,75	1,59	0,26	3,09	2,32
CLASS 3	16	1,53	19,46	15,50	4,80	63,53	44,67
CLASS 4	12	0,70	11,92	18,81	0,85	17,64	41,50
CLASS 5	4	0,08	0,52	0,92	0,08	0,50	0,86
ENTIRE AREA	40	0,87	11,76	12,25	3,07	41,23	36,17

Table F: Coefficient of variation and variance for Site C.

STATISTICS	**C_L**	**C_M**	**C_H**	**s^2_L**	**s^2_M**	**s^2_H**
CLASS 1	0,00	0,00	0,00	0,00	0,00	0,00
CLASS 2	1,20	1,77	1,46	0,07	9,56	5,38
CLASS 3	3,13	3,26	2,88	23,00	4035,75	1995,00
CLASS 4	1,22	1,48	2,21	0,73	311,29	1722,10
CLASS 5	1,09	0,97	0,93	0,01	0,25	0,73
ENTIRE AREA	3,51	3,50	2,95	9,40	1699,63	1308,08

Table G: Number of soil samples, mean values and standard deviations for Site D.

STATISTICS	**n**	**x_L**	**x_M**	**x_H**	**s_L**	**s_M**	**s_H**
CLASS 1	4	0,31	0,42	0,55	0,05	0,13	0,07
CLASS 2	87	0,31	6,62	6,80	1,04	22,69	19,14
CLASS 3	2	0,08	0,22	0,20	0,01	0,20	0,17
CLASS 4	15	0,77	6,95	5,96	1,08	12,25	8,28
CLASS 5	8	0,22	2,03	2,40	0,12	1,99	2,31
ENTIRE AREA	116	0,36	5,97	6,01	0,98	20,06	16,82

Table H: Coefficient of variation and variance for Site D.

STATISTICS	**C_L**	**C_M**	**C_H**	**s^2_L**	**s^2_M**	**s^2_H**
CLASS 1	0,17	0,30	0,13	0,00	0,02	0,01
CLASS 2	3,31	3,43	2,82	1,08	514,84	366,48
CLASS 3	0,09	0,90	0,85	0,00	0,04	0,03
CLASS 4	1,39	1,76	1,39	1,16	150,06	68,58
CLASS 5	0,54	0,98	0,97	0,01	3,94	5,36
ENTIRE AREA	2,72	3,36	2,80	0,97	402,48	282,81

Appendix 7: Basic Statistical Parameters

Table I: Number of soil samples, mean values and standard deviations for Site E.

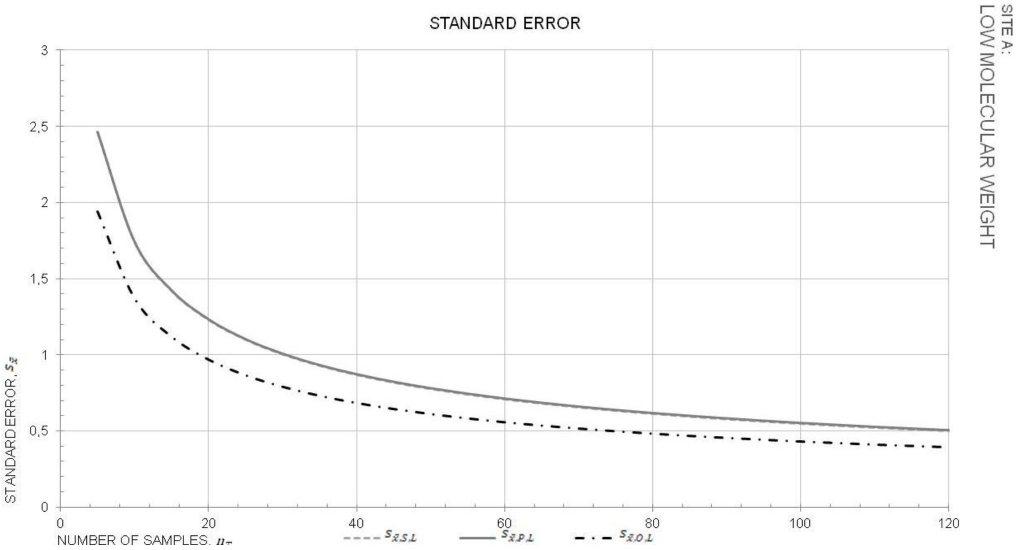
STATISTICS	**n**	**x_L**	**x_M**	**x_H**	**s_L**	**s_M**	**s_H**
CLASS 1	3	0,25	0,16	0,05	0,12	0,07	0,00
CLASS 2	48	2,04	6,59	4,92	4,15	16,99	12,78
CLASS 3	2	3,99	3,74	0,28	2,04	0,40	2,49
CLASS 4	0	-	-	-	-	-	-
CLASS 5	0	-	-	-	-	-	-
ENTIRE AREA	53	1,98	6,16	4,55	4,01	16,38	12,33

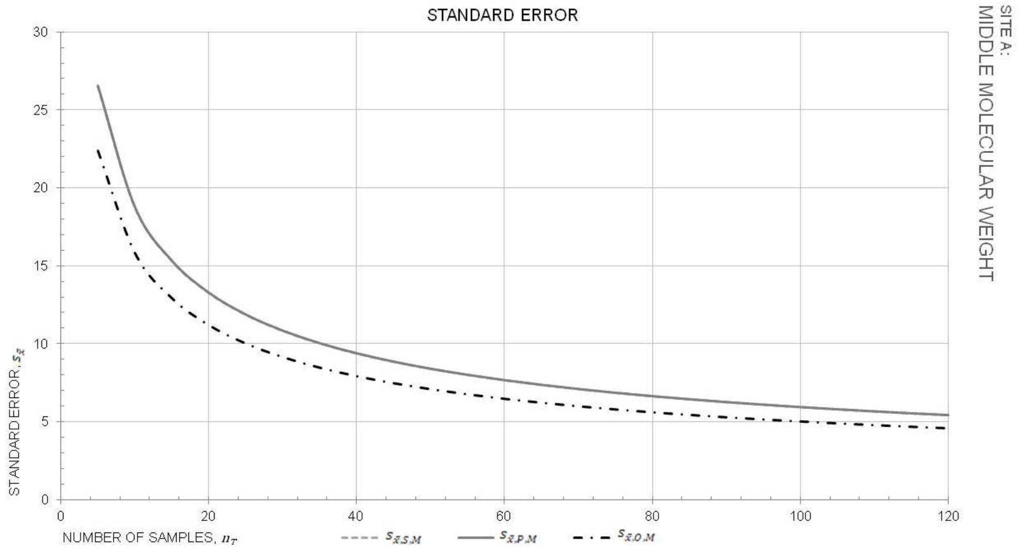
Table J: Coefficient of variation and variance for Site E.

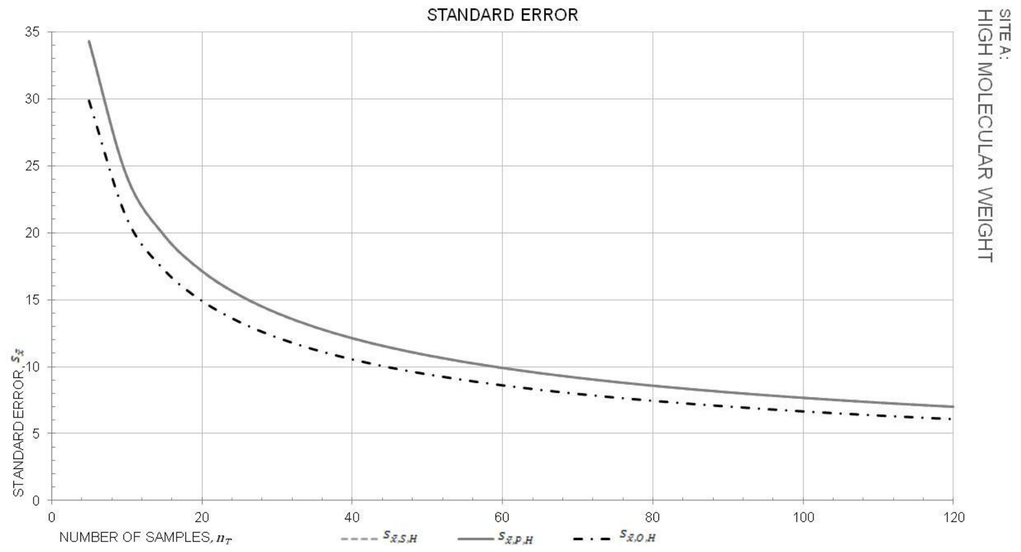
STATISTICS	**C_L**	**C_M**	**C_H**	**s^2_L**	**s^2_M**	**s^2_H**
CLASS 1	0,50	0,47	0,00	0,02	0,01	0,00
CLASS 2	2,03	2,58	2,60	17,20	288,68	163,21
CLASS 3	0,51	0,11	8,89	4,18	0,16	6,20
CLASS 4	-	-	-	-	-	-
CLASS 5	-	-	-	-	-	-
ENTIRE AREA	2,03	2,66	2,71	16,11	268,45	152,09

