

Estimation of Measurement Uncertainty in the Sampling of Contaminated Land

By
Ariadni Argyraki

A thesis submitted for the Degree of Doctor of Philosophy
of the University of London and for the Diploma of Imperial College.

Environmental Geochemistry Research Group
Department of Geology
Royal School of Mines
Imperial College
London



May 1997



to my mother

ABSTRACT

This thesis presents the findings of research on the estimation of measurement uncertainty in the sampling of contaminated land. Sampling is the first step in every geochemical investigation and the need for quality in this process is becoming well recognised. The traditional approach requires the "correct" collection of the samples in order to represent as close as possible the population under study. However, sampling is inevitably an error generating process that contributes to the total measurement uncertainty of the results of contaminated land investigations and affects the interpretation of the data. In this project no attempt is made to provide guidelines for sampling strategies. The process of sampling has been studied in respect to the error that it generates, regardless of the protocol used.

The sampling target identified is always related to the objectives of the environmental sampling plans and the sampling protocols that can be employed are diverse. However, the general principles of sampling theory have always to be kept in mind. The implementation of a methodology for the quantification of sampling uncertainty should have the potential to be applied regardless of the sampling target, with the flexibility to be adapted to different circumstances. Field sampling of contaminated soil has been used as an example for testing the general theoretical methodology developed in this research. This medium was selected as a common sampling material in environmental studies with well developed sampling techniques.

Four methods, of varying complexity, for estimating measurement uncertainty arising from field sampling have been developed in this research and tested at two sites in the UK contaminated with heavy metals. These methods differ in the manner in which the samples are collected and demonstrate substantial values of measurement uncertainty on the average metal concentrations up to ± 84 % for Pb, depending on the degree of heterogeneity of the sampling target. Aspects such as inter-organisational trials have been applied to soil sampling for the first time for estimating and quantifying the uncertainty introduced in the estimation of the average metal concentration in the soil. These techniques, previously developed and used for the estimation of the analytical error, were adapted and applied to sampling. The role of measurement uncertainty in evaluating an in-situ measurement technique (portable X-ray fluorescence) has also been examined. Finally, a probabilistic classification of contaminated land taking into account the estimated uncertainty of the concentration measurements, as well as their values, has been suggested.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Dr. Mike Ramsey for his invaluable advice, guidance, assistance and inspiration throughout the project and also for his great help during the field work.

This research project has been financially sponsored by the Greek State Scholarship Foundation (I.K.Y.) without whose financial backing I could not have carried out this research. A big thank you goes also to my mother and family in Greece not only for their moral but also for the financial support.

On the technical side, I would particularly like to thank Barry Coles for all the ICP-AES analyses and Aldban Doyle for technical assistance during all sample preparation. Also, I would like to thank all the people that participated in the two sampling inter-organisational trials organised during this research, for their time and effort during the field experiments. Also I would like to thank the local authorities of Hounslow for giving permission to use the recreational area of Hounslow Heath as an experimental site and Dr. Phil Potts of the Open University for his invaluable help during the field experimentation using the portable XRF in Derbyshire. Dr. Phil Potts must also be acknowledged for providing his expertise during the analysis of the samples.

Last but not least I would like to thank all the students in the Environmental Geochemistry Research Group for the friendly and very welcome environment during my studying in Imperial College, and all my friends and particularly Christos and Katerina for they have been like family for me during these years.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	ix
List of Figures	xii
List of Abbreviations	xvi
 CHAPTER ONE. INTRODUCTION	 1
1.1 Background of the Research	1
1.2 Context of Research and Objectives	3
1.3 Structure of Thesis	6
 CHAPTER TWO. LITERATURE REVIEW ON SAMPLING AND UNCERTAINTY	 7
2.1 Introduction	7
2.2 Terminology	7
2.3 Sampling Theory	9
2.3.1 Application of Sampling Theory to Sampling Surveys	9
2.3.2 Sampling Strategies	11
2.3.3 Environmental Sampling of Contaminated Land	14
2.3.4 Geostatistical Approach for Sampling and Estimation	16
2.4 Estimation of Sampling Errors	17
2.5 Estimation of Uncertainty from Analytical Measurements	18
2.5.1 Classical Evaluation of Analytical Uncertainty Using the 'Bottom-up' Method	19
2.5.2 Evaluation of Analytical Uncertainty Using Inter-laboratory Trials, the 'Top-down' Method	20
 CHAPTER THREE. METHODOLOGY	 23
3.1 Introduction	23
3.2 Sampling Methodology	
3.2.1 Collection of Soil Samples in Initial Surveys	23
3.2.2 Collection of Samples in Inter-organisational Trials	25
3.2.2.1 The Collaborative Trial in Sampling (CTS)	26
3.2.2.2 The Sampling Proficiency Test	28
3.2.3 Sample Preparation	30
3.3 Analytical Methodology	30
3.3.1 Preparation of Analytical Equipment	30
3.3.2 Multi-element Analysis of Soil Samples	31
3.3.3 Estimation of Analytical Uncertainty and Quality Control	31
 CHAPTER FOUR. ESTIMATION OF SAMPLING UNCERTAINTY AT AN OLD SMELTING CONTAMINATED SITE IN THE ENGLISH MIDLANDS: INITIAL SURVEY	 34
4.1 Introduction	34
4.2 Site Description	34

4.3	One Sampler Using Many Sampling Protocols	37
4.3.1	Statistical Characteristics of Pb and Cu Distributions	39
4.3.2	Estimation of Random Errors from the Initial Survey at the Derbyshire Site	46
4.3.3	Estimation of Systematic Errors from Analysis from the Initial Survey at the Derbyshire Site	50
4.3.4	Estimation of Measurement Uncertainty from the Initial Survey in Derbyshire: Method 1	51
4.3.5	Estimation of Measurement Uncertainty from the Initial Survey in Derbyshire: Method 2	53
4.4	Conclusions from the Initial Survey at the Derbyshire Site	54
CHAPTER FIVE. FIRST COLLABORATIVE TRIAL IN SAMPLING UNDERTAKEN AT THE DERBYSHIRE SMELTING SITE		55
5.1	Introduction	55
5.2	Experimental	57
5.2.1	Sample Collection and Preparation	57
5.2.2	Analytical Quality Control on the Data	58
5.3	Results from the First Collaborative Trial in Sampling	60
5.3.1	Statistical Interpretation of the Data	60
5.3.2	Estimation of Measurement Uncertainty Using the CTS Data: Method 3	64
5.3.3	Comparison of Sampling and Analytical Collaborative Trials	65
5.4	Conclusions of the First Collaborative Trial in Sampling	67
CHAPTER SIX. FIRST SAMPLING PROFICIENCY TEST UNDERTAKEN AT THE DERBYSHIRE SMELTING SITE		69
6.1	Introduction	69
6.2	Experimental	70
6.2.1	Sampling Target and Sampling Procedures	70
6.2.2	Preparation of HRM 31	72
6.3	Results from the First Sampling Proficiency Test	75
6.3.1	The z-score Assessment	76
6.3.2	Separation of Sources of Bias	79
6.3.3	Estimation of Measurement Uncertainty Using the SPT Data: Method 4	81
6.3.4	Effect of the Heterogeneity of the Sampling Target	82
6.3.5	Comparison of Sampling and Analytical Proficiency Testing	84
6.4	Conclusions of the First Sampling Proficiency Test	86
CHAPTER SEVEN. UNCERTAINTY FROM ITERATIVE SAMPLING DESIGNS USING PORTABLE X-RAY FLUORESCENCE		88
7.1	Introduction	88
7.2	Background Information on Portable XRF Methodology	89
7.3	Experimental Design	90
7.3.1	Instrumentation	90
7.3.2	Method Development and Selection of Sampling and Analytical Protocol	91
7.4	Results	94

7.4.1	Analytical and Sampling Variances	95
7.4.2	Bias of Portable XRF Measurements	98
7.4.3	Origin of Bias from the Field Portable XRF Measurements	101
7.4.4	Mapping of Contamination Using P-XRF	104
7.5	Advantages and Disadvantages of the P-XRF Methodology	107
7.6	Conclusions on the Use of the Portable XRF Methodology	109

CHAPTER EIGHT. ESTIMATION OF SAMPLING UNCERTAINTY AT A COMPLETED LANDFILL SITE IN WEST LONDON 112

8.1	Introduction	112
8.2	Reconnaissance Survey	113
8.2.1	Site Description	113
8.2.2	Sampling Strategy for the Pilot Study in Hounslow Heath	115
8.2.3	Statistical Characteristics of Pb and Zn Distributions	116
8.2.3.1	Estimation of Random Errors from the Initial Survey at the Hounslow Heath Site	117
8.2.3.2	Estimation of Systematic Errors from Analysis from the Initial Survey at the Hounslow Heath Site	119
8.2.3.2	Statistical Characteristics of Pb and Zn Distributions	120
8.2.4	Conclusions from the Pilot Study in Hounslow Heath	123
8.3	Second Inter-Organisational Sampling Trial on Contaminated Land, at Hounslow Heath	124
8.3.1	Introduction	124
8.3.1.1	Analytical Quality Control of the Data	125
8.3.2	Second Collaborative Trial in Sampling	127
8.3.2.1	Statistical Interpretation of the Data from the Second CTS	127
8.3.2.2	Estimation of Measurement Uncertainty Using the Data from the Second CTS	131
8.3.3	Second Proficiency Test in Sampling	132
8.3.3.1	Sampling Target and Sampling Procedures	132
8.3.3.2	Statistical Interpretation of the Results from the Second SPT	135
8.3.3.3	Estimation of Measurement Uncertainty Using the Second SPT Data	139
8.3.4	Conclusions from the Second Sampling Trial in Hounslow Heath	140

CHAPTER NINE. DISCUSSION OF ESTIMATING MEASUREMENT UNCERTAINTY FROM SAMPLING 141

9.1	Introduction	141
9.2	Comparison of the Results from the Two Inter-organisational Sampling Trials on Contaminated Land	141
9.2.1	Comparison of the Results from the Two CTS	142
9.2.2	Comparison of the Results from the Two SPTs	144
9.3	Suggested Methodology for the Estimation of Measurement Uncertainty in the Sampling of Contaminated Land	146
9.3.1	Four Methods for the Estimation of Measurement Uncertainty	146
9.3.2	Acceptable Levels of Uncertainty: Fitness for Purpose	151
9.3.3	Uncertainty on the Vertical Variability of Metal Concentration	153
9.4	Probabilistic Classification of Contaminated Land	154

9.4.1	Application of Probabilistic Classification to Contaminated Land	157
9.4.2	Example of Probabilistic Classification	159
9.4.3	Implications of Probabilistic Classification	161
9.5	Spatial Uncertainty	163
9.5.1	Variograms and Kriging	163
9.5.2	Spatial Interpretation of the Experimental Data	166
9.5.3	Comparison of Uncertainty Calculated from Duplicate Samples and the Plotted Variogram	167
9.6	Potential Application of the Suggested Methodology for Uncertainty Estimation for Other Sampling Media	169
9.6.1	Sampling of Bulk Materials	169
9.6.2	Sampling of Temporally Variable Targets, Such as Water	170

CHAPTER TEN. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER STUDY 172

10.1	Introduction	172
10.2	The Initial Surveys	172
10.2.1	The Old Smelting Site in Derbyshire	172
10.2.2	The Completed Landfill Site in Hounslow Heath, London	174
10.3	Portable XRF Methodology and Sampling Uncertainty	174
10.4	Inter-organisational Sampling Trials	176
10.4.1	The Collaborative Trial in Sampling and the Sampling Proficiency Test	176
10.4.2	Four Methods for the Estimation of Measurement Uncertainty in Field Sampling	178
10.4.3	Probabilistic Classification of Contaminated Land	180
10.5	Recommendations for Further Study	181

APPENDIX 1.	Analytical Method	192
APPENDIX 2.	Total geochemical data	194
APPENDIX 3.	Duplicate Data and ANOVA Results	220
APPENDIX 4.	Statistical Tests	233
LIST OF PUBLICATIONS		239

LIST OF TABLES

CHAPTER THREE

Table 3.1. List of Certified reference materials (CRMs) used for the estimation of analytical bias for all ICP-AES analyses in this research. 32

Table 3.2. List of house reference materials (HRMs) used for the estimation of analytical bias for all ICP-AES analyses in this research. 32

CHAPTER FOUR

Table 4.1. Descriptive statistics for Pb and Cu in soil from the initial survey in Derbyshire (number of samples, $n = 119$). 42

Table 4.2. Pb and Cu concentrations estimated from the four different sampling protocols and difference (%) from the grand mean of $6233 \mu\text{g g}^{-1}$ for Pb and $27.9 \mu\text{g g}^{-1}$ for Cu (n , number of samples for each design). 44

Table 4.3. Sampling and analytical precisions (as percentages of the mean) for Pb and Cu in soil calculated from the robust ANOVA results on the duplicate samples and analyses in the Derbyshire site. 47

Table 4.4. Analytical bias for Pb and Cu concentration estimated from six CRMs ($\bar{x}$, mean concentration; s , standard deviation; and c , certified values). 50

CHAPTER FIVE

Table 5.1 Results of the analysis of control materials for the first collaborative trial in sampling in Derbyshire. Concentration and bias values in $\mu\text{g g}^{-1}$. 58

Table 5.2. Elemental concentration values from the first Collaborative Trial in Sampling in Derbyshire. Values in $\mu\text{g g}^{-1}$. 62

Table 5.3: Statistics from the first collaborative trial in sampling for Cu and Pb (mg g^{-1}). (s_0 : repeatability standard deviation; $s_1 = s_{r(s)}$: sampling repeatability standard deviation; s_2 : between-sampler standard deviation; $s_{R(s)}$: sampling reproducibility standard deviation; $\text{RSD}_0, \text{RSD}_{r(s)}, \text{RSD}_{R(s)}$: relative standard deviations). 63

Table 5.4. Standard deviations and expanded uncertainties for Pb and Cu in soil, assuming coverage factor $k=2$). Data from the first CTS in Derbyshire. 64

CHAPTER SIX

Table 6.1. Sampling and analytical procedures used by the nine organisations for the estimation of the average Pb and Cu concentration in soil.	72
Table 6.2. Elemental concentrations and certified values from the analysis of the six CRMs. Concentrations in $\mu\text{g g}^{-1}$. Data from the preparation of HRM 31.	73
Table 6.3. MLFR data for Pb and Cu showing statistically significant rotational bias in the HRM 31 samples.	74
Table 6.4. Reported analytical results for the first sampling proficiency test at the Derbyshire site, (concentrations in $\mu\text{g g}^{-1}$).	75
Table 6.5. Bias of the reported results from the assigned value and from the grand mean for Pb and Cu. Data from the first SPT in Derbyshire. * denotes bias statistically significant with 95% confidence interval.	78
Table 6.6: Bias of HRM 31 reported results from the assigned values of 6895 mg g^{-1} Pb and 30 mg g^{-1} Cu. Data from the first SPT in Derbyshire.	79
Table 6.7. Lead and Cu results after the correction for analytical bias using the HRM 31 results. Data from the first SPT in Derbyshire.	80
Table 6.8: Statistical characteristics of reported and corrected Pb in soil with the outlying result of organisation six included and removed. Data from the first SPT in Derbyshire.	81

CHAPTER SEVEN

Table 7.1. Analytical bias estimated with CRM NBS2710 by the three analytical methods (field P-XRF, lab P-XRF, ICP-AES).	98
---	----