Appendix 8: Standard Error Graphs

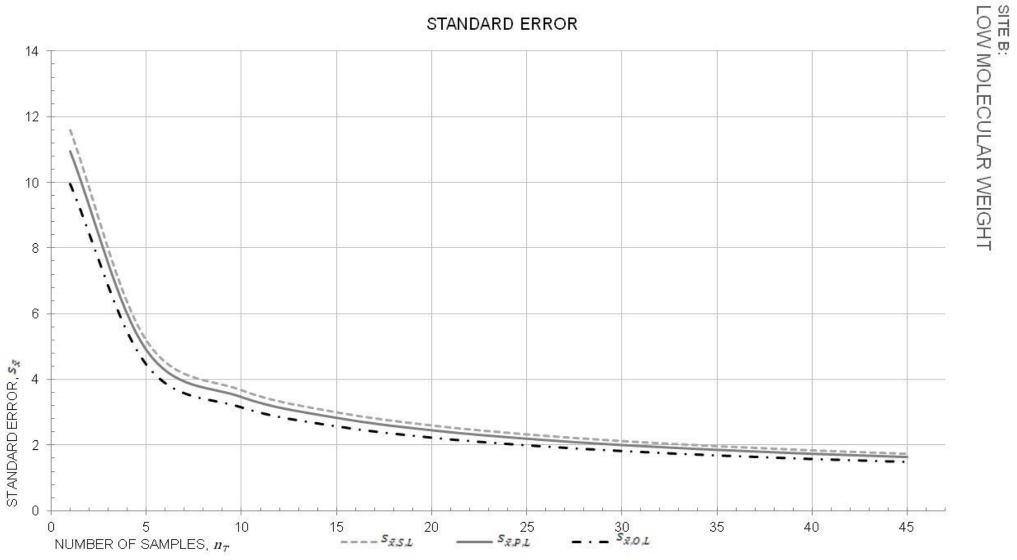
Standard Error Graphs

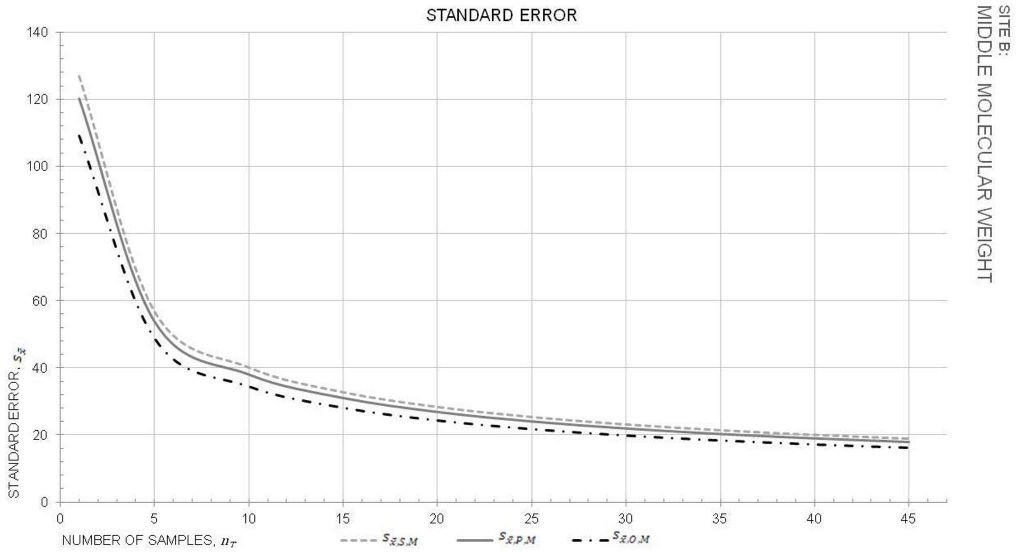


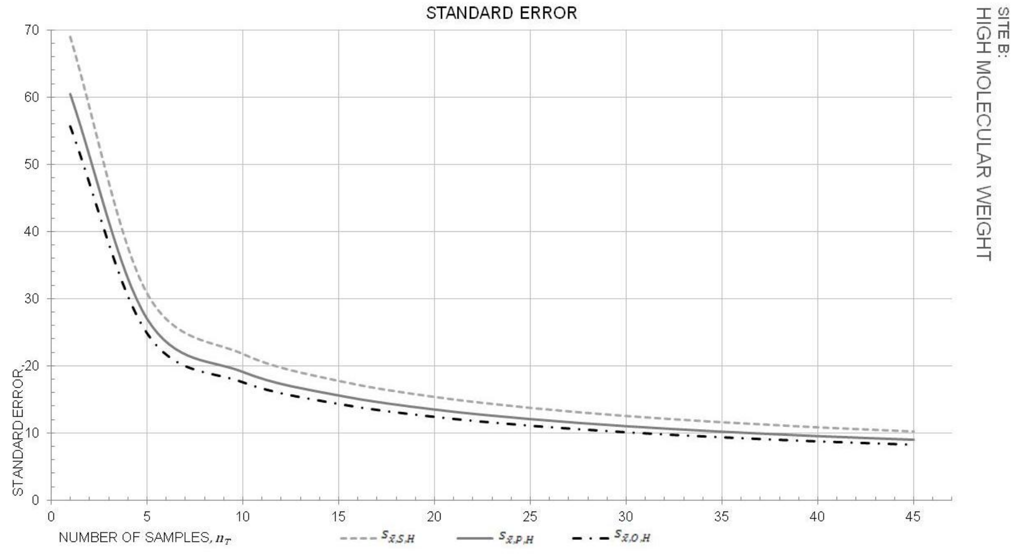




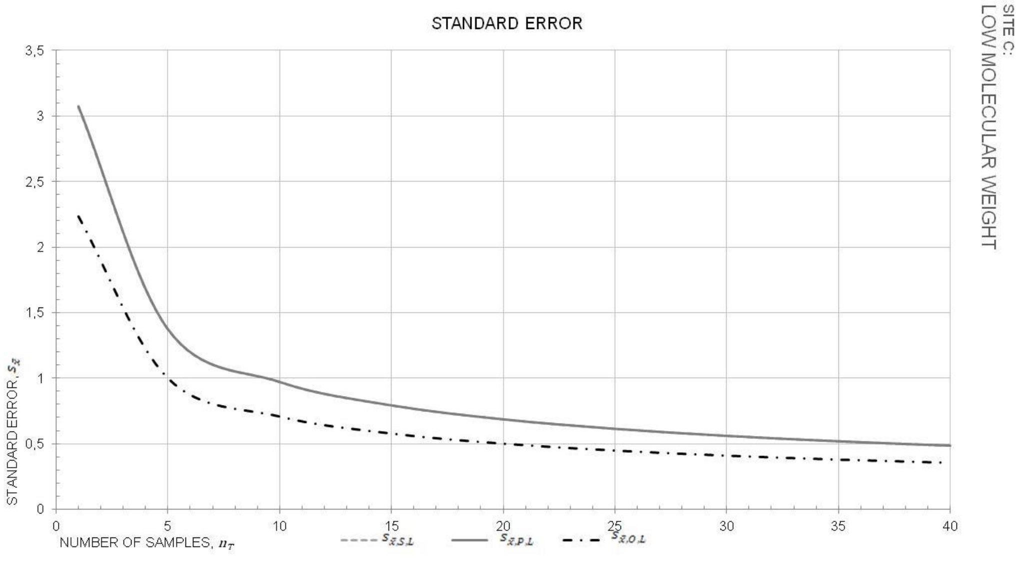
Appendix 8: Standard Error Graphs

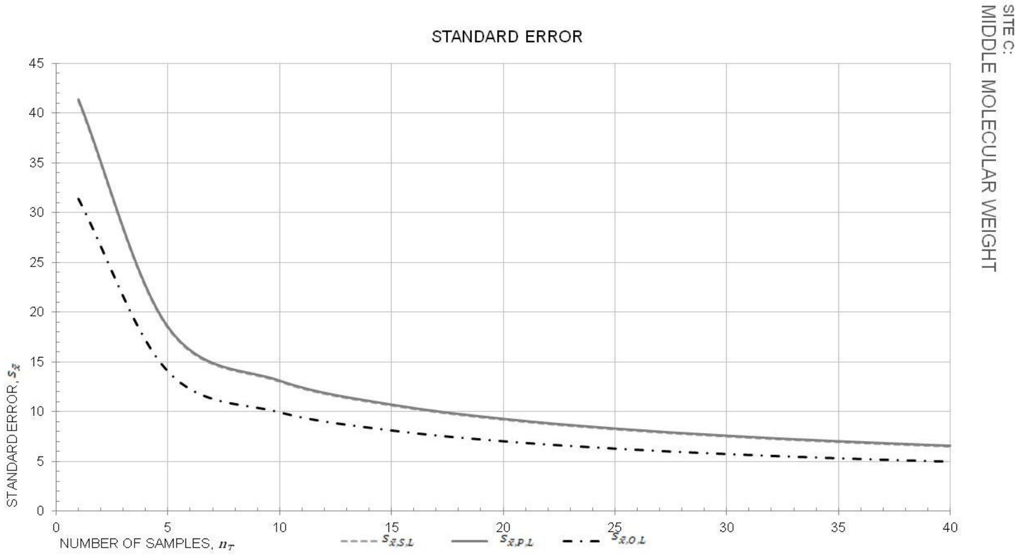


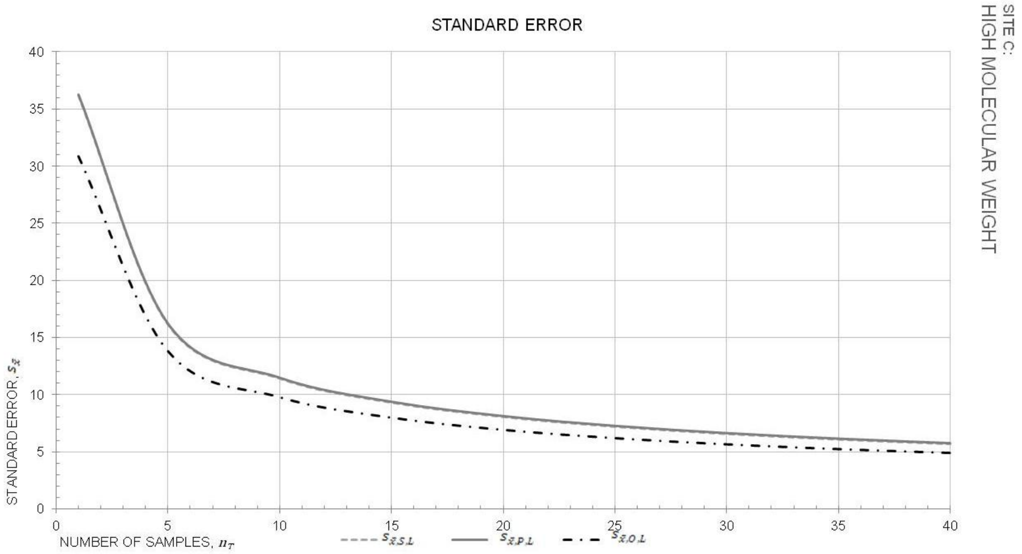




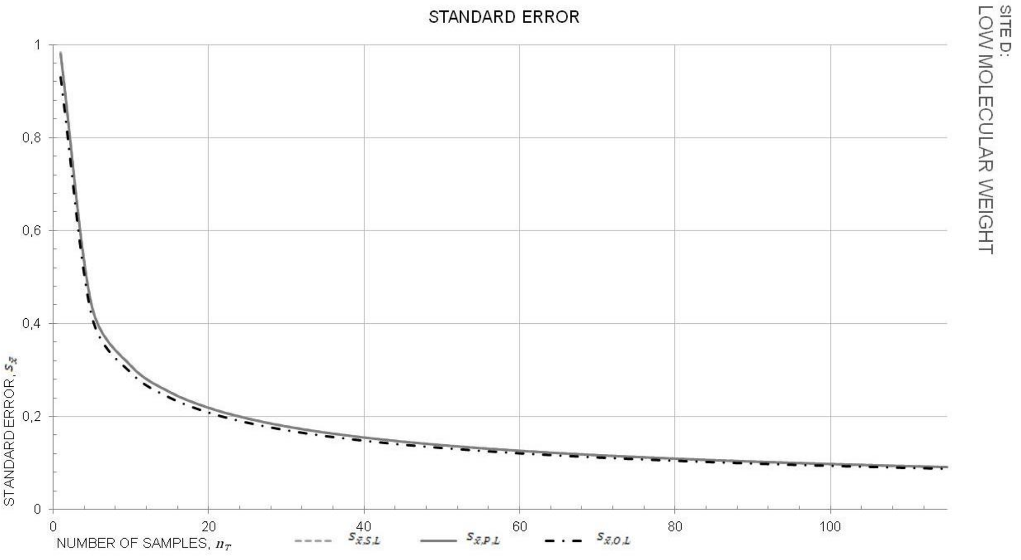
Appendix 8: Standard Error Graphs

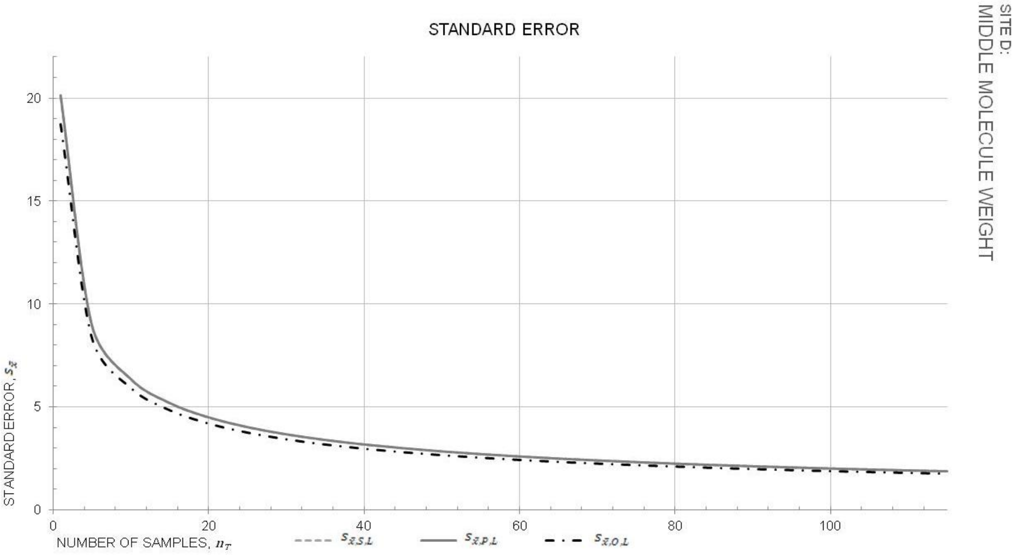


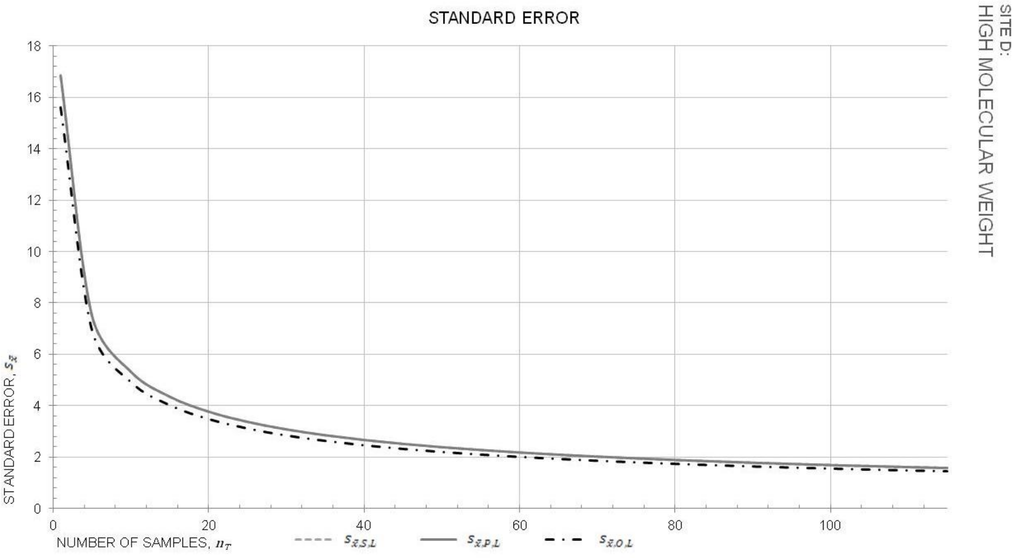




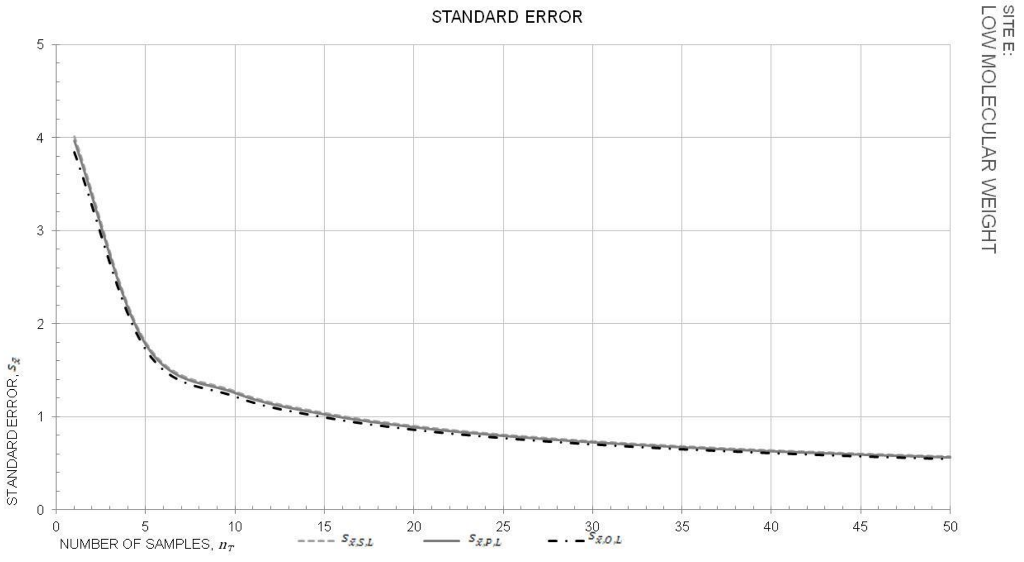
Appendix 8: Standard Error Graphs

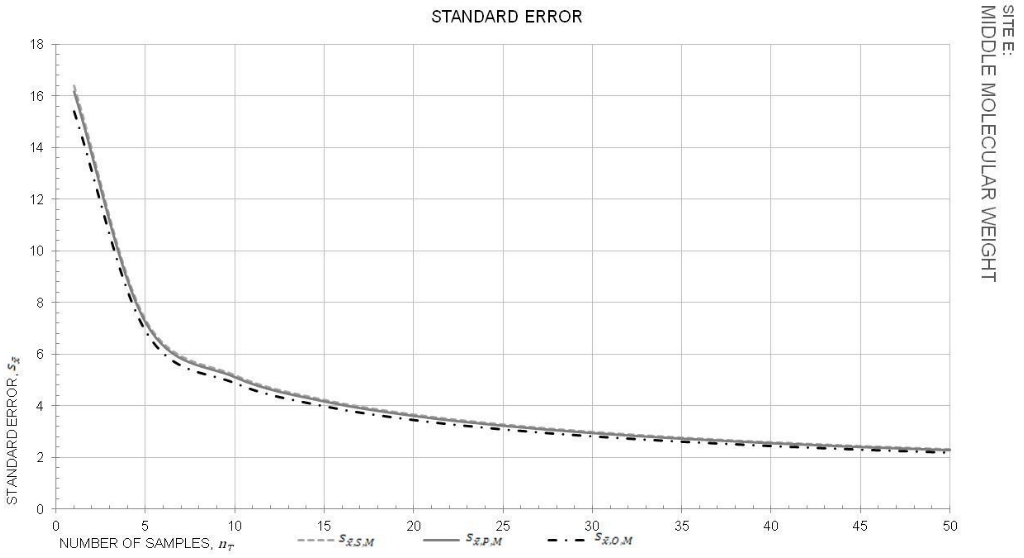


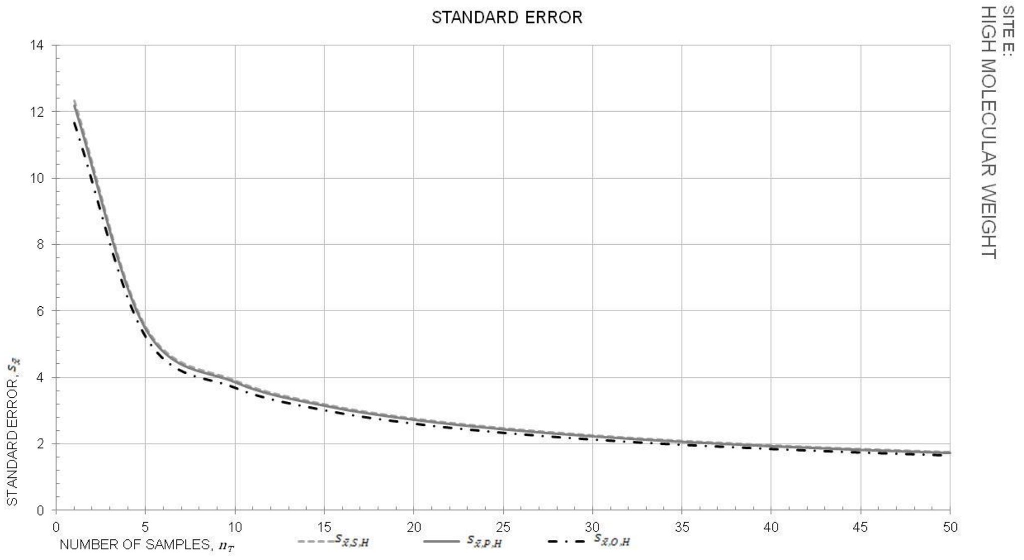




Appendix 8: Standard Error Graphs







Sampling Simulations Data

Table A: Values underlying the Data Sampling Simulation for LMW, Site A.

	LOW MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNOR	Mean	Std	LOGNOR	Mean	Std	LOGNOR
5	1,633565	2,490267	68%	1,815047	2,565369	70%	2,044556	3,533899	67%
10	1,653214	1,78306	75%	1,725584	1,764147	76%	2,138748	2,45733	75%
15	1,609265	1,428279	79%	1,677223	1,391869	81%	2,141655	1,973374	80%
20	1,610589	1,231546	82%	1,589594	1,220341	82%	2,112074	1,72869	83%
25	1,612054	1,106485	85%	1,620198	1,08819	85%	2,131581	1,549552	86%
30	1,608297	0,987643	87%	1,650167	0,989497	88%	2,103813	1,379408	88%
35	1,614028	0,934934	89%	1,669978	0,949989	90%	2,08746	1,295723	89%
40	1,631445	0,879479	91%	1,682942	0,885608	91%	2,099301	1,199164	91%
45	1,626506	0,829571	92%	1,6115	0,813165	92%	2,072932	1,136196	92%
50	1,61379	0,775373	93%	1,656983	0,779421	94%	2,097881	1,076003	93%
55	1,613697	0,7362	94%	1,639219	0,732971	94%	2,073646	1,011171	94%
60	1,628846	0,717752	95%	1,647425	0,706897	95%	2,047013	0,957083	95%
65	1,629151	0,690549	95%	1,657013	0,686138	96%	2,085118	0,947171	96%
70	1,622614	0,660053	96%	1,650671	0,65059	96%	2,097115	0,910966	96%
75	1,622644	0,640402	96%	1,668548	0,631546	97%	2,074326	0,873956	97%
80	1,615803	0,606652	97%	1,639063	0,611893	97%	2,068076	0,849339	97%
85	1,608632	0,590332	97%	1,649299	0,596999	98%	2,092039	0,83152	97%
90	1,617815	0,579905	98%	1,651332	0,586205	98%	2,072327	0,790797	98%
95	1,628028	0,567984	98%	1,65316	0,564132	98%	2,078626	0,772574	98%
100	1,619838	0,547527	98%	1,656484	0,552627	98%	2,094475	0,766576	98%
105	1,617674	0,536002	98%	1,652518	0,539943	99%	2,064958	0,738554	98%
110	1,613615	0,515759	99%	1,643681	0,521742	99%	2,090012	0,717164	99%
115	1,626886	0,515648	99%	1,650198	0,516214	99%	2,082148	0,698313	99%
120	1,622505	0,501564	99%	1,64763	0,501673	99%	2,074654	0,685582	99%

	Mean
True	1,604333333

Appendix 9: Sampling Simulations Data

Table B: Values underlying the Data Sampling Simulation for MMW, Site A.

	MIDDLE MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNOR	Mean	Std	LOGNOR	Mean	Std	LOGNOR
5	20,66803	26,61314	75%	22,95415	27,67835	76%	24,92497	32,45097	75%
10	20,87465	19,10268	83%	21,28053	18,86347	83%	25,63528	24,88263	82%
15	20,42896	15,35193	87%	20,94945	14,89922	89%	24,81473	19,31881	87%
20	20,55068	13,28655	91%	20,48208	13,16127	91%	24,17834	16,63953	90%
25	20,47811	11,83602	93%	20,71715	11,68023	94%	25,00574	15,38918	92%
30	20,48256	10,63519	95%	21,08174	10,71323	95%	24,72107	14,01153	94%
35	20,52322	10,03341	96%	21,13152	10,18272	96%	24,76746	12,68686	96%
40	20,73353	9,446366	97%	21,29484	9,489358	97%	25,02619	12,08217	97%
45	20,67915	8,930435	98%	20,63083	8,784492	98%	24,83489	11,33941	97%
50	20,49102	8,340815	98%	20,95116	8,363555	98%	24,96398	10,92811	98%
55	20,55799	7,920553	99%	20,79641	7,844856	99%	24,96111	10,32896	98%
60	20,69598	7,671386	99%	20,87838	7,631151	99%	24,7165	9,842634	99%
65	20,7108	7,441626	99%	21,01744	7,426783	99%	24,772	9,339793	99%
70	20,6194	7,116166	99%	20,92765	7,008098	99%	24,84198	9,094455	99%
75	20,64468	6,897452	99%	21,07728	6,806449	100%	25,01755	8,778279	99%
80	20,55317	6,54877	100%	20,81738	6,593485	100%	24,86246	8,603402	99%
85	20,49123	6,3457	100%	20,93921	6,418388	100%	24,85015	8,235207	100%
90	20,55717	6,267316	100%	20,90247	6,318135	100%	24,90739	8,086287	100%
95	20,68998	6,113792	100%	20,92596	6,056583	100%	24,94993	7,868826	100%
100	20,59113	5,900558	100%	21,0334	5,970109	100%	24,93484	7,675662	100%
105	20,5841	5,762297	100%	20,93274	5,807784	100%	24,89997	7,418932	100%
110	20,52817	5,547865	100%	20,87124	5,657152	100%	24,96044	7,270436	100%