CHAPTER EIGHT

Table 8.1. Estimates of extended uncertainty ($k = 2$) for Pb and Zn in soil from the initial survey of part of Hounslow Heath.	117
Table 8.2. Analytical bias for Pb and Zn concentration estimated from six CRMs ($\bar{x}$, mean concentration; s , standard deviation; and c , certified values in $\mu\text{g g}^{-1}$ unless otherwise stated). * denotes bias statistically significant with 95% confidence interval.	119
Table 8.3 Descriptive statistics for Pb and Zn in soil in the initial survey of Hounslow Heath.	121

Table 8.4. Analytical bias estimated from six CRMs for the sampling trial in Hounslow Heath. (ACC.V. = accepted value, t_{tab} = Student's t for 95 % confidence., st.si.= statistical significance).	126
--	-----

Table 8.5. Estimated concentration values from the second collaborative trial in sampling. Concentrations in $\mu\text{g g}^{-1}$.	128
--	-----

Table 8.6. Statistics from the second collaborative trial in sampling in Hounslow Heath.	130
---	-----

Table 8.7. Standard deviations and extended uncertainties for Pb, Zn, Cd and Co in soil, assuming coverage factor $k=2$. Data from the second CTS.	131
--	-----

Table 8.8. Sampling procedures used by the nine organisations for the second proficiency test.	134
---	-----

Table 8.9. Analytical results from the second SPT.	136
---	-----

CHAPTER NINE

Table 9.1. Expanded uncertainties for Pb, Cu, Co in soil from the two CTS, assuming coverage factor $k = 2$.	143
--	-----

Table 9.2. Standard deviations and expanded uncertainties for Pb, Cu, Co concentrations in soil, assuming coverage factor of 2 ($U= 2s$). Data from the two SPTs.	146
--	-----

Table 9.3. Probabilistic classification of soils with measured concentration (C) and uncertainty (U), based on the probability that the contaminant concentration is greater than a particular threshold value (T). Uncertainty values are stated for 95% confidence interval, but could be recalculated for what ever confidence interval is considered appropriate.	157
--	-----

Table 9.4 Average Pb concentration ($\bar{x}$) and total variance (s_{tot}^2) calculated from ANOVA and the plotted variograms.(n; number of samples).	168
---	-----

LIST OF FIGURES

CHAPTER THREE

- Figure 3.1.** Schematic sampling designs used in the initial surveys of the research.(a) regular grid, (b) composite regular grid, (c) stratified random, (d) herringbone design. 25
- Figure 3.2.** Schematic presentation of the sampling designs for the two CTS. (a) ‘W’ design for the first CTS, (b) regular grid for the second CTS. 26
- Figure 3.3.** Schematic sampling protocols used by the participants in the two SPTs organised for this project. 29

CHAPTER FOUR

- Figure 4.1.** Geographical location of the study area in Derbyshire. 35
- Figure 4.2.** The four sampling protocols applied to sample top soil at the Derbyshire site. (A) simple regular grid, (b) composite regular grid, (c) herringbone design, (d) stratified random design. 38
- Figure 4.3.** Frequency distribution of Pb concentration in soil at the Derbyshire site, showing positively skewed population (a). Data from all the single auger sampling designs. (b), Log-transformed data from all the designs. 40
- Figure 4.4.** Frequency distribution of Cu concentration in soil at the Derbyshire site, showing positively skewed population (a). Data from all the single auger sampling designs. (b), Log-transformed data from all the designs. 41
- Figure 4.5.** Dot maps of Pb (a) and Cu (b) concentration at the sampling points in the Derbyshire site. 43
- Figure 4.6.** Schematic balanced design of sampling and analytical duplicates. 46
- Figure 4.7.** Relative importance of measurement errors (from sampling and analysis) and geochemical variability, in the Derbyshire site both expressed as proportions of total variance. 49

CHAPTER FIVE

- Figure 5.1.** Schematic design of the collaborative trial in sampling. 56
- Figure 5.2** Map of the sampling target for the first collaborative trial in sampling in Derbyshire, showing the four possible “W” patterns for

collecting the sample increments.	57
Figure 5.3. Graphic representation of the results of the first collaborative trial in sampling. Each point shows the mean of the duplicated analytical determinations.	61
 CHAPTER SIX	
Figure 6.1: Graphic presentation of the nine different sampling designs used by the participants to sample top soil in the rectangular field in Derbyshire for the first sampling proficiency test.	71
Figure 6.2. z-scores for Pb and Cu in soil from the first sampling proficiency test, showing unsatisfactory performance for participants 6 and 2 for Pb and Cu respectively.	77
 CHAPTER SEVEN	
Figure 7.1. Schematic representation of P-XRF instrumentation for in-situ analysis.	91
Figure 7.2. Sampling design for the P-XRF measurements and collection of auger samples in the Derbyshire site. 30 m, 10 m, 1 m ² grids.	93
Figure 7.3. ANOVA pie charts for Pb in soil, showing acceptable proportions of measurement variances over the total for all three analytical methods. (a) P-XRF field data (n=24), (b) P-XRF lab data (n=10), (c) ICP-AES data (n=10). (n = number of duplicate samples).	96
Figure 7.4. ANOVA pie charts for Zn in soil, showing acceptable proportions of measurement variances over the total for all three analytical methods. (a) P-XRF field data (n=24), (b) P-XRF lab data (n=10), (c) ICP-AES data (n=10). (n = number of duplicate samples).	97
Figure 7.5. (a) Simple linear regression of Pb field P-XRF measurements against Pb ICP-AES measurements, (b) Simple linear regression for Zn.	99
Figure 7.6. Schematic representation of the 'core experiment' set-up.	101
Figure 7.7. P-XRF core experiment: Pb in soil as determined in the cores by the two analytical methods. (a) core 1, (b) core 2. Zn in soil as determined in the cores; (c) core 1, (d) core 2.	103
Figure 7.8. Spatial mapping of Pb concentration in soil determined by P-XRF, showing the iterative sampling design with initial 30 m grid followed by the 10 m grid centred on candidate 'hot spots' (Pb > 8000 µg g ⁻¹). The eastern	

'hot spot' is confirmed by a cluster of high values whereas the western 'hot spot' is not confirmed. 105

Figure 7.9. Concentration of Pb in soil in the 1 m² area over the contamination hot spot on the eastern side of the field, showing no systematic trend in the highest values. 106

Figure 7.10. Concentration of Pb in soil in the 1 m² area on the western side of the field, where the lowest metal concentrations were observed, showing no systematic trend in the highest values. 107

CHAPTER EIGHT

Figure 8.1. Geographical location of the study area in West London. 114

Figure 8.2. 30 x30 m grid for the pilot study in Hounslow Heath. 116

Figure 8.3. ANOVA pie charts for Pb and Zn in soil from the ten duplicate samples collected in the pilot study in Hounslow Heath. 118

Figure 8.4. Frequency distributions for Pb (a) and Zn (b) from the samples of the pilot study in Hounslow Heath. 120

Figure 8.5. Dot maps of Pb (a) and Zn (b) concentrations in soil from the pilot study in Hounslow Heath, classified using the probabilistic approach that allows for measurement uncertainty. 122

Figure 8.6. Graphical presentation of the results from the second CTS in Hounslow Heath. Each point shows the mean of the duplicated analytical determinations. 129

Figure 8.7. Graphic presentation of the nine different sampling designs used by the participants to sample top soil in the square area for the second proficiency test. 133

Figure 8.8. Graphic representation of the analytical results from the second SPT. 135

Figure 8.9. z-scores from the second sampling proficiency test in Hounslow Heath. 138

CHAPTER NINE

Figure 9.1. Probabilistic classification of contaminated soils, allowing for uncertainty on either the measured concentration values (*a*), or around the threshold value (*b*) for a rapid method of classification. 156

Figure 9.2. Geochemical maps of Pb contamination in the Hounslow Heath site, using a threshold value of $500 \mu\text{g g}^{-1}$. The deterministic approach (*a*), shows a few isolated locations of contamination over the threshold. The probabilistic map (*b*), shows that almost the entire area is possibly contaminated over the same threshold, because of the very high measurement uncertainty of the survey.

160

Figure 9.3. Experimental variograms and the fitted spherical models of the data from the Derbyshire site (*a*) and the Hounslow Heath site (*b*).

167

LIST OF ABBREVIATIONS

Pb	Lead
Zn	Zinc
Cu	Copper
Co	Cobalt
Cd	Cadmium

ANOVA	Analysis of variance
MLFR	Maximum likelihood functional relationship

XRF	X-ray fluorescence spectroscopy
P-XRF	Portable X-ray fluorescence spectroscopy
ICP-AES	Inductively coupled plasma atomic emission spectrometry
AAS	Atomic absorption spectroscopy
CRM	Certified reference material
HRM	In-House reference material

RST	Reference sampling target
CTS	Collaborative trial in sampling
SPT	Sampling proficiency test

mm	Millimetre
cm	Centimetre
m	Meter
km	Kilometre
$\mu\text{g g}^{-1}$	Microgram per gram

CHAPTER ONE. INTRODUCTION

1.1 Background of the Research

The scope of geochemistry as defined by Goldschmidt is the study of the distribution and the amounts of the chemical elements in minerals, ores, rocks, soils, waters and the atmosphere, and the circulation mechanisms of the elements in nature (Goldschmidt, 1954). Therefore the development of methods for measuring any property of the chemical elements is the basic and essential step for understanding their geochemical behaviour. Probably more important than simple estimation of concentration values is a realistic assessment of the uncertainty on these estimates, for avoiding the possible misinterpretation of the geochemical data.

The surface environment, closest to man, has dominant importance. Environmental geochemistry is concerned with the sources, distribution and interactions of chemical elements in the system rock-soil-water-air-life (Thornton, 1983). The redistribution of the elements in the surficial environment results in lower concentrations and therefore demands sophisticated measurement methodologies for quantification. The estimation of the true values of analyte concentration often plays a significant role. For example in contaminated land investigations regulatory limits are the guidance for costly decisions concerning remediation of a contaminated site. It is important therefore that concentration estimates are determined with a specified level of confidence. A sound methodology for the quantification of the uncertainties on concentration measurements will therefore have financial implications.

This research is concerned with the estimation of uncertainty in the measurement process during contaminated land investigations. Contaminated land can be defined as land containing harmful substances (pollutants) introduced into the environment by man, liable to cause hazards to human health, harm to living resources and ecological systems, damage to structures or amenity, or interference with legitimate uses of the environment (Alloway and Ayres, 1997).

The process of measuring the concentration of pollutants usually comprises the collection of individual field samples (called sample increments), the preparation of test material in the laboratory and the chemical analysis of the derived test portions. The objective is usually to estimate the true values of the analyte concentration. The mean concentration across a whole site indicates the existence of geochemical anomalies when it is compared to a threshold value. However there is always some error affecting the estimated concentration. The error, originating from both sampling and chemical analysis (measurement error), may significantly affect all subsequent interpretation of the geochemical data.

Clarification is needed for the use of the terms error and uncertainty. These terms are defined in the next chapter of this thesis, however it is sufficient to clarify here that uncertainty is a parameter associated to the result of the measurement process and error is related to the methodology leading to the quantification of the measurement value. There are two types of measurement errors that contribute to uncertainty in concentration estimates. Random errors are related to the precision of repeated measurements which expresses the closeness of agreement between independent test results obtained under prescribed conditions (ISO Guide 30: 1992). Random errors, when they are normally distributed, are quantified with the standard deviation. Systematic errors are related to the closeness of the agreement between the result of a measuring procedure and the accepted reference value. Systematic errors are quantified with bias which is the difference between the expectation of the test results and an accepted reference value (ISO 3534- 1: 1993).

Methodologies for the estimation of the random and systematic error in analysis have been well established. Duplicate analyses for the estimation of precision (Garrett, 1969) and the use of reference materials for the estimation of bias (Ramsey et al., 1987) can give a sound estimate of the two types of error. A methodology comprising the collection of duplicate samples, has been developed for the estimation of random error in sampling (Garrett, 1969; Ramsey, 1993). The estimation of the systematic error in environmental sampling is a subject hardly covered in literature. Gy (1992), in his work on sampling of particulate materials considers sampling as an error generating process and clearly states that sampling correctness is the only effective way to avoid sampling bias. However, no method has been described for the estimation of sampling bias in environmental investigations.

1.2 Context of Research and Objectives

The context of this project is the estimation of the uncertainty introduced during the sampling process in contaminated land investigations. For the purposes of this research two examples of contaminated land sites bearing heavy metal contamination, have been used but the knowledge gained from these studies is valuable for more general aspects for the quantification of measurement uncertainty.

Both random and systematic errors have been under consideration in this project. Investigations of these two components cannot be conducted in isolation as they are often interrelated. Fitness for purpose, i.e. the degree to which data produced by a measurement process enables a user to make technically correct decisions for a stated purpose (AMC, 1995) has also to be kept in mind. Gains in accuracy for example, have always to be in balance with the effort and the cost of the measurement method.

The importance of this study to environmental investigations is apparent as the errors introduced into the measurements during sampling are related to:

- a) the classification of contaminated land which has financial implications,
- b) the potential toxicity of the soil with possible effects on human health,
- c) the analytical measurements that can be effectively invalidated due to sampling errors.

The primary objectives of the research are listed here and discussed below:

1. Development and implementation of methods for the quantification of measurement uncertainty arising from field sampling, in contaminated land investigations.
2. Evaluation of the contribution of sampling uncertainty for two specific sites by these methods of varying complexity, for estimating the average metal concentration in soil.
3. Assess the role of measurement uncertainty in evaluating fitness-for-purpose of an in-situ measurement technique on contaminated land (i.e. portable x-ray fluorescence).
4. Consider the general applicability of these methodologies for sampling other media.

Secondary objectives of this research, more related to the practice of the suggested methodology are listed below:

1. Test the feasibility of applying methodologies of uncertainty estimation, originally devised for use with chemical analysis, to sampling.
2. Investigate the relationship between the results obtained from the new methods.
3. Consider the feasibility of estimating absolute sampling bias in environmental investigations by establishing a Reference Sampling Target (RST) (e.g. a standard field).

The need for quality in the sampling process of contaminated land is well recognised. The first objective of this research is the development of a methodology for the quantification of uncertainty due to field sampling. Sampling bias is an especially difficult component of error to estimate, mainly because the 'true' value of an analyte concentration at a site can never be known exactly. The use of field blanks, control sites and field spike samples suggested by Black (1988) was a first attempt at this goal, but it is incomplete and not always practical depending on the type of the analyte. However the analogy with chemical analysis suggests that it is possible to adapt the methodology regardless the sampling medium (Ramsey, 1994, Thompson and Ramsey, 1995).

In chemical analysis the main method used for the estimation of bias is the analysis of samples with known analyte concentrations, namely certified reference materials (CRMs). Essential features of these materials are homogeneity, stability and a certified property value stated with its uncertainty at a level of confidence (ISO Guide 30: 1992). Extensive testing for each one of these characteristics comprises repeated analysis over time. Certification trials which involve a number of experienced laboratories employing a variety of analytical methods are often used to provide the assigned or certified value. The analogous approach in sampling would be to define sites or piles of bulk material functioning as reference sampling targets (Thompson and Ramsey, 1995). Methodologies such as comparative sampling and analysis in the form of inter-organisational trials could be performed at these sites. The feasibility of applying such techniques to sampling, that have previously developed and used for the estimation of analytical error, is a secondary objective of the research, linked to the primary objective of finding a way to estimate measurement uncertainty in the sampling of contaminated land.

The simplest and easiest objective for such trials is to estimate the average analyte concentration at a site. The estimation of uncertainty could be approached successively

from the simplest case where one sampler is using one or several protocols to collect the samples, to more elaborate ones, where several independent samplers could sample the field by either one or several different designs. For the estimation of the average concentration, one composite sample from each sampler would be enough. Such compositing saves time for sampling, preparation and analysis and reduces the cost. A best estimate of the average analyte concentration could be established in this way. Experimentation towards this objective in two sites of contaminated land would give practical solutions to arising problems. However what is usually wanted in environmental investigations is information on the spatial distribution and the extent of the contamination. The potential, but not the practice, of such an approach to this more complex objective has been examined in this research and relevant problems have been recognised. Also the relationships between the uncertainty estimates made from the different methods are examined and compared to the results of the plotted experimental variograms of the data.

In-situ measurements of soils at contaminated sites is a developing methodology in recent years (Keith, 1991). Such methodologies for screening the contamination have the potential of giving both a rapid assessment of the concentration of the contaminant and immediate information on its spatial distribution. In this research a portable X-ray fluorescence device was tested to assess the magnitude of uncertainty when such methodology is used. Comparison of the results with laboratory based methods would determine the fitness-for-purpose of the portable method measurements for environmental interpretation.

The consideration of the general applicability of the methodology in sampling other media is an other objective of this research. The general principles of sampling theory define this general applicability. However, the special characteristics of the different sampling media (e.g. temporal variation in the sampling of waters) have to be considered in each instance. A longer term objective of the project will be considered after reviewing the overall results of the experimentation. Theoretically the establishment of a 'standard field' or RST is possible. If such a field could be established, absolute sampling bias could be measured for new samplers or new protocols by comparison with certified measurements. This could be applied either to the average analyte concentration or to the spatial distribution of the contamination (e.g. contamination hot spot location).

1.3 Structure of Thesis

This thesis contains ten chapters. General information on the subject is covered in the first chapter and a literature review on sampling and uncertainty is presented in the second chapter. The sampling and analytical methodology followed throughout the experiments conducted in this project are presented in the third chapter. Chapters Four, Five, Six and Seven are related to the experiments at the first contaminated site in Derbyshire. The initial survey on this contaminated site is described in Chapter Four, the first collaborative trial in sampling (CTS) in Chapter Five and the first sampling proficiency test (SPT) in Chapter Six. Experimentation with in-situ measurement using a portable x-ray fluorescence analyser on the same site is described in Chapter Seven. In Chapter Eight the experiments on the second contaminated site, on Hounslow Heath, London are described. This comprises the initial survey at this site and the second CTS and SPT organised to validate the methodology. The general discussion of the suggested methodology for the estimation of sampling uncertainty is presented in Chapter Nine. Finally the conclusions of the research and some recommendations for further study are presented in Chapter Ten.

Six papers related to the work of this research have been published (or are in press). These papers are listed at the end of the thesis.

CHAPTER TWO. LITERATURE REVIEW ON SAMPLING AND UNCERTAINTY

2.1 Introduction

In this chapter a review of the literature on sampling and uncertainty is presented. After the description of some useful terms used in this thesis, some information from the bibliography on sampling and uncertainty is given. Analytical and sampling errors are then linked to measurement uncertainty as a whole. The existing methodologies for the estimation of analytical uncertainty are presented as a base for the development of methodologies for the estimation of measurement uncertainty due to sampling, which is the main problem studied in this research.

2.2 Terminology

It is particularly important to clearly define the terms used in the subject of sampling and uncertainty in order to explain the research with the least ambiguity. The terms described below have been classified according to their relation to i) sampling, ii) uncertainty and error and iii) practices for data quality.

i) Terms related to sampling:

Sample. A portion of material taken in the process of sampling, with the intent of being representative (Ramsey, 1994). In this research the term refers to field sampling rather than laboratory sampling.

Sampling target. A lot or consignment of a raw material or manufactured product, or a certain area or site from where the samples are to be collected (Thompson and Ramsey, 1995).

Sampling protocol. A prescribed fixed procedure for the collection of samples from a sampling target (Thompson and Ramsey, 1995).

Sample increment. A segment from the sampling target selected on a basis defined in the sampling protocol (Thompson and Ramsey, 1995).

Aggregate sample. The sample derived from the combination of the sample increments (Thompson and Ramsey, 1995).

Laboratory sample. The material that is to be analysed, derived from the aggregate sample after treatment in the laboratory (Thompson and Ramsey, 1995).

Test material. The prepared material in the laboratory ready for analysis (Ramsey, 1994).

Test portion. The portion of a test material that is prepared for an individual analysis (Ramsey, 1994).

ii) Terms related to uncertainty and error:

Uncertainty. A parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand (Eurachem, 1995).

Error. The difference between an individual result and the true value of the measurand (Eurachem, 1995). Error is a single value, whereas uncertainty takes the form of a range.

Trueness. The closeness of the agreement between the average value obtained from a large series of test results and an accepted reference value (ISO 3534-1:1993). Because trueness refers to an average value of a large series of results the random element is removed (Thompson and Ramsey, 1995).

Accuracy. The closeness of the agreement between a test result and the accepted reference value (ISO, 1993). Accuracy refers to a single measurement and thus includes both random and systematic elements.

Bias. The difference between the expectation of the test results and an accepted reference value (ISO 3534-1:1993). Bias refers to the mean or expectation of a large series of results which removes the random element (Thompson and Ramsey, 1995).

Precision. The closeness of agreement between independent test results obtained under prescribed conditions (ISO Guide 30, 1992).

iii) Terms related to practices for data quality in analysis

Repeatability conditions. Conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time (Thompson and Ramsey, 1995).

Reproducibility conditions. Conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment (Thompson and Ramsey, 1995).

Reference material. A material or substance one of whose property values are sufficiently homogeneous and well established to be used for the calibration of apparatus, the assessment of a measurement method, or for assigning values to materials (ISO, 1993).

Certified reference material. A reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure which establishes its traceability to an accurate realisation of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence (ISO, 1993).

Proficiency test. A method of checking laboratory testing performance by means of interlaboratory tests (Thompson and Wood, 1993).

Collaborative trial. An interlaboratory study in which each laboratory uses the defined method of analysis to analyse identical portions of homogeneous materials to assess the performance characteristics obtained for that method of analysis (Horwitz, 1988).

Fitness for purpose. The property of data introduced by a measurement process that enables a user of the data to make technically correct decisions for a stated purpose (Thompson and Ramsey, 1995).

2.3 Sampling Theory

2.3.1 Application of Sampling Theory to Sampling Surveys

The purpose of sampling theory is to make sampling more efficient (Cochran, 1977). Specifically, the theory aims to develop of methods of sampling selection and estimation, of specified precision at minimum cost. Cochran (1977) indicates two differences between standard sample survey theory and the classical sampling theory. In

sample surveys the information about the mathematical form of the frequency distribution is not assumed known (distribution-free approach) as opposed to the known mathematical form of frequency distribution in classical theory. A second difference is that the populations in survey work contain a finite number of units. Sampling from a finite instead of an infinite population makes the results slightly more complicated. However the application of statistical methods to sampling for chemical analysis is relatively common (Kratochvil et al., 1984). Generally, the traditional approach in sampling for surveys is the collection of a representative sample for the studied population. To achieve this, every effort is made to optimise the sampling protocol (e.g., number of samples taken, distance between sampling points, size of sample) so as to minimise measurement uncertainties. However, in practice it is difficult to follow the theoretically derived protocol in all instances. Alternatively, the development of methods for the estimation of the measurement uncertainties introduced by the sampling process has been suggested (Ramsey, 1994).

General statistical sampling theory has been applied in the area of geochemical and mineral evaluation by several workers. Most of their work has been aimed toward the solution of specific types of problems. Gy (1979) and Visman (1969), for example, have worked on commercial bulk applications of sampling theory.

Gy (1979) gives equations for the prediction of sampling variance in a sample of a given weight or the sample weight required for any desired uncertainty level. This is based on several factors related to the particle shape, particle size distribution the composition of the phases comprising the particulates and the degree to which the substance sought is liberated from the remainder material in the instance of sampling of particulate materials. The working equation of Gy's theory is:

$$s_s^2 = \rho d^3 \cdot x (100 - x) \cdot (1/w - 1/W) \cdot l \cdot f \cdot h$$

where: s_s^2 = sample variance

ρ = weighted mean density of all particles

d = side of cubic particle

x = weight percentage of element x in the material as a whole

w = weight of sample

W = the weight of the whole consignment or lot

l = a mineral liberation factor

f = a particle shape factor

h = a particle size distribution factor

Two basic assumptions are made for the application of this equation:

- i) The material is thoroughly mixed and there are no errors inherent in the sampling tool or crushing machines
- ii) Individual particles can be selected with equal probability.

Gy has also considered in a systematic way, all the potential sources of error in sampling of heterogeneous materials (Gy, 1992). It should be noted here that in this work, he refutes the possibility of measuring sampling bias and suggests that sampling correctness is the only effective way to avoid sampling bias.

Visman (1969) has developed a general theory of sampling that takes into account the effects of heterogeneity in both well-mixed and segregated populations. He proposed that the sampling variance s_s^2 could be related to individual increment size w and number of increments n by:

$$s_s^2 = A/wn + B/n$$

The constant A is called a homogeneity constant and is a function of the total weight of sample collected. The constant B is called a segregation constant. Values for A and B can be obtained experimentally for a bulk population.

Sampling theory applied to soil sampling has been considered by different workers for several purposes such as nutrient content and for soil mapping and classification (Webster, 1977). The application of sampling theory for the investigation of contaminated land gives useful information on the required sampling strategy and it is used for the optimisation of sampling design in respect to the number of samples, the sample size and the distance between the sampling points (Ferguson, 1992; Holcombe, 1988).

2.3.2 Sampling strategies

Sampling strategies vary according to the medium being sampled, the objectives of the sampling plan and the available analytical methodology. Specifically when sampling solids, for example soil, the sampling protocols differ widely.

There are two different approaches to field sampling; probability sampling which is based on statistical theory and judgmental sampling which is based on prior knowledge and information on the sampling target. Generally, in the class of probabilistic strategies there are two types of sampling patterns; systematic and random. In systematic sampling the sample increments are collected on a grid. Regular grids usually employ a rectangular grid, are easy to set up and interpolation between sampling points is potentially easy. The grid dimensions will differ according to a) the objective of sampling, e.g., estimation of average concentration of an analyte, or location of hot spots of contamination and b) the dimensions of the sampling target area. A disadvantage of this design is the possible regular periodicities in the population that might coincide with the period of the grid (Webster, 1977). The herringbone design (Ferguson, 1992) is an example of unaligned systematic sampling grid that has been shown by simulation to be the most effective for locating contamination hot spots. This is because it is very weakly influenced by target orientation. Triangular grids have the advantage of leaving a smaller unsampled area compared to rectangular grids.

Non-systematic patterns shaped like letters such as 'W' in which the samples are collected along the legs of the 'W' are commonly used by soil samplers for some purposes (MAFF, 1979). A limitation of these patterns is that heterogeneity of the sampling target might prove the sampling design not to be 'fit-for-purpose'.

In simple random sampling, the coordinates of the sampling points are taken from tables of random numbers. Random sampling is the most statistically appropriate design but it is time consuming to set up and the resulting uneven coverage makes this scheme inefficient for describing areas (Webster, 1977) or locating contamination hot spots (Ferguson, 1992). The best instance in which this design is effective is for totally exploratory sampling plans or irregular occurrences of contaminated zones (Keith, 1991). In the stratified random design the sampling target area is subdivided into a number of grid cells (strata). The sampling points are selected randomly in these smaller areas. As a result the sampling target is more evenly covered with gains in precision and efficiency (Webster, 1977). Also this design is less affected by regular periodicities than the regular grid.

The collection of composite samples, i.e. a sample made up by mixing a number of 'n' increments should improve the sampling precision. When the objective of sampling is

the estimation of the average concentration of an analyte in a sampling target, it is common to mix the collected increments to make a composite sample. This saves time and reduces the cost of analysis. However in some instances composition is inappropriate and can lead to wrong evaluation e.g., in the analysis of volatile organic compounds where a decrease in the content of volatiles can occur due to the process of mixing (Paetz and Crossmann, 1994).

The optimum number, spacing and frequency of samples can be estimated statistically to meet the minimum confidence levels in sample results. When large numbers of heterogeneous samples are involved, geostatistical and kriging approaches can be most helpful (Barcelona, 1988). Generally, the necessary number N of single samples is determined according to the equation (Hoffmann, 1994):

$$N = (ts / L)^2$$

where: t = the Student's factor,

s = the estimate of the standard deviation of the arithmetic mean value of all single values from preliminary examination of at least 20 single samples,

L = the tolerable uncertainty (stated in %) of the result.

In an other approach to sampling of contaminated land, the number of sample points required depends on the target size and shape for a fixed probability of hitting the target (Ferguson, 1992). Specifically the number of sampling points (N) needed to ensure 0.95 probability of hitting targets of various sizes with a herringbone sampling design is given by:

$$N = k A/a$$

where: A = the total site area,

a = the target area and

k = the target shape constant.

The formula assumes that the target area is a single, uniform contamination hot spot of a geometric shape. The assessor should choose a k value to reflect his belief in the likely target shape. The shape constant k takes values as follows:

Circular target $k = 1.08$

Plume shaped target $k = 1.25$

Elliptical target $k = 1.80$

For more than one hot spot in the site, a larger number of samples are predicted to be needed to ensure a 0.95 probability of detecting them all.

The size of the single sample (i.e. the required sample mass) depends on the expected concentration of the analyte, the analytical method, the required precision of the result and the state of distribution of the analyte (Hoffmann, 1994). In the instance of sampling solids the grain size and the degree of the homogeneity also affect the size of the sample. Considerations of manageable sample size for transportation, storage and analysis have also a significant role.

Often, preliminary sampling (screening) of a sampling target is desirable to help delineate the extent of the contamination and the variations in contaminant levels in the area under study (Keith, 1991). Screening facilitates the use of test methods defined as fast, inexpensive and portable methods for measuring the degree of contamination in the soil (Stock, 1995). Field portable equipment is very useful in this situation as it may be possible to collect, analyse and map data in-situ. The results from such rapid screening techniques (e.g. portable X-ray fluorescence) may not be as accurate as the laboratory analysis, but can potentially identify contamination hot spots rapidly and with minimal preparation (Moore, 1994). Such methods are discussed further in Chapter Seven of this thesis. However, the use of such methods depends on the specific requirements of the investigation and can be not useful in situations such as the determination of the identity of contamination. In these instances compound specific capabilities are required. Also rapid field methods are of little help for monitoring, where high precision and continual measurement are essential (Stock, 1995).

2.3.3 Environmental Sampling of Contaminated Land

The process of sampling in environmental geochemical investigations usually has the objective of the collection of a representative portion of the population under study. The planning of such sampling surveys depends on different factors related to the sampling media, the required data quality (i.e. measurement uncertainty), the available analytical methods and financial considerations.

The objectives of environmental sampling can be divided into two types: monitoring or research (exploratory) (Keith, 1991). Monitoring sampling plans provide a spatial or temporal summary of analyte concentrations and are used mostly for regulatory purposes. Research sampling plans have goals such as providing information on the speciation of the analyte under study, the environmental pathways, or the investigation of the relationships that control the levels and variability of pollution concentrations.