115	20,67033	5,554838	100%	20,91353	5,52162	100%	25,01274	7,144209	100%
120	20,62798	5,37519	100%	20,90798	5,42799	100%	24,93073	7,01057	100%

	Mean
True	20,394875

Appendix 9: Sampling Simulations Data

Table C: Values underlying the Data Sampling Simulation for HMW, Site A.

	HIGH MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	mean	std	LOGNOR	mean	std	LOGNOR	mean	std	LOGNOR
5	29,3683	34,41994	77%	32,64763	35,86737	79%	34,79847	41,94959	77%
10	29,65275	24,62053	86%	29,91253	24,47208	86%	33,96905	29,16728	85%
15	29,09946	19,85924	90%	29,47328	19,23174	91%	34,72138	24,81704	90%
20	29,22268	17,14956	93%	28,76228	17,07299	93%	34,77334	21,47215	93%
25	29,13322	15,22562	95%	29,14341	15,11867	96%	34,68854	19,4578	95%
30	29,21932	13,80046	97%	29,74767	13,88926	97%	34,13569	17,12351	96%
35	29,23414	12,96295	98%	29,67823	13,16241	98%	34,67232	16,2499	97%
40	29,51278	12,17135	98%	30,01961	12,27199	99%	34,37274	14,99944	98%
45	29,45397	11,53547	99%	29,03702	11,33563	99%	33,92228	13,92956	99%
50	29,18351	10,76551	99%	29,50934	10,82728	99%	33,97812	13,45949	99%
55	29,30895	10,21809	99%	29,31537	10,15383	99%	33,8431	12,61672	99%
60	29,44421	9,89526	100%	29,45514	9,882112	100%	34,3032	12,42809	99%
65	29,47994	9,622847	100%	29,61181	9,591275	100%	34,00941	11,79045	100%
70	29,32992	9,224005	100%	29,49504	9,057013	100%	34,09762	11,30492	100%
75	29,40452	8,887974	100%	29,67798	8,82988	100%	33,87046	10,89556	100%
80	29,28608	8,494988	100%	29,3111	8,522819	100%	34,18027	10,59238	100%
85	29,19777	8,207123	100%	29,49417	8,313118	100%	34,02839	10,22287	100%
90	29,3	8,08532	100%	29,43895	8,15853	100%	34,03032	9,959642	100%
95	29,46377	7,884949	100%	29,47468	7,836167	100%	34,00117	9,736283	100%
100	29,29797	7,635014	100%	29,63043	7,71967	100%	34,18339	9,485525	100%
105	29,3351	7,462464	100%	29,45982	7,504893	100%	33,98589	9,21952	100%
110	29,26996	7,177161	100%	29,37654	7,328428	100%	34,14847	8,852721	100%
115	29,41876	7,186328	100%	29,42695	7,121638	100%	34,12249	8,777216	100%
120	29,38838	6,921704	100%	29,47036	7,029789	100%	34,347	8,684909	100%

	Mean
True	29,05395833

Appendix 9: Sampling Simulations Data

Table D: Values underlying the Data Sampling Simulation for LMW, Site B.

	LOW MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNOR	Mean	Std	LOGNOR	Mean	Std	LOGNOR
5	2,90	5,25	64%	2,37	4,87	61%	2,67	5,37	62%
10	2,80	3,58	69%	2,52	3,82	65%	2,76	3,86	67%
15	2,88	2,98	73%	2,50	2,96	69%	2,69	3,13	70%
20	2,88	2,59	76%	2,44	2,53	71%	2,68	2,69	73%
25	2,86	2,30	79%	2,52	2,35	74%	2,69	2,41	75%
30	2,86	2,10	81%	2,50	2,13	75%	2,72	2,21	78%
35	2,87	1,96	83%	2,52	2,00	77%	2,70	2,05	79%
40	2,88	1,85	84%	2,50	1,83	79%	2,72	1,91	81%
45	2,83	1,71	85%	2,48	1,70	80%	2,67	1,79	82%

	Mean
True	2,812276596

Table E: Values underlying the Data Sampling Simulation for MMW, Site B.

	MIDDLE MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNOR	Mean	Std	LOGNOR	Mean	Std	LOGNOR
5	31,14	57,53	69%	25,61	53,55	66%	28,90	59,00	67%
10	30,02	39,25	75%	27,21	41,97	71%	29,82	42,43	73%
15	30,98	32,57	80%	26,96	32,56	77%	29,04	34,40	77%
20	30,93	28,42	83%	26,33	27,83	79%	28,95	29,57	81%
25	30,71	25,14	86%	27,20	25,79	82%	29,09	26,50	83%
30	30,67	22,96	88%	26,98	23,37	84%	29,39	24,27	86%
35	30,86	21,52	90%	27,15	21,91	86%	29,84	22,85	88%
40	30,94	20,25	91%	26,92	20,11	88%	29,37	21,00	89%
45	30,38	18,72	93%	26,71	18,71	89%	28,82	19,68	90%

	Mean
True	30,17910638

Appendix 9: Sampling Simulations Data

Table F: Values underlying the Data Sampling Simulation for HMW, Site B.

	HIGH MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNOR	Mean	Std	LOGNOR	Mean	Std	LOGNOR
5	21,02	31,34	72%	17,00	27,01	70%	18,35	29,67	70%
10	21,33	21,47	80%	17,64	21,13	75%	18,81	21,32	77%
15	21,80	17,72	85%	17,49	16,38	81%	18,42	17,31	81%
20	21,79	15,45	89%	17,26	14,02	84%	18,35	14,86	85%
25	21,62	13,90	91%	17,63	12,95	87%	21,28	14,55	89%
30	21,60	12,51	93%	17,57	12,51	88%	18,40	11,97	90%
35	21,71	11,69	94%	17,62	11,01	91%	20,84	12,27	93%
40	21,74	10,99	96%	17,52	10,10	92%	20,57	11,39	94%
45	21,43	10,16	96%	17,43	9,41	94%	20,06	10,52	95%

	Mean
True	21,23301915

Table G: Values underlying the Data Sampling Simulation for LMW, Site C.

	LOW MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNOR	Mean	Std	LOGNOR	Mean	Std	LOGNOR
5	0,896777	1,375783	77%	0,859734	1,350127	76 %	1,424496	1,919205	79%
10	0,905907	0,96548	85%	0,947494	0,978784	86%	1,52941	1,428736	87%
15	0,905245	0,801994	90%	0,920032	0,790270	90%	1,497922	1,144386	92%
20	0,900900	0,690668	93%	0,928179	0,686977	93%	1,472609	0,982902	94%
25	0,892649	0,60585	95%	0,923659	0,609389	95%	1,4631	0,886000	96%
30	0,893270	0,554692	96%	0,940744	0,561455	97%	1,480065	0,816867	97%
35	0,896654	0,517814	97%	0,922719	0,517184	98%	1,479013	0,737634	98%
40	0,894509	0,489241	98%	0,933393	0,485215	98%	1,473584	0,691851	99%

	Mean
True	0,8742

Appendix 9: Sampling Simulations Data

Table H: Values underlying the Data Sampling Simulation for MMW, Site C.

	MIDDLE MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNORM	Mean	Std	LOGNORM	Mean	Std	LOGNORM
5	12,04095	18,49341	70%	11,29495	18,06111	69%	18,93929	25,47816	74%
10	12,20349	12,98024	77%	12,78501	13,16845	78%	19,24047	18,18501	82%
15	12,17852	10,75763	82%	12,29959	10,61417	82%	19,07049	14,69237	87%
20	12,10815	9,299292	85%	12,53755	9,246386	86%	18,61896	12,71594	90%
25	12,03174	8,137088	88%	12,41264	8,203301	89%	18,53952	11,34553	92%
30	12,00833	7,452760	90%	12,7065	7,544537	91%	18,69947	10,30584	94%
35	12,07139	6,951156	92%	12,41074	6,956146	92%	18,79350	9,649221	95%
40	12,04637	6,575549	93%	12,60464	6,525613	94%	18,90725	9,066411	97%

	Mean
True	11,762775

Table I: Values underlying the Data Sampling Simulation for HMW, Site C.

	HIGH MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNORM	Mean	Std	LOGNORM	Mean	Std	LOGNORM
5	12,50055	16,13054	73,2%	11,2122	14,99963	71,9%	18,18906	19,39076	79,1%
10	12,66366	11,39714	81,4%	13,24964	11,42737	82,6%	18,03518	13,63821	87,3%
15	12,62988	9,314687	86,1%	12,57162	9,107499	86,5%	17,93501	11,07663	91,8%
20	12,52382	8,113006	89,2%	13,06967	8,068512	90,5%	18,10938	9,680055	94,7%
25	12,56915	7,121501	92,1%	12,8737	7,128442	92,7%	17,98782	8,681749	96,3%
30	12,50694	6,556995	93,6%	13,28035	6,590095	94,9%	17,48649	7,693832	97,4%
35	12,53361	6,084951	95,0%	12,89153	6,056859	95,7%	17,57215	7,161516	98,2%
40	12,53074	5,739131	96,0%	13,17012	5,668431	97,0%	17,60037	6,734756	98,7%

	Mean
True	12,2525

Appendix 9: Sampling Simulations Data

Table J: Values underlying the Data Sampling Simulation for LMW, Site D.

	LOW MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNORM	Mean	Std	LOGNORM	Mean	Std	LOGNORM
5	0,37	0,445	100%	0,407	0,472	100%	0,4	0,464	100%
10	0,36	0,31	100%	0,35	0,31	100%	0,4	0,33	100%
15	0,36	0,25	100%	0,37	0,25	100%	0,37	0,27	100%
20	0,36	0,21	100%	0,37	0,21	100%	0,38	0,23	100%
25	0,37	0,2	100%	0,36	0,19	100%	0,39	0,21	100%
30	0,36	0,18	100%	0,36	0,17	100%	0,37	0,19	100%
35	0,36	0,17	100%	0,37	0,17	100%	0,37	0,17	100%
40	0,36	0,16	100%	0,35	0,15	100%	0,38	0,16	100%
45	0,36	0,14	100%	0,36	0,15	100%	0,38	0,15	100%
50	0,36	0,14	100%	0,36	0,14	100%	0,39	0,15	100%
55	0,36	0,13	100%	0,36	0,13	100%	0,37	0,14	100%
60	0,36	0,13	100%	0,37	0,13	100%	0,39	0,13	100%
65	0,36	0,12	100%	0,36	0,12	100%	0,38	0,13	100%
70	0,36	0,12	100%	0,36	0,11	100%	0,38	0,12	100%
75	0,36	0,11	100%	0,37	0,11	100%	0,38	0,12	100%
80	0,36	0,11	100%	0,36	0,11	100%	0,39	0,11	100%
85	0,36	0,10	100%	0,36	0,11	100%	0,39	0,11	100%
90	0,36	0,10	100%	0,37	0,10	100%	0,39	0,11	100%
95	0,36	0,10	100%	0,36	0,10	100%	0,38	0,10	100%
100	0,36	0,10	100%	0,36	0,10	100%	0,39	0,10	100%
105	0,36	0,10	100%	0,36	0,10	100%	0,39	0,10	100%
110	0,36	0,09	100%	0,36	0,09	100%	0,38	0,10	100%
115	0,36	0,09	100%	0,36	0,09	100%	0,35	0,09	100%

	Mean
True	0,362241379

Appendix 9: Sampling Simulations Data

Table K: Values underlying the Data Sampling Simulation for MMW, Site D.