Environmental sampling can have a random, systematic or judgmental approach depending on the objectives set, the needed quality of the information and the complexity of the site (Keith, 1991). The various approaches have different implication for the number of samples and the magnitude of bias introduced. Judgmental sampling introduces the largest amount of bias but requires the smallest number of samples, while random sampling designs, theoretically introduce the smallest bias but require the largest number of samples.

The choice of sampling devices is related to the sampling medium and the analytes to be measured. They play a very important role because they can contribute large errors especially when the analytes of interest are at low concentrations. Disturbance of the environment is also possible when using an inappropriate sample collection method. The frequency of sampling and the number of samples needed depend on the time or spatial variability of the sampling site and the precision required (Keith, 1991).

The depth of sampling can be either to a specified metric interval or to a specified soil-horizon related (Paetz and Crossmann, 1994). In the instance of metric depth sampling soil may be sampled at a few fixed depths. This system is used for agricultural purposes and screening analysis of potentially contaminated land. It is also appropriate in cases where the soil profile is not very differentiated. Sampling of soil horizons is sometimes suggested for investigations concerning soil and groundwater protection.

International standardisation of terminology and investigation methodology is very important for the improvement of the international communication and collaboration. For environmental sampling and especially for soil sampling the International Organisation for Standardisation (ISO) established a Technical Committee for Soil Quality (ISO TC 190) in 1985. The scope of the Committee comprises the investigation of soil in respect to the design of sampling programmes, guidelines on soil sampling techniques and sampling strategies (Bosman, 1994).

2.3.4 Geostatistical Approach for Sampling and Estimation

Geostatistical techniques allow for use of the spatial information of the sampling data (Isaaks and Srivastava, 1989). Spatial continuity is an essential characteristic of many natural phenomena and is not accounted for in classical statistical sampling theory. Geostatistics have been extensively used in mining for ore evaluation (Journel and Huijbregts, 1978). However, geostatistical techniques can be applied for the description and estimation of every spatially related variable. There is a clear analogy between the objectives of ore grade evaluation and that of estimating the mean concentration of a contaminant at a site. Also between the cut-off level in mining and the regulatory limits in contaminated land investigations. Geostatistical approaches to the design of field sampling of contaminated sites have been suggested (Flatman et al., 1988).

Geostatistical methods for estimation involve weighted linear combinations of the data values (Isaaks and Srivastava, 1989):

$$estimate = \hat{v} = \sum_{i=1}^n w_i \cdot v_i$$

where $v_1, \dots, v_n$ are the n available data values and w_i is a weight assigned to the value v_i . The assignment of the weights to the data values is usually based on ^{geo-}statistical theory.

There is a distinction between global estimation related to the characteristics of the distribution of the data values over the whole area of interest (e.g. estimation of ore grades in a mineral deposit or estimation of pollutant concentrations in a waste site) and local estimation which gives detailed local information. The latter is often more useful, especially in the instance of the study of contamination, as the main concern is to find the areas with unacceptable high concentrations of contaminants. However global estimation is used in the early stages of studies as it gives a good general picture of the situation.

The consideration of geostatistical techniques in field sampling studies becomes vital when the objective is the delineation of areas with particular characteristics (e.g. contamination hot spots). Concentrations estimated from field samples are usually spatially correlated to some extent, which means that samples taken close together are more similar in value than samples taken farther apart (Flatman et al., 1988). Subsequently, in addition to classical statistics for the description of the population under

study, spatial variables must be described by a measure of the correlation between neighbouring locations. Variography and kriging are two classic methods for describing the spatial distribution of one or more attributes. Variograms provide a measure of spatial variability of the data and kriging gives an average estimation of the property under study. Although in this project the geostatistical approach has not been considered in detail, some information on these techniques is given in Chapter Nine.

2.4 Estimation of Sampling Errors

Field sampling is an integral part of the measurement process that is often neglected in terms of the errors and uncertainty that it generates. Systematic error in sampling, expressed as sampling bias, affects the results of every environmental survey to some extent. Specifically, undetected sampling bias within measurements for the classification of contaminated land can have implications that are both financial and medical due to the potential toxicity of the soils to human health. For example, if a particular site is examined for Pb contamination in the soil and the analytical result of a representative sample gives $450 \mu\text{g g}^{-1}$ of Pb, which is below a guideline limit of $500 \mu\text{g g}^{-1}$, then the site may be judged not to be contaminated. However the presence of an undetected systematic error of 14% for example, would lead to the wrong decision. The presence of sampling bias can also effectively invalidate the interpretation of chemical analyses no matter how high the analytical quality.

There are four main types of measurement error in environmental geochemistry. Analytical and sampling precision describe the random measurement errors and are often expressed as variances, σ^2 estimated by s^2 (the square of the standard deviation, σ estimated by s). The analytical and sampling bias however, quantify the systematic errors from these two sources (sampling and analysis). Methods have now been developed for the estimation of the first three types of error (Ramsey, 1994). Such methods utilise the collection of duplicate samples with duplicated chemical analyses for the estimation of sampling and analytical precision (Garrett, 1969; Ramsey, 1993) and use certified reference materials for the estimation of analytical bias (Ramsey et al., 1987). Estimates of sampling precision are crucially dependant on the distance chosen between the original and the duplicate sample. As a general rule the distance between the initial sampling point

and the duplicate sample point represents the uncertainty in the position of the location of the sampling point in respect of its neighbours and is typically equal to 15% of the distance between the sampling points (Ramsey, 1994). No method has yet been described for the estimation of sampling bias in general or for this particular application of environmental sampling.

There are many texts that give information on general sampling practice in different environmental media (Keith, 1988) but the emphasis up to now has been on how to improve sampling technique rather than on estimating the errors and uncertainty introduced. Gy has suggested that sampling bias can be avoided by ensuring correctness in the sampling procedure and, in effect, he denies the possibility of estimating systematic error in sampling (Gy, 1992). However the idea of devising methods for the estimation of sampling uncertainty that are analogous to the methods already in use for the estimation of analytical uncertainty has recently been suggested (Ramsey, 1994; Thompson and Ramsey, 1995).

2.5 Estimation of Uncertainty from Analytical Measurements

Uncertainties originating from chemical analysis are linked to sampling uncertainties in the consideration of measurement uncertainty as a whole (Ramsey, 1994). Methods for the estimation of analytical uncertainty are generally well established and widely applicable.

Generally there are two approaches to the problem (AMC, 1995). The 'bottom-up' estimates of uncertainty, requires the dissection of the analytical system, the estimation of the uncertainties in all of the separate parts of the system and the combination of all these individual uncertainties to form the 'combined standard uncertainty'. The alternative approach is the 'top-down' view of uncertainty estimation, where the laboratory is viewed as a member of a population of laboratories. In this way systematic and random errors occurring within single laboratories effectively become random errors in this context and can be easily quantified as variances. Information from inter-laboratory trials is extremely useful in this latter instance.

2.5.1 Classical Evaluation of Analytical Uncertainty Using the 'Bottom-up'

Method

The procedures for evaluating uncertainty in analytical measurements are well documented in 'Quantifying Uncertainty in Analytical Measurement' produced by EURACHEM in 1995. The different steps start with the specification of what is being measured and the analytical methodology being employed. The next step is the identification of possible sources of uncertainties for each parameter in the analytical system. Sampling, instrument bias, reagent purity, measurement conditions, sample and computational effects, contamination, operator bias and random effects are mentioned in the EURACHEM's report as potential error-generating processes.

The quantification of uncertainty from the different sources is made either through experimental determination (e.g. the effect of temperature can be investigated by making measurements at different temperatures), or by using data from other sources, i.e. information from reference materials and previous results obtained from inter-laboratory studies. In some instances the estimation of uncertainty is based on judgement of the analyst. This subjective probability expresses the 'degree of belief' and is influenced by common sense and expert knowledge gained by earlier experiments and observations.

The combined standard uncertainty (u_c) on the measurement result is obtained by combining the individual components, taking the square root of the 'sum of the squares' of the individual components. It is recommended that the final analytical result is stated together with the expanded uncertainty (U) which is calculated by multiplying the standard uncertainty by a coverage factor k (e.g. $k = 2$). The use of this coverage factor gives a specified level of confidence (e.g. approximately 95% for $k = 2$).

The shortcomings of the described methodology are the long calculations required and the need to recognise all the sources of uncertainty in the different steps which might be a difficult task. Also because of the summation of the squares for the calculation of the combined uncertainty, the latter is often dominated by the largest term. The correction of all systematic errors is also a requirement that may be difficult to achieve. Internal quality control is essential for the application of this methodology in order to continually justify the applicability in the analytical system (AMC ..., 1995).

2.5.2 Evaluation of Analytical Uncertainty Using Inter-laboratory Trials, the 'Top-down' Method

The philosophy in this approach for the estimation of uncertainty is that errors, both systematic and random, occurring within single laboratories become random errors between laboratories when seen from a higher level and can be simply quantified (Analytical Methods Committee, 1995). Subsequently, an effective way to monitor and reduce uncertainty is to replicate at the highest level in the top-down analysis. Inter-laboratory trials can provide the required information to an end-user of the data by supplying data from different samples and laboratories. Moreover the estimation of measurement uncertainty not only from analysis, but also from sampling should be possible, by applying analogous methods in the area of sampling quality.

There are two types of inter-laboratory trials that are applied in the area of analytical data quality that could be useful for the estimation of measurement uncertainty. These are the collaborative trial and the proficiency test (Analytical Methods Committee, 1995). They have different purposes and require different methods for interpreting the results.

The collaborative trial is traditionally applied at the final stage in determining the performance characteristics of an analytical method. The analytical method in question would be extensively tested in a single laboratory beforehand to ensure that conducting a trial, which is an expensive operation, is worthwhile. The specific purpose of the trial is to estimate the within-laboratory (repeatability) precision, the between-laboratory (reproducibility) precision and, if possible, to verify that the bias of the method is insignificant. The trial is conducted by the distribution of a number of samples (>5) of the same test materials to a minimum of eight competent laboratories, each of which analyses the samples in blind duplicate according to the supplied protocol.

The statistical interpretation is made using one-way analysis of variance (ANOVA) to form the estimates of the within-group standard deviation ($s_r = s_0$), the between-group standard deviation ($s_L = s_I$), and the reproducibility standard deviation ($s_R = \sqrt{(s_r^2 + s_L^2)}$). The standard deviation s_r represents repeatability conditions under which independent test results are obtained with the same method on identical test material in the same laboratory by the same operator using same equipment within short intervals of time

(Thompson and Ramsey, 1995). The standard deviation s_R represents reproducibility conditions of precision, under which independent results of measurement of the same measurand are obtained with the same measurement procedure in different laboratories with different observers using different equipment (Thompson and Ramsey, 1995).

The treatment of outliers in such trials has been discussed by several workers. In general although the rejection of outlying results by applying statistical tests is suggested in the protocol of analytical collaborative trials (Horwitz, 1988) the alternative of robust ANOVA (Analytical Methods Committee, 1989) can be used on the results of a collaborative trial (Thompson and Ramsey, 1995). Robust ANOVA provides an alternative to outlier rejection by retaining outlying observations, giving them less weight. Such an approach has been also applied successfully for the estimation of sampling precision in environmental studies (Thompson and Maguire, 1993; Ramsey, 1993).

An uncertainty for the method bias can be obtained if an assigned value X_a for the analyte concentration is available. This uncertainty is given by:

$$u_b^2 = \text{var}(x) + u_a^2$$

where x is the grand mean of the results and u_a is the uncertainty of the assigned value. The value of $\text{var}(x)$ is a function of the repeatability and reproducibility variances of the participating laboratories.

A proficiency test is a method of testing the bias of laboratories by the regular analysis of typical materials with unknown concentration of analyte that are distributed by the organiser (Thompson and Wood, 1993). The ethos of such tests is that the laboratories use their own methods under typical conditions (i.e., no special effort) so that the outcome is representative of normal operations. The participants report the concentration estimates to the organiser by a fixed date, and in exactly the same form as they would to a normal client. The organiser then converts each result to a score that characterises the performance of the laboratory and reports back the scores of all the participants. Hence each participant laboratory can compare its performance with that of its peers, with an externally determined standard of performance, and with its own performance in the past. The purpose of the proficiency test is to alert the participant to unexpected sources of error, so that remedial actions can be instigated of which the efficacy can be subsequently tested.

Anonymity of the participants is usually ensured in proficiency tests. When a reference material with an assigned value of X_a is available the standard uncertainty of the method bias u_b can be estimated by:

$$u_b = \sqrt{\text{var}(x) + u_a^2}$$

where $\text{var}(x)$ is a function of the scores achieved by the different laboratories and u_a the uncertainty on the assigned value of the reference material.

CHAPTER THREE. METHODOLOGY

3.1 Introduction

The general sampling and analytical methodology followed in all the experiments of this research is presented in this chapter. The sampling designs used for the initial surveys at the two contaminated sites used for experimentation, as well as the sampling protocols for the inter-organisational trials, are described with reference to the relevant chapters. The sample treatment and preparation is also described and the analytical procedure is presented.

3.2 Sampling Methodology

3.2.1 Collection of Soil Samples in Initial Surveys

Two contaminated sites were used for experimentation in this research. The first one is an old smelting site, situated in Derbyshire in the English Midlands (described in detail in Chapter Four, section 4.2). The second one is a part of a completed landfill site on Hounslow Heath, West London (described in Chapter Eight, section 8.2.1). These, were selected as sampling targets after the consideration of several criteria listed below:

1. Substantial concentrations of heavy metals in the soil,
2. Stability of contamination, i.e. absence of factors that might cause short term variation of the metal concentration,
3. Dimensions of sampling area typical of contaminated land investigations,
4. Geographical location to ensure ease of access for sampling.

The surveying technology for the initial surveys utilised metric tapes for measuring the distances between the sampling locations and canes for marking the nodes of the sampling grids. A compass was used for the orientation of the grids and for the determination of reference points, e.g. the origin of each sampling grid. The topography of the sampling targets affected the accuracy in locating the sampling points

in the field. Especially in the Derbyshire site the presence of a slope strongly affected the lay out of the sampling grids. To allow for the location errors an appropriate distance for collecting duplicate samples was selected as described later in this section.

Top-soil samples (0-15 cm) were collected using a hand auger with a 2.5 cm diameter stainless steel screw. The samples were collected after removing the turf of vegetation and were either taken using single auger hole or by taking five fold composites within an area of 1m² as shown in Fig. 3.1. The auger was carefully cleaned after the collection of each sample and the auger holes were covered over, to preserve the original state of the field after sampling. The samples were placed in wet strength paper bags for transportation to the laboratory and for oven drying.

Various sampling protocols were used in the initial surveys of this research at the two contaminated sites, used as sampling targets (Fig. 3.1). They were selected on the basis of representativity and common usage. These protocols were:

- a) a simple regular grid design taking a single auger sample every 20 m (described by Webster, 1977).
- b) a composite regular grid being the same as a) but taking five fold composite auger samples in a shape of "X" over an 1 m² area.
- c) a stratified random design (Webster, 1977).
- d) a herringbone design with an alternate 5m offset from the regular square grid (Ferguson, 1992).

Duplicate samples were collected in all instances for the estimation of sampling precision (Ramsey, 1993). The distance between the duplicate samples was selected to represent the spatial uncertainty in the location of the sampling point and subsequently it depends on surveying technology employed, the total area of the sampling target and also the spacing between the sampling points (Ramsey, 1994). In this research duplicate samples were collected 2 m away from the defined sampling point in all surveys. This represents an estimate of the accuracy with which the sites were located using measuring tapes on rough ground.

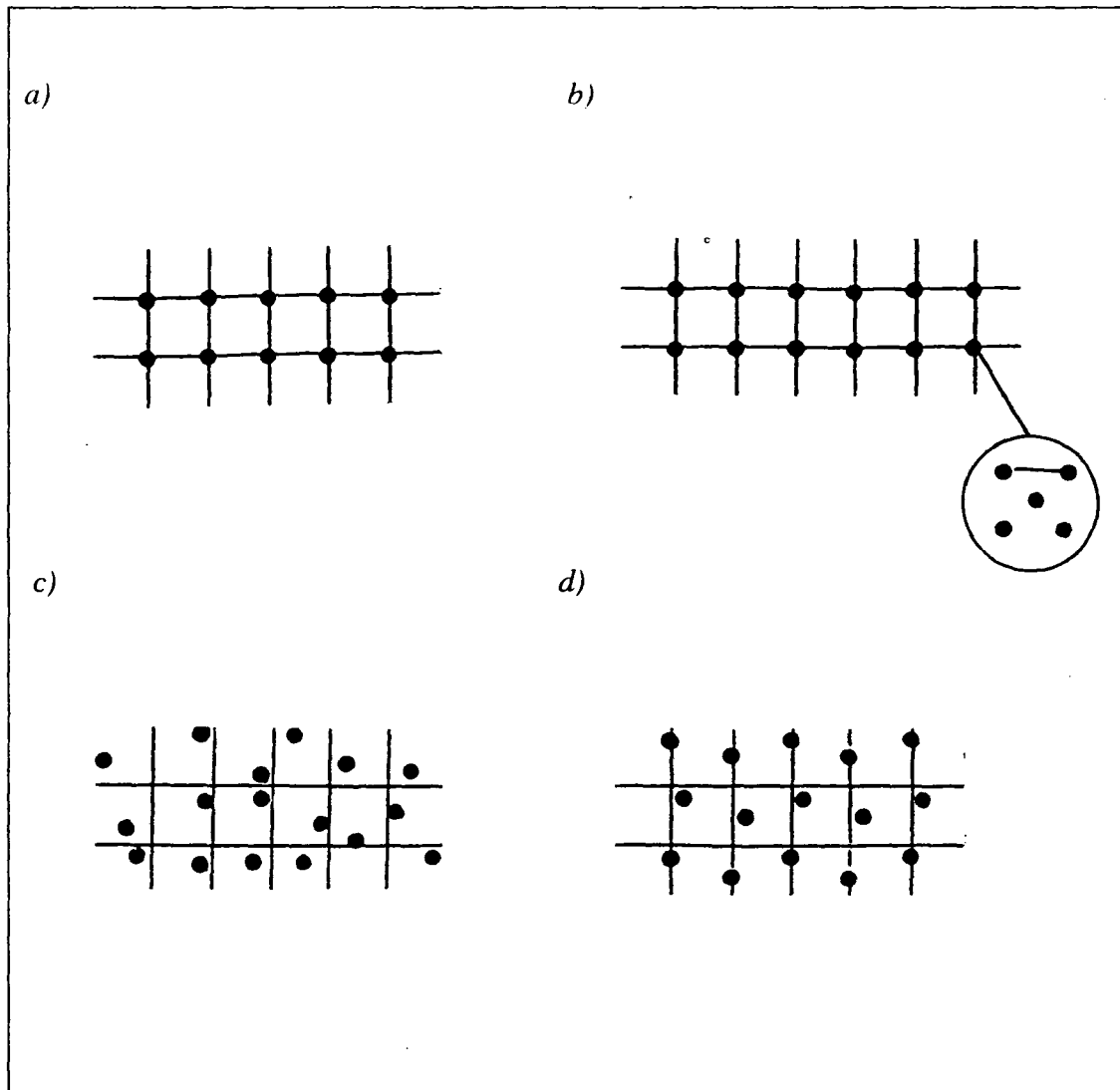


Figure 3.1. Schematic sampling designs used in the initial surveys of the research. (a) regular grid, (b) composite regular grid, (c) stratified random, (d) herringbone design.

3.2.2 Collection of Samples in the Inter-organisational Trials

Two inter-organisational sampling trials were organised for the purposes of this research at the two experimental contaminated sites. Each one of them comprised a collaborative trial in sampling (CTS) and a sampling proficiency test (SPT). The two CTS are described in detail in Chapters Five and Eight and the two SPTs in Chapters Six and Eight. It is sufficient to say here that in such sampling trials, different samplers (the participants of the trials) are collecting the samples either according to a common

sampling protocol provided by the organisers (CTS), or according to different protocols of their own choice (SPT). In these particular experimental trials organised for the purposes of this research the participants were asked to collect composite top soil samples from the sampling targets for the estimation of the average concentration of metals in soil. The sampling procedures are described in the following sections.

3.2.2.1 The Collaborative Trial in Sampling (CTS)

The collaborative trial in sampling required that the samples are collected according to a fixed sampling protocol. Nine samplers separately visited the sites (sampling targets) and used the protocol supplied, twice in a random fashion, i.e. the orientation and starting point of the sampling were selected using a table with random numbers, generated with the MINITAB statistical software. The eighteen samples so produced from each trial were analysed by the organisers in a single run. The results were calculated by nested ANOVA, to give the within and between sampler precision.

The sampling protocols used in the two collaborative trials organised for this project were different (Fig. 3.2). A 'W' shaped design was selected for the site old smelting site in Derbyshire, taking the sampling increments at the vertices of the 'W' and at three roughly equidistant points along each leg, and a regular grid with 20 m intervals for the completed landfill site in Hounslow Heath, London.

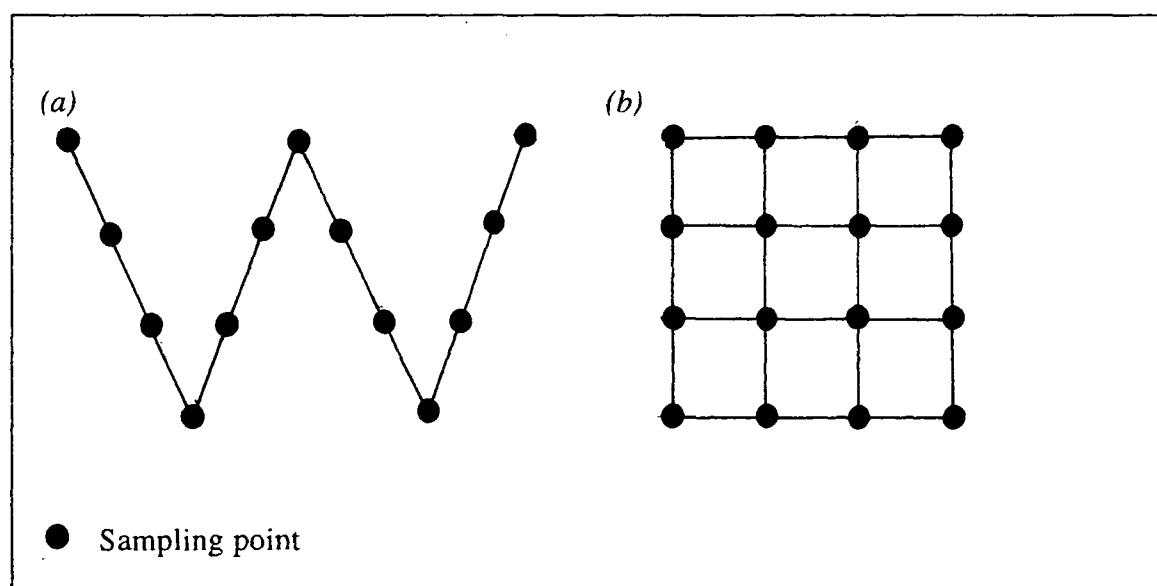


Figure 3.2. Schematic presentation of the sampling designs for the two CTS. (a) 'W' design for the first CTS, (b) regular grid for the second CTS.

These protocols were selected on the basis of representability and common usage as well as time constraint considerations. Specifically, it was judged that the time for each participant to spend on the collection of the sample increments on the CTS should not exceed two hours. The participants were asked to provide a composite sample for each of the trials for the specified area of the field. This compositing saves time for sampling and is sufficient for giving an estimate of the mean metal concentration in soil over the site. The sampling protocols given to the participants can be summarised as follows:

Sampling instructions

- 1. Sample the designated sampling area using the provided sampling design. Select the corner to commence sampling from the random number table provided.*
- 2. Use the metric tape and the canes to mark the sampling points in the field, if required.*
- 3. Use the auger provided to take the increments of 0-15 cm top soil. To do this first remove the surface vegetation from the sampling point and then rotate the auger to drill into the soil. Pull out the auger, trying to avoid smearing, and collect all the soil on the auger.*
- 4. Put the increment into the bag provided and move to the next sampling point. Make sure that you cover over the auger hole at each sampling point, leaving no sign of activity to preserve the statistical independence of the samplers.*
- 5. Follow the same procedure at each sampling point and put all the increments in the same bag, making one composite sample.*
- 6. Take a second sample, repeating the whole procedure. Select a different starting corner for sampling from the random number table. Use a second bag to collect this sample.*
- 7. Give the two sampling bags (marked) to the organisers.*

The equipment provided for the samplers was as follows:

- 1. One copy of the map showing the outline of the sampling area.*
- 2. One metric tape to measure the distances between the sample points in the field.*

3. *Canes.*
4. *Two sample bags.*
5. *One 2.5 cm diameter auger.*
6. *One marker pen.*

The instructions given to the samplers were designed so that would be easy to explain but also not totally unambiguous. For example the optional use of the metric tape and the canes allowed the participants to interpret the sampling protocol according to their own judgement. This way the CTS gave a realistic picture of the usual practice of samplers of contaminated land in interpreting a sampling protocol. Differences in the interpretation of the sampling protocol were judged to be important in this instance, because they contribute to the uncertainty of the measurement. This latter point was one of the objectives of the CTS in this research (see Chapter Five, section 5.1).

3.2.2.2 The Sampling Proficiency Test

In the sampling proficiency test (SPT) the participants visited the sampling targets and took one composite top soil sample by their own choice of sampling protocol. The samplers visited each one of the sites on separate days in order to preserve the statistical independence between them. Half ^a working day was allowed for each one of them to survey and sample the site. The collection of the composite sample for the SPT was done before they received the common sampling protocol for the CTS to avoid any bias in their decision on how to sample the site. No particular depth of sampling was defined in order to be able to see how the term 'top soil' is interpreted by soil samplers.

A variety of sampling protocols were used comprising regular grids, parallel traverses, triangular grids, 'W' shaped designs, herringbone and stratified random designs (Fig. 3.3). The sampling equipment comprised different types of augers and trowels and the depth of sampling varied from 0-10 to 0-15 cm.

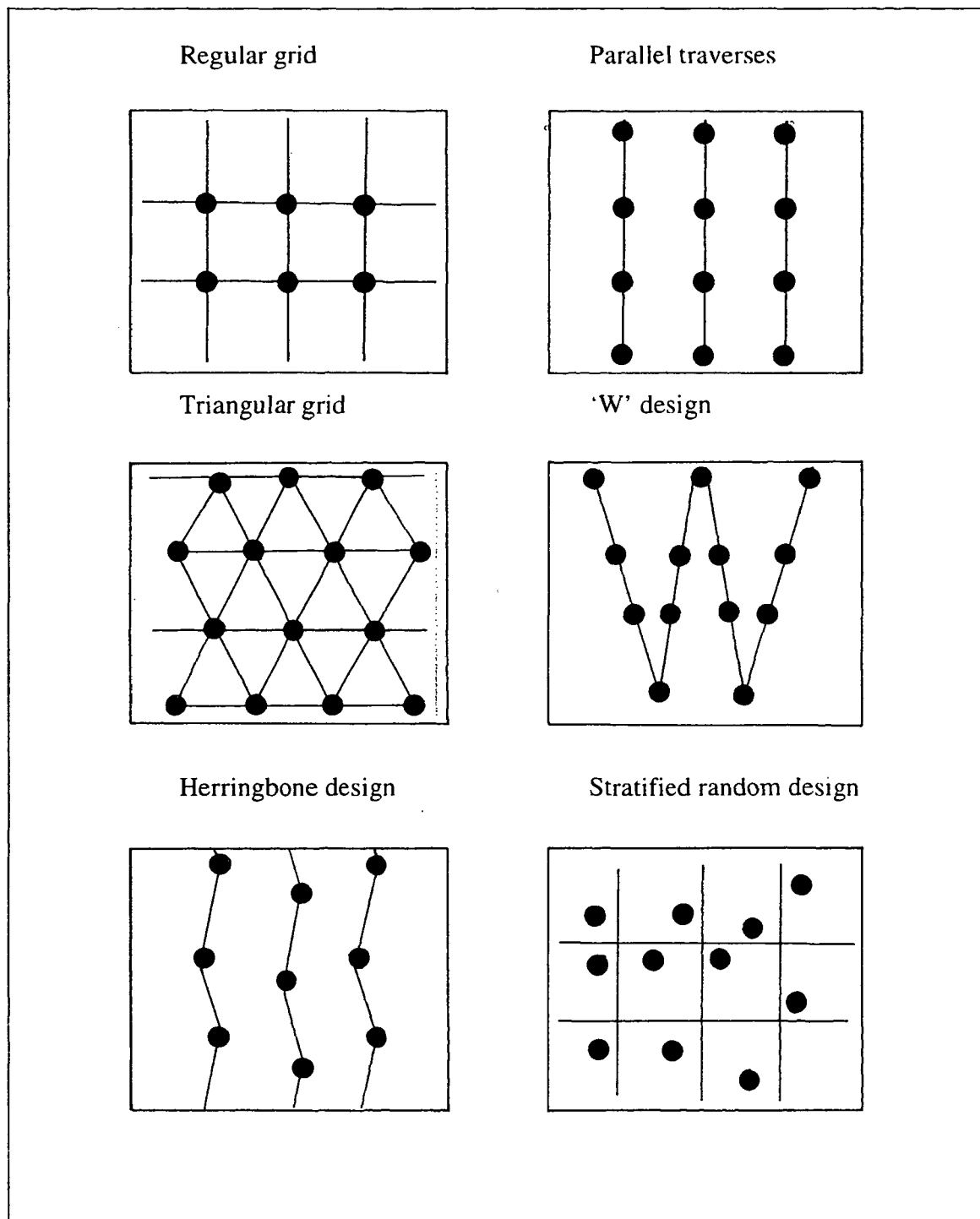


Figure 3.3. Schematic sampling protocols used by the participants in the two SPTs organised for this project.

3.2.3 Sample Preparation

All soil samples were carried to the laboratory in wet strength paper bags and dried in a drying oven overnight at 65 °C. The dried samples were disaggregated using a porcelain pestle and mortar so as to pass through a 2 mm (10 mesh) nylon sieve. A 50ml portion of the <2 mm soil fraction of each sample was mechanically ground in a terna mill to <100 µm and placed in plastic bags properly labelled. From this, test portions of 0.100 ± 0.001 g were weighed using a top-pan balance and placed in test tubes ready for acid dilution.