	MIDDLE MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNORM	Mean	Std	LOGNORM	Mean	Std	LOGNORM
5	6,11	8,985	68%	6,784	9,373	70%	6,52	9,929	68%
10	6,02	6,32	75%	6,18	6,5	75%	6,72	6,91	76%
15	5,95	5,11	79%	5,98	5,1	79%	6,61	5,7	80%
20	5,98	4,44	83%	6,09	4,44	83%	6,66	4,88	84%
25	6,1	4,09	85%	6,04	3,94	86%	6,59	4,36	86%
30	6,05	3,69	88%	5,98	3,58	88%	6,63	3,95	89%
35	6,03	3,38	90%	6,16	3,44	90%	6,68	3,73	91%
40	6,03	3,23	91%	6,00	3,14	91%	6,68	3,47	92%
45	5,98	3,00	92%	6,08	2,96	93%	6,67	3,26	93%
50	5,99	2,84	93%	5,97	2,82	93%	6,75	3,16	94%
55	5,98	2,70	94%	6,03	2,70	94%	6,71	2,92	95%
60	5,98	2,57	95%	6,12	2,58	95%	6,69	2,78	96%
65	5,99	2,46	96%	5,97	2,47	96%	6,67	2,66	97%
70	5,97	2,38	96%	6,00	2,38	96%	6,67	2,59	97%
75	5,98	2,30	97%	6,10	2,34	97%	6,59	2,52	97%
80	5,97	2,23	97%	6,05	2,24	97%	6,62	2,45	98%
85	5,98	2,19	97%	6,02	2,17	97%	6,63	2,35	98%
90	6,00	2,05	98%	6,07	2,08	98%	6,67	2,29	98%
95	5,95	2,04	98%	6,00	2,04	98%	6,62	2,24	98%
100	5,97	2,00	98%	6,01	1,98	98%	6,67	2,19	99%
105	5,98	1,95	98%	5,98	1,96	98%	6,65	2,13	99%
110	5,95	1,89	99%	6,03	1,91	99%	6,62	2,05	99%
115	5,99	1,89	99%	6,01	1,84	99%	6,65	2,02	99%

	Mean
True	5,97137931

Appendix 9: Sampling Simulations Data

Table L: Values underlying the Data Sampling Simulation for HMW, Site D.

	HIGH MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNORM	Mean	Std	LOGNORM	Mean	Std	LOGNORM
5	6,049	7,498	71%	6,662	7,791	73%	6,839	8,462	72%
10	6,05	5,3	79%	6,18	5,45	79%	6,69	5,78	80%
15	5,96	4,29	83%	5,94	4,27	83%	6,79	4,82	85%
20	6,02	3,75	87%	6,05	3,72	87%	6,69	4,03	89%
25	6,1	3,42	90%	6,03	3,3	90%	6,71	3,66	91%
30	6,06	3,08	92%	5,95	3,00	92%	6,73	3,36	93%
35	6,05	2,84	93%	6,09	2,87	93%	6,70	3,07	94%
40	6,05	2,71	94%	5,98	2,64	94%	6,74	2,91	96%
45	6,01	2,53	95%	6,04	2,48	96%	6,66	2,74	96%
50	6,02	2,38	96%	6,01	2,36	96%	6,78	2,67	97%
55	6,02	2,27	97%	6,07	2,26	97%	6,72	2,49	98%
60	6,01	2,15	98%	6,14	2,15	98%	6,78	2,46	98%
65	6,02	2,06	98%	6,01	2,06	98%	6,68	2,29	98%
70	6,00	1,99	98%	6,04	2,00	98%	6,72	2,19	99%
75	6,02	1,92	99%	6,12	1,96	99%	6,69	2,13	99%
80	6,00	1,87	99%	6,08	1,87	99%	6,68	2,03	99%
85	6,02	1,84	99%	6,05	1,81	99%	6,69	1,98	99%
90	6,04	1,71	99%	6,09	1,74	99%	6,71	1,95	99%
95	5,98	1,70	99%	6,05	1,71	99%	6,69	1,89	100%
100	6,01	1,68	99%	6,04	1,65	99%	6,72	1,86	100%
105	6,02	1,63	100%	6,02	1,64	100%	6,71	1,79	100%
110	5,99	1,59	100%	6,06	1,60	100%	6,71	1,76	100%
115	6,02	1,59	100%	6,04	1,54	100%	5,69	1,57	99%

	Mean
True	6,004956897

Appendix 9: Sampling Simulations Data

Table M: Values underlying the Data Sampling Simulation for LMW, Site E.

	LOW MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNORM	Mean	Std	LOGNORM	Mean	Std	LOGNORM
5	1,979216	1,791645	77%	2,057801	1,841834	77%	2,046949	1,844921	77%
10	1,973767	1,261313	85%	1,881888	1,264636	83%	2,088765	1,316873	86%
15	1,979370	1,022095	90%	1,949101	1,036052	89%	2,057167	1,068997	90%
20	1,950876	0,882688	93%	2,074920	0,875715	95%	2,062440	0,941620	93%
25	1,973217	0,794255	95%	2,064501	0,804227	96%	2,133586	0,807876	96%
30	1,976910	0,730326	96%	1,996015	0,717388	97%	2,129756	0,746134	97%
35	1,984873	0,675492	97%	2,024310	0,669630	98%	2,106531	0,684423	98%
40	1,973011	0,633985	98%	2,066334	0,625410	99%	2,103978	0,641793	99%
45	1,972567	0,594045	99%	2,066363	0,594567	99%	2,096416	0,612167	99%
50	1,973328	0,561502	99%	2,036157	0,557589	99%	2,101203	0,575821	99%

	Mean
True	1,960566038

Table N: Values underlying the Data Sampling Simulation for MMW, Site E.

	MIDDLE MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNORM	Mean	Std	LOGNORM	Mean	Std	LOGNORM
5	6,222396	7,379920	72%	6,323336	7,560193	72%	6,24779	7,54990	72%
10	6,155965	5,165464	80%	5,711757	5,064270	78%	6,24720	5,29330	80%
15	6,210655	4,200320	85%	5,885990	4,229153	83%	6,27169	4,33222	85%
20	6,120187	3,696887	88%	5,844302	3,571406	87%	6,25981	3,72772	88%
25	6,131989	3,276602	91%	5,901407	3,227914	90%	6,23754	3,38720	90%
30	6,190528	3,004520	93%	5,773975	2,944067	91%	6,31324	3,08584	93%
35	6,180564	2,723015	95%	5,845936	2,711599	93%	6,21318	2,84604	94%
40	6,162350	2,590410	95%	5,846144	2,512918	95%	6,28561	2,63990	95%
45	6,121972	2,416386	96%	5,875009	2,429665	95%	6,25893	2,49755	96%
50	6,133938	2,285384	97%	5,799930	2,241340	96%	6,30109	2,41300	97%

	Mean
True	6,12509434

Appendix 9: Sampling Simulations Data

Table O: Values underlying the Data Sampling Simulation for HMW, Site E.

	HIGH MOLECULAR WEIGTH								
	SRS			PSRS			OSRS		
	MEAN			MEAN			MEAN		
n	Mean	Std	LOGNORM	Mean	Std	LOGNORM	Mean	Std	LOGNORM
5	4,581186	5,517012	71%	4,430894	5,489180	70%	4,349165	5,404086	70%
10	4,536013	3,887651	78%	3,994892	3,677159	75%	4,413447	3,858142	77%
15	4,602687	3,174011	83%	4,099499	3,061213	80%	4,386440	3,145077	82%
20	4,518839	2,776336	86%	3,955699	2,587939	83%	4,429348	2,770099	85%
25	4,540941	2,469792	89%	4,030067	2,345227	86%	4,473309	2,468895	88%
30	4,571020	2,254608	91%	3,952401	2,126468	87%	4,371211	2,204803	90%
35	4,568444	2,051872	93%	4,001363	1,964536	90%	4,413528	2,082722	92%
40	4,561278	1,959254	94%	3,969505	1,818705	91%	4,387998	1,914063	93%
45	4,522112	1,824459	95%	3,996741	1,760793	92%	4,404665	1,824118	94%
50	4,523850	1,715929	96%	3,949942	1,629642	93%	4,375469	1,712529	95%

	Mean
True	4,535283019

Table P: Values underlying the Probabilistic Sampling Simulation for LMW, Site A.

	LOW MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	1,63	1,90	73%	1,69	1,55	79%	2,08	2,26	76%
20	1,61	1,37	80%	1,57	1,24	81%	2,04	1,71	82%
30	1,60	1,06	86%	1,60	1,28	81%	2,06	1,36	88%
40	1,61	0,82	92%	1,64	0,89	91%	2,00	1,10	92%
50	1,60	0,77	93%	1,61	0,74	94%	2,04	1,03	94%
60	1,61	0,71	95%	1,61	0,69	95%	2,03	0,94	95%

Mean	1,604333

Appendix 9: Sampling Simulations Data

Table Q: Values underlying the Probabilistic Sampling Simulation for MMW, Site A.

	MIDDLE MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	20,43	18,49	83%	21,27	24,69	77%	24,29	21,93	83%
20	20,40	14,89	88%	20,27	12,57	92%	24,02	16,82	89%
30	20,26	10,23	95%	20,14	9,16	97%	24,09	13,32	94%
40	20,35	9,01	97%	20,62	9,23	97%	24,44	11,77	97%
50	20,41	8,43	98%	20,44	8,71	98%	24,35	10,93	97%
60	20,49	7,71	99%	20,38	7,48	99%	24,37	10,06	98%

Mean	20,39488

Table R: Values underlying the Probabilistic Sampling Simulation for HMW, Site A.

	HIGH MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	28,76	26,01	84%	29,73	23,23	87%	34,03	27,22	87%
20	29,31	18,12	92%	28,52	15,52	95%	34,90	23,37	91%
30	28,80	13,20	97%	29,26	13,73	97%	34,01	17,02	96%
40	29,08	11,85	98%	29,55	12,71	98%	34,17	14,81	98%
50	29,02	11,50	99%	29,14	11,07	99%	33,60	13,90	99%
60	28,99	9,54	100%	28,97	9,82	100%	33,87	12,88	99%

Mean	29,05396

Table S: Values underlying the Probabilistic Sampling Simulation for LMW, Site B.