In some instances subsampling was required for the reduction of large composite samples. Coning and quartering was employed in order to take a representative sample according to the following process:

The disaggregated sample after sieving to pass through the 2 mm nylon sieve was mixed and heaped in a form of a cone. When the cone was formed the top was flattened and a cross was marked on top. Opposite quarters were selected and the two remaining quarters were rejected. When further reduction was needed the whole process was repeated after mixing again the selected two quarters.

3.3 Analytical Methodology

3.3.1 Preparation of Analytical Equipment

All analytical reagents used during this research were of 'AnalaR' quality and all equipment were kept clean. Deionised water was used routinely throughout the laboratory work for all washing and preparation of equipment, preparation of solutions and any other general purposes.

Test tubes were initially washed in water using a detergent (Decon) and left overnight in a 2% Decon solution. This was followed by rinsing with water and soaking overnight in nitric acid (10% v.v.). The glassware was finally rinsed three times in deionised water and dried in a filtered air cabinet at 40 °C before use.

3.3.2 Multi-element Analysis of Soil Samples

A nitric-perchloric acid digestion (Thompson and Walsh, 1989) was used on 0.100 g of finely milled soil in Pyrex test tubes (180 mm x 18 mm). The detailed procedure is described in Appendix 1. The solutions were presented for analysis to an Inductively Coupled Plasma Atomic Emission Spectrometer (ARL 3580). The calibration of the ICP-AES was based on commercial solutions of known concentration as described in Thompson and Walsh (1989) p. 146-147. This instrument measures the concentration of a suit of 25 elements with low detection limits and good precision. For Pb for example the detection limit is $6 \mu\text{g g}^{-1}$ and the instrumental precision is about 6% for concentrations well above the detection limit. The operation principles of ICP-AES are described by Thompson and Walsh, 1989. The main objective of this research required that the uncertainties introduced into the measurements by the ICP-AES and the sample preparation were quantified and controlled.

3.3.3 Estimation of Analytical Uncertainty and Quality Control

Internal, laboratory quality control was provided in every analytical run by the inclusion in the run of blank solutions, reference materials and analytical duplicates (Ramsey et al., 1987).

Blank solutions were treated in the same manner as the rest of the samples in order to detect bias caused by any external contamination in the laboratory. Reference materials were also used for the estimation of analytical bias and comprised a variety of certified reference materials (CRMs) (Table 3.1). These were selected to match the samples in respect to the nature of the samples, i.e. the type of soil and also the presence of the analytes over the range of their concentration in the samples.

Table 3.1. List of certified reference materials (CRMs) used for the estimation of analytical bias for all ICP-AES analyses in this research.

CRM	Type of Material	Concentration characteristics	Certified values and uncertainties ($\mu\text{g g}^{-1}$)
BCR141	calcareous loam soil	Low element concentrations	Cd 0.36 ± 0.10 ; Cu 32.6 ± 1.4 ; Hg $56.8 \times 10^{-3} \pm 4.3 \times 10^{-3}$; Pb 29.4 ± 2.6 ; Zn 81.3 ± 3.7
BCR142	light sandy soil	Low element concentrations	Cd 0.25 ± 0.09 ; Cu 27.5 ± 0.6 ; Hg $104 \times 10^{-3} \pm 12.3 \times 10^{-3}$; Ni 29.2 ± 2.5 ; Pb 37.8 ± 1.9 ; Zn 92.4 ± 4.4
BCR143	sewage sludge amended soil	Moderately elevated element concentrations	Cd 31.1 ± 1.2 ; Cu 236.5 ± 8.2 ; Hg 3.92 ± 0.23 ; Ni 99.5 ± 5.5 ; Pb 1333 ± 39 ; Zn 1272 ± 30
NBS2709	San Joaquin agricultural soil	Baseline trace element concentrations	Cd 0.38 ± 0.01 ; Cu 34.6 ± 0.7 ; Co 13.4 ± 0.7 ; Pb 18.9 ± 0.5 ; Zn 106 ± 3
NBS2710	Montana soil	Highly elevated trace element concentrations	Cd 21.8 ± 0.2 ; Cu 2950 ± 130 ; Pb 5532 ± 80 ; Zn 6952 ± 91
NBS2711	Montana agricultural soil	Moderately elevated trace element concentrations	Cd 41.7 ± 0.25 ; Cu 114 ± 2 ; Pb 1162 ± 31 ; Zn 350.4 ± 4.8

Some house reference materials (HRMs) prepared in the laboratory (Table 3.2) were also analysed so as to cover the full range of expected analyte concentrations. Analytical duplicates provided the method for the estimation of analytical precision as part of an integrated sampling and analytical quality scheme, utilising ANOVA.

Table 3.2 List of house reference materials (HRMs) used for the estimation of analytical bias for all ICP-AES analyses in this research.

HRM	Type of material	Concentration characteristics
HRM1	Agricultural soil from East Anglia	Low concentration of elements
HRM2	Mixture of soil and stream sediments	High concentration of elements
HRM31	Agricultural soil from Derbyshire	High concentration of heavy metals

All the quality control materials were analysed anonymously, in random positions within every analytical batch of samples. In this way the possible instrumental drift would have minimal effect on the analytical control measurements.

CHAPTER FOUR. ESTIMATION OF SAMPLING UNCERTAINTY AT AN OLD SMELTING CONTAMINATED SITE IN THE ENGLISH MIDLANDS: INITIAL SURVEY

4.1 Introduction

The feasibility of measuring sampling uncertainty, can be assessed initially by considering the simplest case which is the estimation of sampling uncertainty between different sampling protocols applied by a single sampler on a single area of contaminated land. The method can be also viewed as an application of the 'bottom-up' approach for the estimation of uncertainty in field sampling. This approach uses the sum of the random errors from all the component operations which in the instance of geochemical investigations comprise errors due to sampling and analysis. The findings are then used to assess the experimental design that would be required for the estimation of sampling uncertainty between different organisations. The ultimate estimation of absolute sampling bias using such an area as a Reference Sampling Target (RST), analogous to the use of a Reference Material (RM) for the estimation of analytical bias, is also considered.

4.2 Site Description

A pilot study was conducted at one site of contaminated land about 1 km east of Wirksworth (Grid Reference SK 298542), Derbyshire in the English Midlands (Fig. 4.1), originally described by Maskall and Thornton (1993).

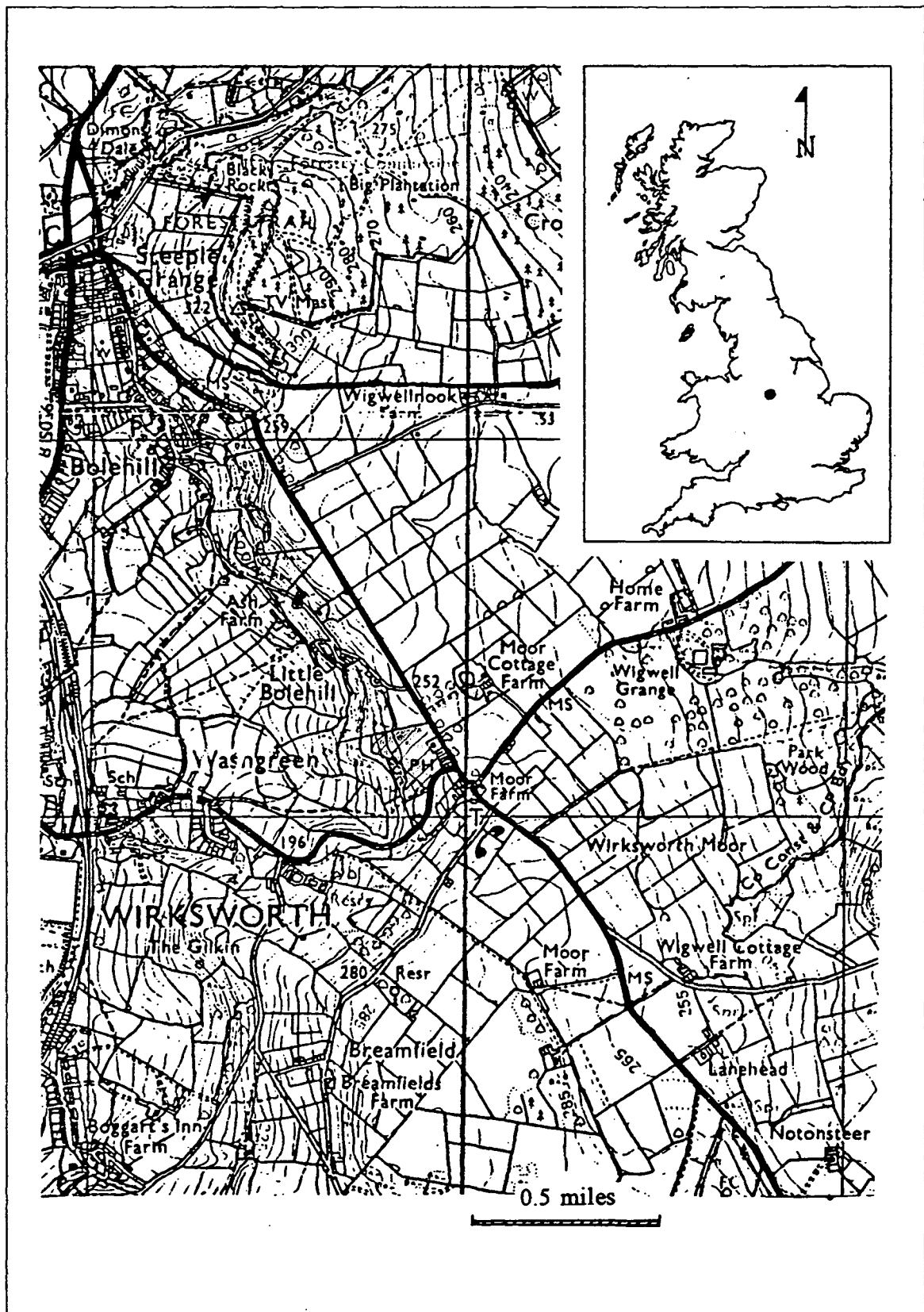


Figure 4.1. Geographical location of the study area in Derbyshire.

The sampling area has dimensions of 180x170 m (1.8 hectares) at an altitude of 230-250 m. The major sampling area is situated on a scarp slope which drops away to the south and southwest. The eastern part of the field is flat and lies along a ridge at an altitude of 250 m. The local geology is predominantly Millstone Grit with some associated siltstones and clays. Soils are classified as generally being loamy sands with little organic matter developed on Millstone Grit. The heavy metal contamination is due to the use of the site for Pb smelting in medieval times (1300-1550 AD). The physical location of the site is typical of a medieval bole smelter site, fuelled by wood. Such smelting sites were often located on high scarps. Boles relied on the prevailing wind to create the necessary air blast (Maskall and Thornton, 1993). In this particular field however, no evidence of previous smelting activities is visible.

The field has remained unploughed since 1948 and has been used as pasture for grazing horses in recent years. Evidence of man's activities comprise a septic tank on the flat area to the east, abandoned agricultural machinery at the bottom of the slope to the west, an old shed by the fence on the north side and two manholes over the road drainage pipe that crosses the field parallel to the north side fence. These might have caused disturbance in the original soil distribution and also affect the distribution of the contaminants in the ground.

This site was considered appropriate for use as an example of contaminated land in this project when judged by several criteria, i.e.

- i) substantial concentrations of heavy metals at the site,
- ii) the absence of factors that might cause short term variation of the metal concentration (e.g. flooding),
- iii) no recent or current cultivation or remediation,
- iv) dimensions that are typical of contaminated land investigations,
- v) the co-operation of the owner.

4.3 One Sampler Using Many Sampling Protocols

Four different sampling protocols were applied to sample the topsoil (0-15 cm) (Fig. 4.2). They were selected on the basis of representativeness and common usage. These protocols were:

- a) a simple regular grid design taking a single auger sample every 20 m, as generally described by Webster (1977)
- b) a composite regular grid being the same as a) but taking five fold composite auger samples in a shape of 'X' over an 1m^2 area,
- c) a herringbone design with an alternate 5m offset the regular square grid (Ferguson, 1992) and
- d) a stratified random design (Webster, 1977). The coordinates of the stratified random sampling within each 20 m square stratum were taken from a uniform random number distribution generated within a statistical spread sheet (MINITAB).

Designs (a), (c) and (d) all used single auger samples at each sampling point. The sampling points on the regular grid were located using a measuring tape. The regular grid was marked with canes and all the other designs were based on these canes, finding the exact sampling locations by using the measuring tape. In all the designs the sampling precision was estimated by taking sampling duplicates at 2 m distance away from the initial sampling point in random direction, in order to make a realistic estimate of the sampling repeatability variance.

Test portions of all soil samples (0.100 g) were prepared for chemical analysis by ICP-AES as described in Chapter Three for a suite of twenty-four elements (Thompson and Walsh, 1989). The raw analytical data are presented in Appendix 2. Analytical duplicates were prepared for each sampling duplicate, for the estimation of sampling and analytical precision (Ramsey, 1993). Reagent blanks and six certified soil reference materials were analysed at random positions between the sample test portions for the estimation of analytical bias (Ramsey et al., 1987).

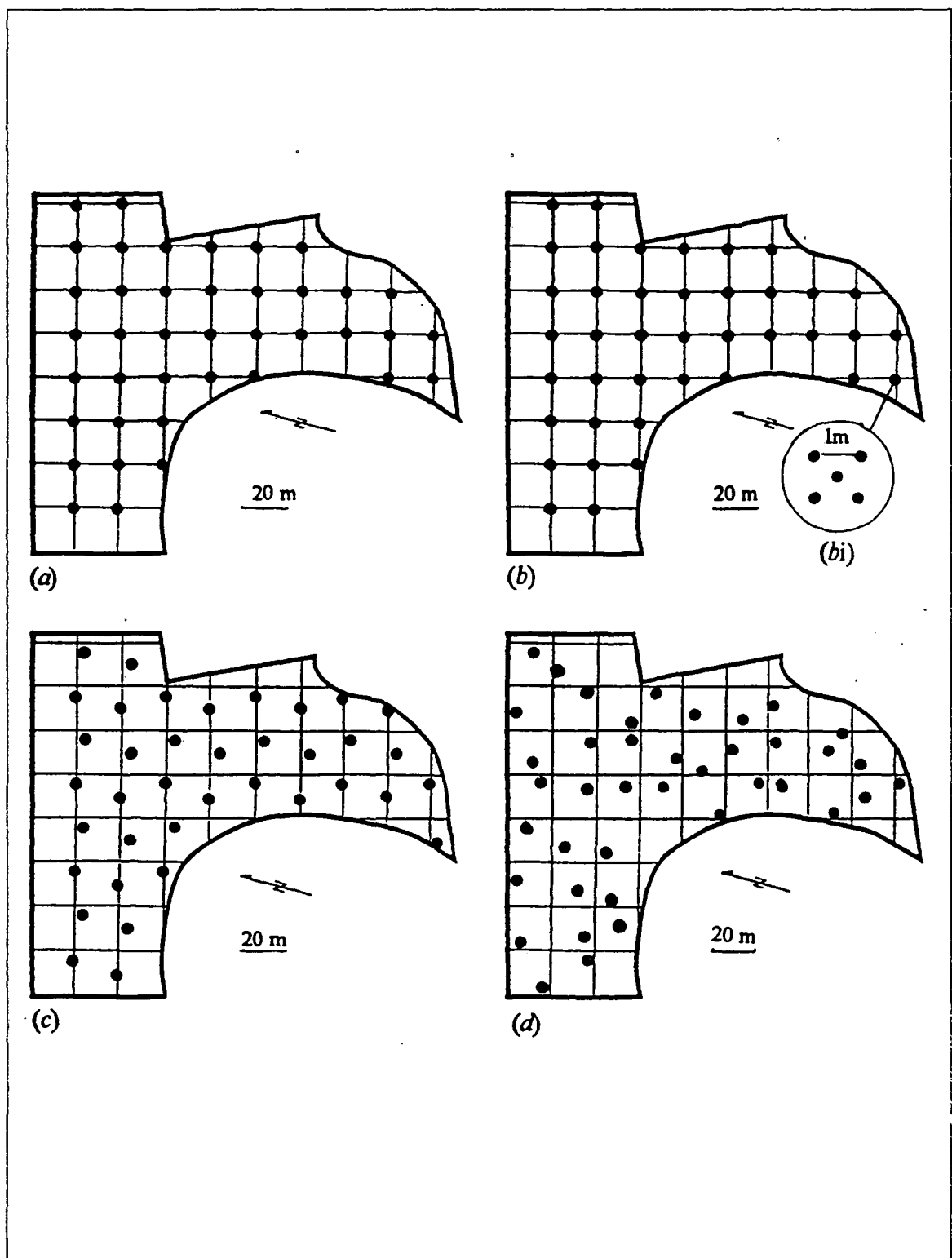


Figure 4.2. The four sampling protocols applied to sample top soil at the Derbyshire site. (a) simple regular grid, (b) composite regular grid, (c) herringbone design, (d) stratified random design.

4.3.1 Statistical Characteristics of Pb and Cu Distributions

The statistical interpretation of analytical measurements focused on the study of two elements with contrasting properties at this site; Pb and Cu. These elements present equal recovery in the samples (both ~ 100%) with the acid dilution used and were selected so that comparisons of the results of the applied sampling methodology could be made. The statistical interpretation of the data was made after applying the quality control procedures discussed below.

The frequency distribution for the mean Pb concentration at each sampling point show a positive skew and approaches a log-normal distribution (Fig. 4.3 (a), (b)). The concentration of Pb is generally high with a geometric mean of around 6000 $\mu\text{g g}^{-1}$ and a maximum of around 30000 $\mu\text{g g}^{-1}$ in the soil. The distribution of Cu is less skewed (Fig. 4.4 (a)) with a mean of around 28 $\mu\text{g g}^{-1}$, about double to the natural background concentration expected in a soil of 15 $\mu\text{g g}^{-1}$ (Rose et al., 1979).

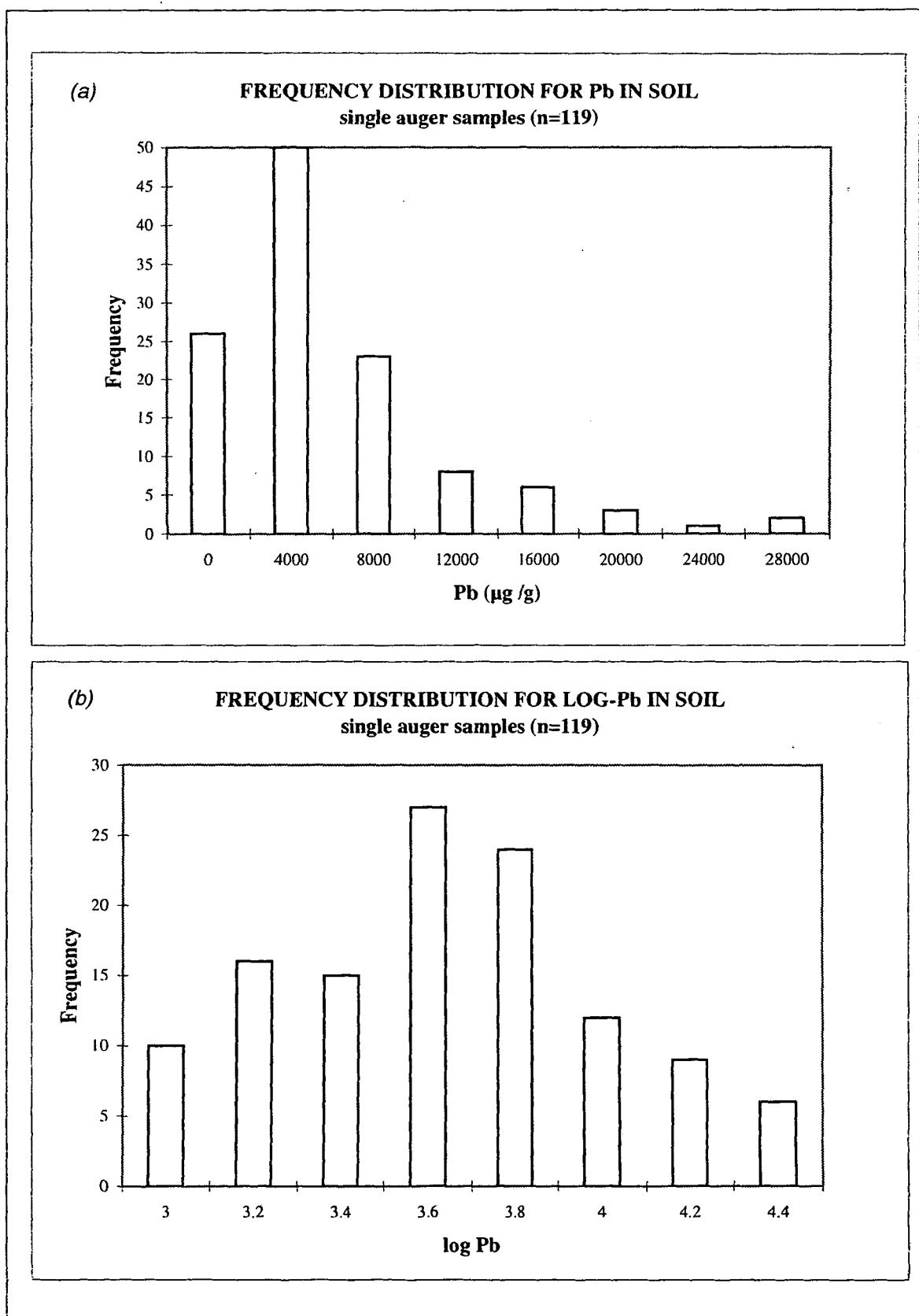


Figure 4.3. Frequency distribution of Pb concentration in soil at the Derbyshire site, showing positively skewed populations, (a). Data from all the single auger sampling designs. (b), Log-transformed data from all the designs.

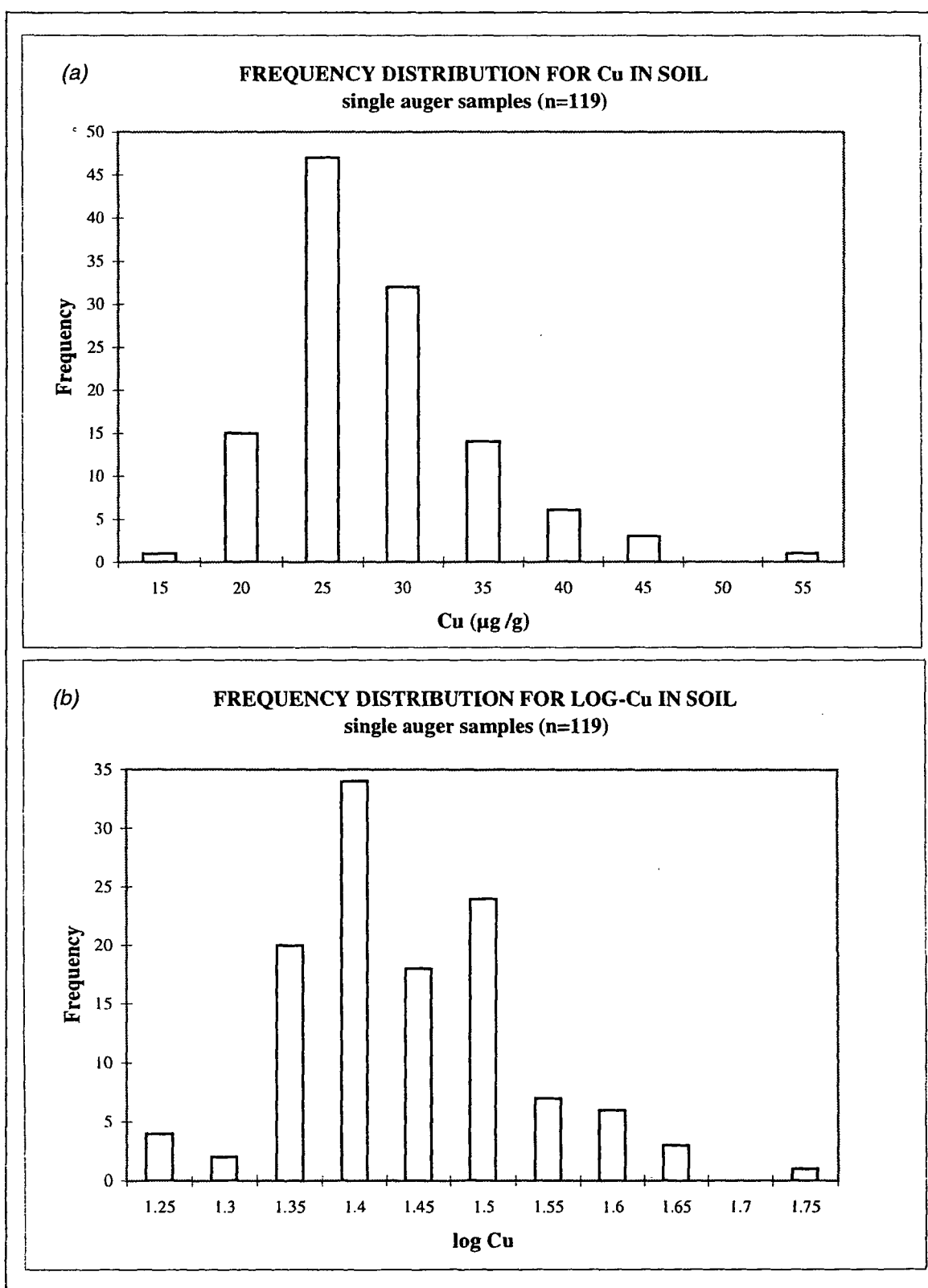


Figure 4.4. Frequency distribution of Cu concentration in soil at the Derbyshire site, showing positively skewed populations, (a). Data from all the single auger sampling designs. (b), Log-transformed data from all the designs.

The descriptive statistics for Pb and Cu concentrations in the field, estimated from the combined data of the single auger sample designs are summarised in Table 4.1:

Table 4.1. Descriptive statistics for Pb and Cu in soil from the initial survey in Derbyshire (number of samples, n = 119).

<i>Statistic</i>	<i>Pb</i>	<i>log-Pb</i>	<i>Cu</i>	<i>log-Cu</i>
Mean $\mu\text{g g}^{-1}$	6202	3.6	28.2	1.4
Standard Error	513.2	0.03	0.6	0.008
Median $\mu\text{g g}^{-1}$	4470	3.65	26.6	1.4
Mode $\mu\text{g g}^{-1}$	1860	3.3	32.3	1.5
Standard Deviation $\mu\text{g g}^{-1}$	5599	0.4	6.2	0.09
Sample Variance	31345085	0.1	38	0.008
Kurtosis	3.91	-0.62	3.4	0.7
Skewness	1.90	0.08	1.3	0.5
Range	28636	1.53	39.5	0.5
Minimum $\mu\text{g g}^{-1}$	864	2.9	17.3	1.2
Maximum $\mu\text{g g}^{-1}$	29500	4.5	56.8	1.8

Pb shows a high geochemical variability (i.e. high variance of the metal concentration between the sampling points) across the site while Cu shows very little geochemical variation. This would apply to laterally variability but also to the vertical variance of metal concentration in this soil profile discussed by Maskall et al.(1995). Dot maps for Pb and Cu were produced from the whole data set (Fig. 4.5). For the production of these maps different symbols were used for every concentration class. The concentration classes were selected based on the mean value and the standard deviation for each data set.

In this way the extent of contamination and the general trends of the metal distribution across the field were studied and useful information for the subsequent experiments was gained.

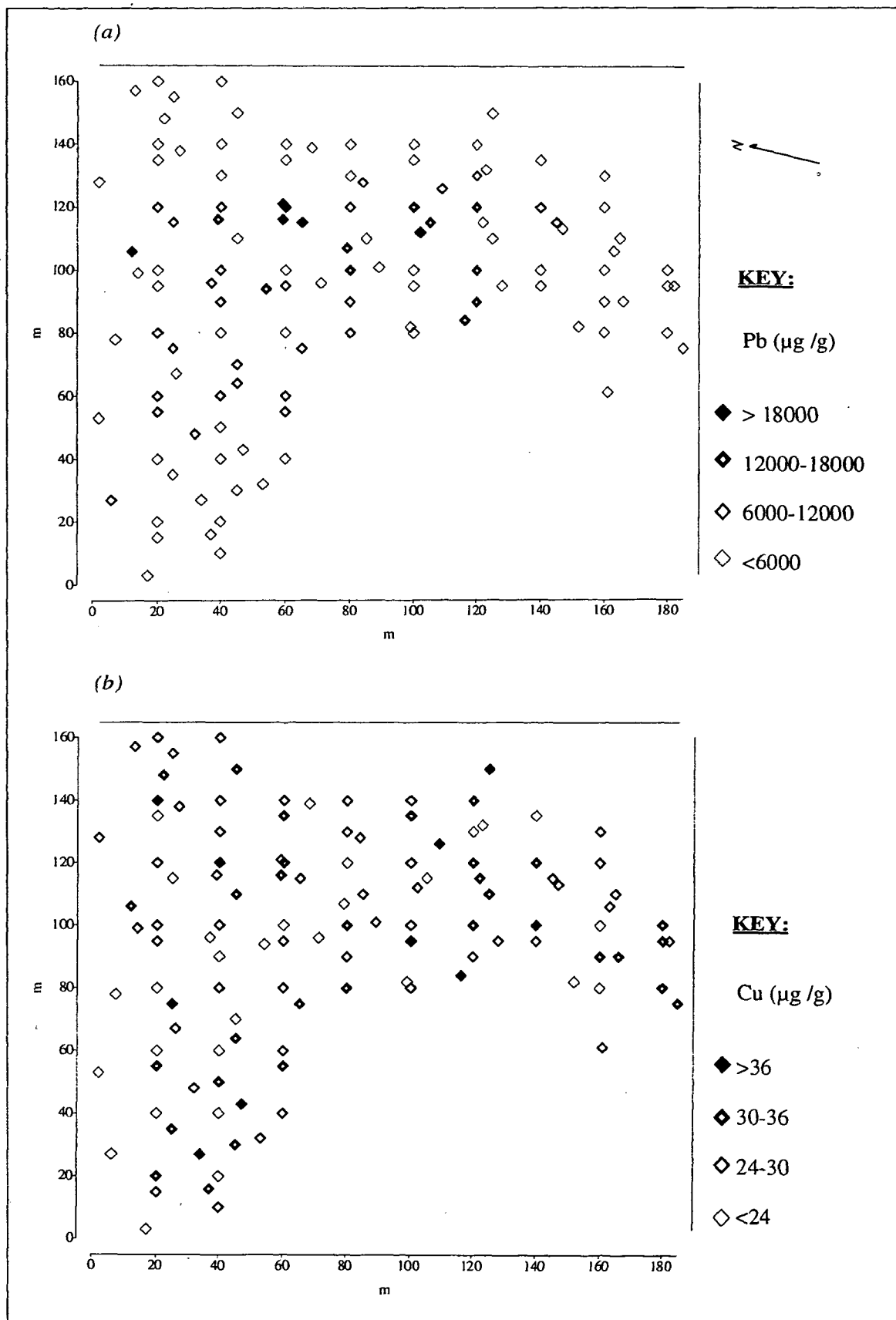


Figure 4.5 Dot maps of Pb (a) and Cu (b) concentration at the sampling points in the Derbyshire site.