	LOW MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	2,89	3,63	0,70	2,43	5,23	0,61	2,63	3,24	0,69
20	2,77	2,24	0,78	3,65	2,28	0,87	2,67	2,70	0,73
30	2,80	1,90	0,82	3,45	1,82	0,91	2,66	2,09	0,78
40	2,77	1,61	0,86	3,85	1,77	0,94	2,70	2,10	0,79
50	2,83	1,83	0,84	3,68	1,52	0,96	2,61	1,62	0,83
60	2,80	1,35	0,90	3,73	1,47	0,97	2,62	1,39	0,87

Mean	2,812277

Appendix 9: Sampling Simulations Data

Table T: Values underlying the Probabilistic Sampling Simulation for MMW, Site B.

	MIDDLE MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	29,98	37,89	76%	24,61	32,88	74%	28,80	40,97	73%
20	29,79	25,37	85%	34,92	22,55	92%	29,29	40,93	74%
30	29,93	20,63	90%	31,63	19,51	93%	28,57	21,68	88%
40	30,48	21,26	90%	34,77	16,69	97%	28,84	21,20	88%
50	30,18	16,77	94%	33,17	15,91	97%	28,08	17,68	92%
60	30,15	15,72	96%	31,53	13,75	98%	28,52	15,63	95%

Mean	30,17911

Table U: Values underlying the Probabilistic Sampling Simulation for HMW, Site B.

	HIGH MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	21,17	19,61	82%	14,36	17,22	74%	16,57	20,23	75%
20	21,43	16,60	87%	25,13	13,43	95%	16,47	14,81	82%
30	21,34	13,24	92%	21,24	9,89	97%	16,31	12,41	86%
40	21,30	10,49	96%	25,11	9,61	99%	16,32	10,04	91%
50	21,28	9,90	97%	22,87	7,86	99%	16,33	8,86	93%
60	21,24	8,81	98%	21,65	7,66	99%	16,57	8,58	94%

Mean	21,23302

Table V: Values underlying the Probabilistic Sampling Simulation for LMW, Site C.

	LOW MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	0,88	0,96	85%	0,88	1,01	84%	1,43	1,55	84%
20	0,88	0,64	94%	0,87	0,61	95%	1,40	0,95	95%
30	0,87	0,50	98%	0,88	0,60	95%	1,40	0,77	98%
40	0,87	0,44	99%	0,89	0,50	98%	1,40	0,69	99%
50	0,87	0,42	99%	0,87	0,42	99%	1,41	0,62	99%
60	0,88	0,39	100%	0,88	0,38	100%	1,41	0,58	100%

Mean	0,874

Appendix 9: Sampling Simulations Data

Table W: Values underlying the Probabilistic Sampling Simulation for MMW, Site C.

	MIDDLE MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	11,66	12,19	77%	11,70	11,82	78%	17,80	17,10	82%
20	11,87	8,90	85%	11,64	9,23	84%	17,47	12,30	89%
30	11,87	7,42	90%	11,74	7,21	90%	17,61	10,33	93%
40	11,85	6,72	92%	11,68	6,17	93%	17,66	9,06	95%
50	11,72	5,63	95%	11,73	5,48	95%	17,81	8,21	97%
60	11,79	5,10	97%	11,80	5,32	96%	17,76	6,78	99%

Mean	11,763

Table X: Values underlying the Probabilistic Sampling Simulation for HMW, Site C.

	HIGH MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	12,12	11,04	81%	12,29	10,48	82%	16,96	14,47	84%
20	12,22	7,73	90%	12,36	8,42	88%	16,86	9,74	93%
30	12,23	6,41	94%	12,33	6,50	93%	16,32	7,64	96%
40	12,23	5,63	96%	12,26	5,51	96%	16,44	6,78	98%
50	12,28	4,90	98%	12,30	5,11	97%	16,58	5,91	99%
60	12,24	4,78	98%	12,33	5,11	97%	16,62	5,51	99%

Mean	12,253

Table Y: Values underlying the Probabilistic Sampling Simulation for LMW, Site D.

	LOW MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	0,36	0,31	100%	0,35	0,28	100%	0,31	0,38	100%
20	0,36	0,24	100%	0,38	0,21	100%	0,31	0,22	100%
30	0,36	0,19	100%	0,36	0,18	100%	0,31	0,19	100%
40	0,36	0,16	100%	0,36	0,15	100%	0,31	0,17	100%
50	0,36	0,13	100%	0,36	0,16	100%	0,31	0,14	100%
60	0,36	0,13	100%	0,36	0,12	100%	0,38	0,13	100%

Mean	0,362241

Appendix 9: Sampling Simulations Data

Table Z: Values underlying the Probabilistic Sampling Simulation for MMW, Site D.

	MIDDLE MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	5,98	6,34	75%	6,16	6,00	77%	6,61	6,40	77%
20	6,02	6,48	74%	6,21	6,01	77%	6,62	4,42	86%
30	6,02	3,77	87%	5,96	3,59	88%	6,66	3,71	91%
40	5,97	2,94	92%	5,94	2,98	92%	6,65	3,39	92%
50	5,97	2,82	93%	6,04	3,18	91%	6,66	2,94	95%
60	6,02	2,59	95%	6,07	2,68	95%	6,61	2,69	96%

Mean	5,971379

Table AA: Values underlying the Probabilistic Sampling Simulation for HMW, Site D.

	HIGH MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	5,93	4,68	81%	6,31	6,07	77%	6,74	5,94	80%
20	6,03	3,98	86%	6,06	3,64	88%	6,67	3,83	90%
30	5,99	2,93	92%	6,00	2,98	92%	6,72	3,45	92%
40	6,00	2,64	94%	6,05	2,61	95%	6,75	3,07	95%
50	5,96	2,25	97%	6,07	2,27	97%	6,73	2,61	97%
60	5,98	2,07	98%	6,10	2,12	98%	6,69	2,47	98%

Mean	6,004957

Table BB: Values underlying the Probabilistic Sampling Simulation for LMW, Site E.

	LOW MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	1,99	1,40	83%	1,84	1,17	84%	2,05	1,35	84%
20	1,99	0,93	92%	2,04	0,85	94%	2,05	0,92	93%
30	1,97	0,72	96%	1,98	0,70	97%	2,10	0,73	97%
40	1,98	0,63	98%	2,05	0,62	99%	2,08	0,64	99%
50	1,97	0,56	99%	2,01	0,57	99%	2,09	0,60	99%
60	1,98	0,51	99%	1,98	0,51	99%	2,07	0,52	100%

Mean	1,977115

Appendix 9: Sampling Simulations Data

Table CC: Values underlying the Probabilistic Sampling Simulation for MMW, Site E.

	MIDDLE MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	6,16	5,16	80%	6,02	5,25	79%	6,52	5,21	82%
20	6,14	3,51	89%	6,12	3,39	90%	6,59	3,72	90%
30	6,16	2,94	93%	6,05	2,88	93%	6,61	3,12	94%
40	6,17	2,52	96%	6,14	2,60	95%	6,58	2,65	96%
50	6,14	2,25	97%	6,09	2,30	97%	6,57	2,43	97%
60	6,18	2,11	98%	6,06	2,05	98%	6,58	2,13	99%

Mean	6,160192

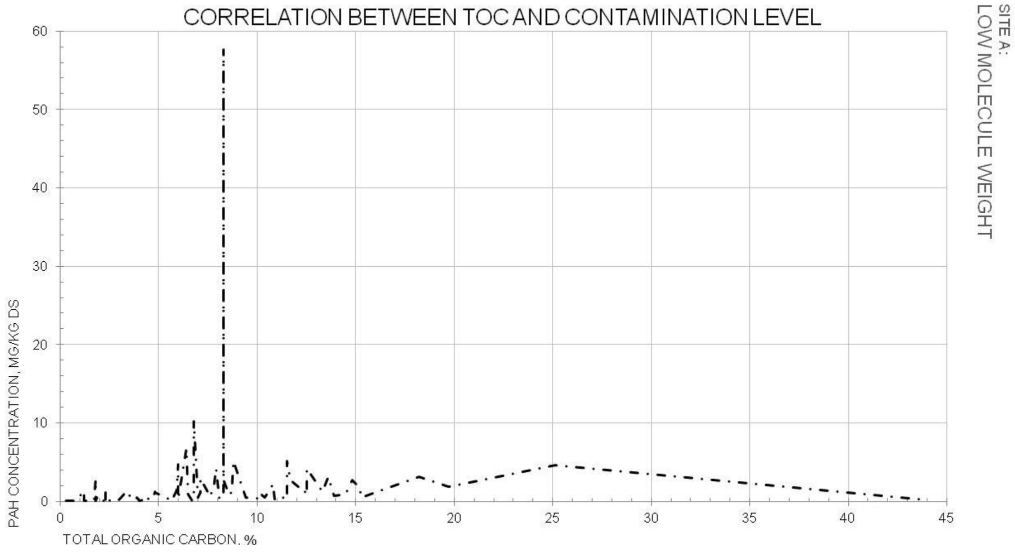
Table DD: Values underlying the Probabilistic Sampling Simulation for HMW, Site E.

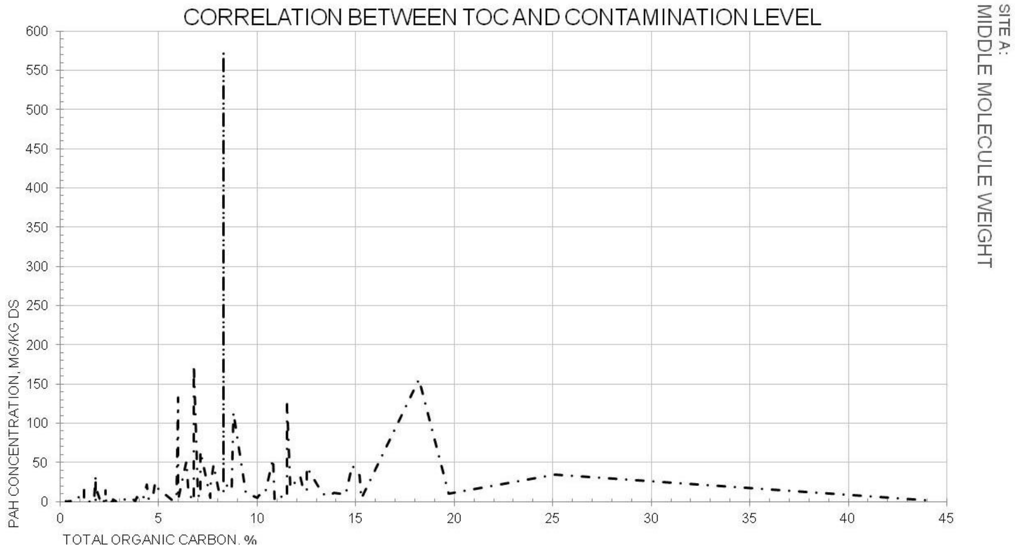
	HIGH MOLECULAR WEIGHT								
	SRS			PSRS			OSRS		
n	Mean	Std	Lognorm	Mean	Std	Lognorm	Mean	Std	Lognorm
10	4,59	4,49	75%	4,42	3,56	79%	4,99	4,77	77%
20	4,57	2,69	87%	4,43	2,54	87%	4,92	2,80	89%
30	4,54	2,35	90%	4,47	2,24	91%	4,94	2,44	92%
40	4,54	1,93	94%	4,45	1,85	94%	4,90	2,04	95%
50	4,53	1,72	96%	4,44	1,70	96%	4,90	1,72	98%
60	4,57	1,59	97%	4,46	1,58	97%	4,83	1,63	98%

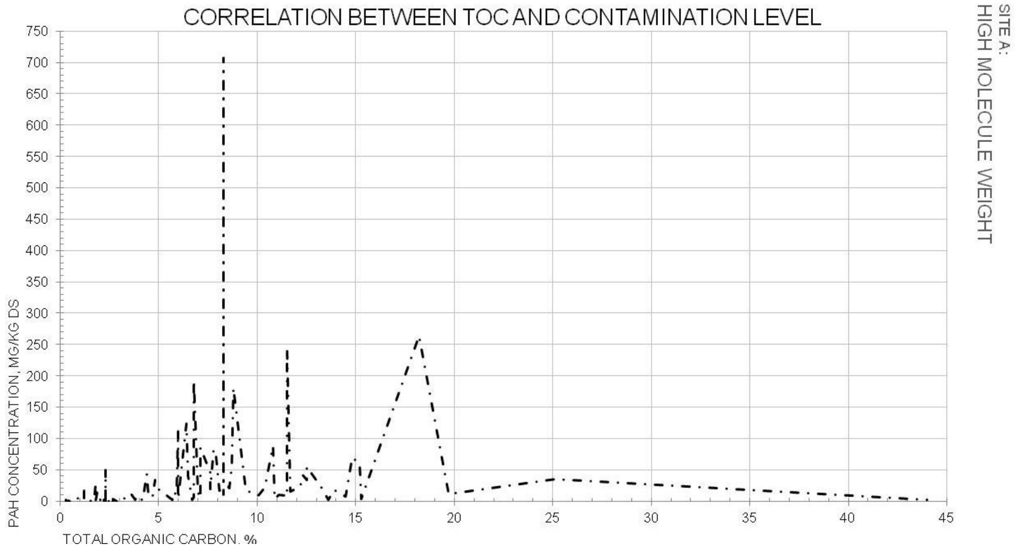
Mean	4,549423

Appendix 10: TOC Analysis

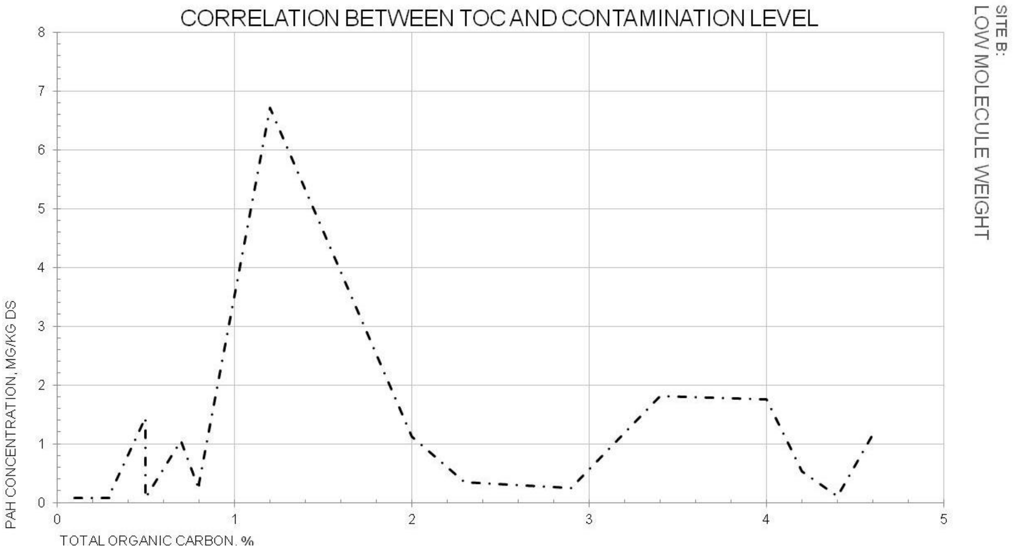
TOC Analysis

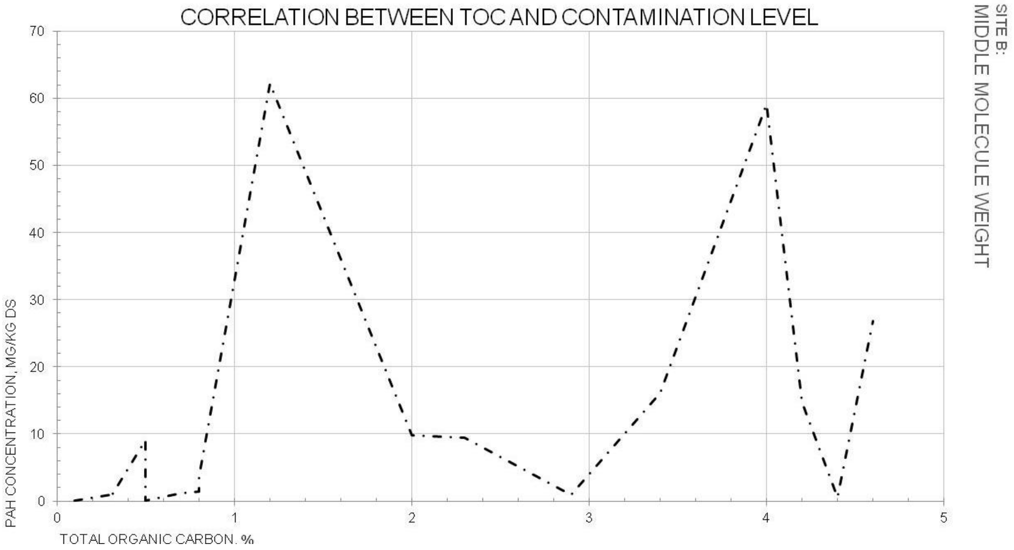


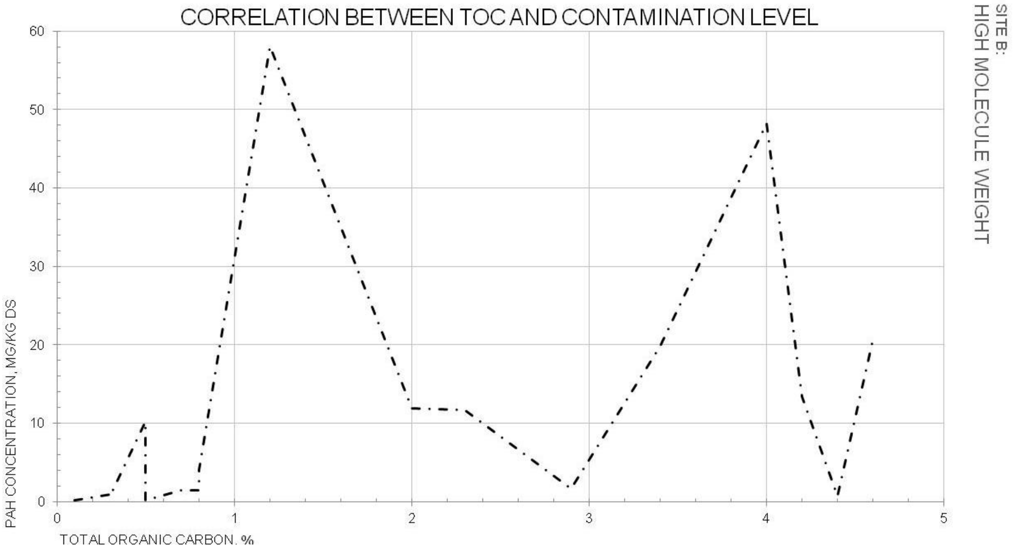




Appendix 10: TOC Analysis







Appendix 10: TOC Analysis