From the map for Pb it is apparent that the highest concentrations in the soil are located along the north-south direction on the top of the map. The location of the highest concentrations is probably corresponding to the location of ancient smelting sites along the crest of the scarp slope. The highest Cu concentrations were observed on the north of the field, near the old shed and they are probably related to more recent human impact at the site.

The distributions of Pb and Cu were also studied separately for each one of the sampling designs. All the histograms for Pb show positively skewed frequency distributions (histograms in Appendix 2). The statistics for each one of the sampling designs and the differences from the grand mean are summarised in Table 4.2:

Table 4.2. Pb and Cu concentrations estimated from the four different sampling protocols and difference (%) from the grand mean of $6233 \mu\text{g g}^{-1}$ for Pb and $27.9 \mu\text{g g}^{-1}$ for Cu (*n*, number of samples for each design)

Sampling design	<i>n</i>	Mean ($\mu\text{g g}^{-1}$)	Median ($\mu\text{g g}^{-1}$)	Standard Deviation ($\mu\text{g g}^{-1}$)	Bias (%) from grand mean
Pb					
Regular grid composite samples	40	6322	4320	5633	+1.4
Regular Grid single samples	40	6046	3915	5321	-3.0
Herringbone	39	5680	4470	4420	-8.9
Stratified random	40	6868	4860	6837	+10.2
Combined designs	159	6233	4420	5590	0
Cu					
Regular grid composite samples	40	27.4	26.1	4.9	-1.8
Regular Grid single samples	40	26.9	25.7	5.3	-3.6
Herringbone	39	28.2	27.3	5.6	+1.1
Stratified random	40	29.4	28.3	7.3	+5.4
Combined designs	159	27.9	27.0	5.8	0

The mean concentrations of Pb measured by the four different sampling protocols vary widely between 5680 $\mu\text{g g}^{-1}$ and 6868 $\mu\text{g g}^{-1}$. When the differences are calculated between each of these values and the best estimate of the 'true value' (the grand mean of 6233 $\mu\text{g g}^{-1}$) the bias values range from -8.9 to +10.2%. However when the differences between the geometric mean Pb concentrations are calculated by selected pairs of the four sampling protocols, using the two-sample *t*-test (Davis, 1986) (see Appendix 4) on the log-transformed data, the bias values are not statistically significant. This is primarily the result of the high natural variability in the Pb concentration (i.e. high geochemical variance). The calculation is based on equation 4.1:

$$G_i - G_j = t \times s / \sqrt{n}$$

where; G_i, G_j : geometric means to be compared,

s : standard deviation,

n : total number of results,

When the medians are compared using the Mann-Whitney test (Davis, 1986) (see Appendix 4) they are similarly indistinguishable statistically. Using the equation 4.1 it is possible to calculate the minimum value for a statistically significant difference of the estimated mean that could be detected by comparison with the grand mean Pb concentration. Taking a *t* value of 1.64 for this particular data set, this difference is calculated as 726.5 $\mu\text{g g}^{-1}$ or 11.7%.

For Cu the geochemical variance was much lower, equivalent to 4.1 $\mu\text{g g}^{-1}$ or 68.6% and therefore smaller amounts of significant difference could be detected (0.8 $\mu\text{g g}^{-1}$ or 2.7%). A significant difference of 2.45 $\mu\text{g g}^{-1}$ (8.7%) was detected between two protocols (the simple regular grid and the stratified random protocol) but not between the other protocols.

4.3.2 Estimation of Random Errors from the Initial Survey at the Derbyshire Site

Sampling and analytical quality control was applied to estimate only the random measurement errors of Pb and Cu as previously described in Chapter Three. For this purpose a nested design of sample and analytical duplicates was used (Fig. 4.6).

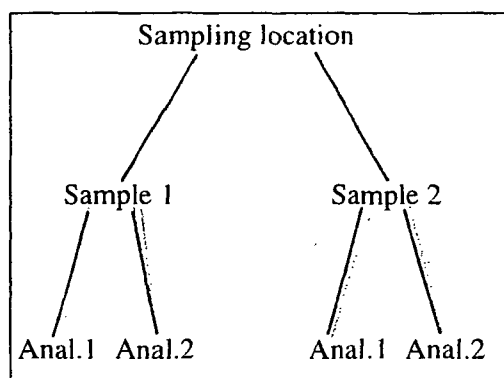


Figure 4.6. Schematic balanced design of sampling and analytical duplicates.

The sampling and analytical precisions for the whole data set as well as for each one of the sampling schemes are presented in Table 4.3:

Table 4.3. Sampling and analytical precisions (as percentages of the mean) for Pb and Cu in soil calculated from the robust ANOVA results on the duplicate samples and analyses in the Derbyshire site.

Sampling Design	Sites Duplicated	Mean Robust	% 2s _{anal.}	% 2s _{samp.}
<i>Pb</i>				
All protocols	27	4953	4	50
Regular grid composite samples	9	6093	4	30
Regular Grid single samples	7	7516	4	50
Herringbone	5	3847	4	50
Stratified random	6	5063	4	70
<i>Cu</i>				
All protocols	27	26.3	4	20
Regular grid composite samples	9	26.2	4	16
Regular Grid single samples	7	25.3	4	28
Herringbone	5	26.1	6	14
Stratified random	6	29.8	4	30

Robust analysis of variance (ANOVA) (Analytical Methods Committee, 1989) was applied to the concentrations measured for the sampling and analytical duplicates so as to estimate separately the geochemical, sampling and analytical variances ($s^2_{\text{geochem.}}$, $s^2_{\text{samp.}}$, $s^2_{\text{anal.}}$ respectively). The full output of the ANOVA results is given in Appendix 3. The total variance is:

$$s^2_{\text{total}} = s^2_{\text{geochem.}} + s^2_{\text{samp.}} + s^2_{\text{anal.}} \quad \text{or,}$$

$$s^2_{\text{total}} = s^2_{\text{geochem.}} + s^2_{\text{meas.}}$$

where s^2_{meas} is the measurement variance.

The proportions of the variances in the total variance were then calculated and displayed as pie charts (Fig. 4.7). Pie charts for the Pb variance in soil using the duplicate data from all four protocols show that the analytical precision is acceptable as it

contributes less than 4% to the overall variance and less than 20% to the measurement variance (Ramsey et al., 1992). Similarly the combined sampling and analytical precision was also acceptable contributing less than 20% to the overall variance. The pie charts also give a clear indication of the effect of taking composite samples, showing a significant improvement of sampling variance over the total variance (from 5.2% for single auger samples to 2.7% for composite samples) in the instance of Pb. This change in variance was shown to be statistically significant using the *F*-test (Davis, 1986) (see Appendix 4) at the 95% confidence level.

For the Cu data from all the sampling designs, the analytical variance (1.7% over total variance) is well within the 4% limit but the big sampling variance of 29.7% of total, makes the measurement precision to be over the 20% acceptable for spatial interpretation of the concentration estimates. However, the mean concentration results can be interpreted within their stated uncertainties. The apparently high proportion of sampling error, in this case, is due primarily to the relatively low geochemical variance of Cu (68.6 % of total variance).

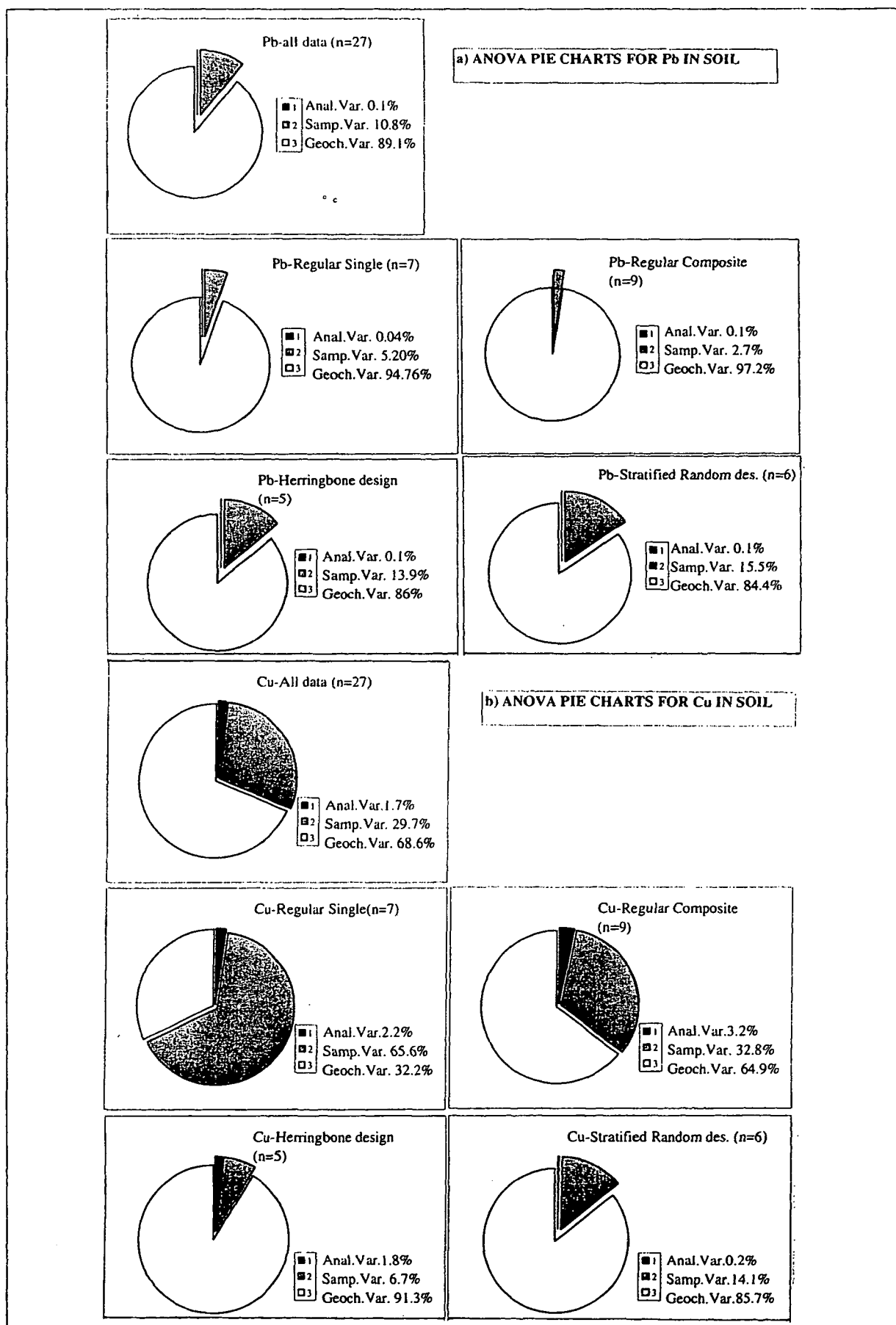


Figure 4.7. Relative importance of measurement errors (from sampling and analysis) and geochemical variability, in the Derbyshire site both expressed as proportions of total variance.

The estimates of random uncertainty for the different sampling protocols used in this initial survey allow the assessment of 'fitness-for-purpose' for each one of the protocols for this sampling target. Since the proportion of measurement uncertainty contributes less than 20% to the total variance for all the sampling designs, logistical factors, related mostly to the ease of applying the sampling protocol, have the main role in the selection of the most appropriate sampling scheme for the objective of estimating the mean metal concentration in the soil. For the particular field the optimal sampling scheme for this objective appears to be the regular grid because it is fast and simple to set up and should therefore be least prone to location errors.

4.3.3 Estimation of Systematic Errors from Analysis, from the Initial Survey at the Derbyshire Site

For the systematic measurement errors from analysis the estimates of analytical bias made separately from the six certified reference materials were generally small (<5%) for both Pb and Cu (Table 4.4).

Table 4.4 Analytical bias for Pb and Cu concentration estimated from six CRMs ($\bar{x}$, mean concentration; s , standard deviation; and c , certified values)

CRM ($n=4$)	$\bar{x}$ $\mu\text{g g}^{-1}$	s $\mu\text{g g}^{-1}$	$s / \bar{x}$ (%)	c $\mu\text{g g}^{-1}$	Bias $\mu\text{g g}^{-1}$	Bias (%)
Pb						
NBS2709	19.7	3.2	16	18.9	+0.8	+4
NBS2710	5352	138	3	5532	-180	-3
NBS2711	1121.4	14.7	1	1162	-40.6*	-3*
BCR141	34.4	3.9	11	29.4	+5	+17
BCR142	36.2	4.6	13	37.8	-1.6	-4
BCR143	127.5	33	3	1333	-35.5	-3
Cu						
NBS2709	32.7	1	3	34.6	-1.9*	-6*
NBS2710	2878.8	82.1	3	2950	-71.2	-2
NBS2711	108.5	0.6	0.6	114	-5.5	-5
BCR141	31.3	1.2	4	32.6	-1.3	-4
BCR142	25.4	0.5	2	27.5	-2.1*	-8
BCR143	235.4	6.8	3	236.5	-1.1	-1

* Bias statistically significant with 95% confidence interval.

Most of the estimates of bias were not statistically different from zero, using the *t*-test with 95% confidence interval. The analytical bias was also estimated over a range of concentrations by fitting a maximum likelihood functional relationship (MLFR) between the measured and certified elemental concentrations (Ripley and Thompson, 1987). The MLFR technique resolves the bias into its translational component (i.e., fixed $\mu\text{g g}^{-1}$ bias) and its rotational component (i.e., the bias as a fixed proportion of the concentration). It also takes into account the uncertainties on both the measured and the certified values, unlike least squares regression. No translational bias was detected (using the MLFR technique), but significant rotational bias of -3.3% was found for both Pb and Cu, suggesting the cause might be volumetric.

4.3.4 Estimation of Measurement Uncertainty from the Initial Survey in Derbyshire: Method 1

From the initial survey in the old smelting site in Derbyshire estimates of the measurement uncertainty for the concentration values can be calculated taking into account the random and systematic errors from analysis, as well as the random errors due to sampling. This method (Method 1) for the estimation of measurement uncertainty follows the ‘bottom up’ approach where estimates of uncertainty from different sources are added together to give the total uncertainty.

The random measurement uncertainty (u_R) can be estimated using this ‘bottom up’ approach, from the ‘holistic’ combination of the sampling and analytical variance described above (section 4.3.2), giving the measurement variance (s_{meas}) as:-

$$u_R = s_{\text{meas}} = \sqrt{(s_{\text{samp}}^2 + s_{\text{anal}}^2)}$$

To express the extended random uncertainty (U_R), with a coverage factor, $k = 2$ (for 95% confidence) this gives:-

$$U_R = k u_R = 2s_{\text{meas}}$$

As uncertainty relative to the mean concentration $\bar{x}$, becomes:-

$$U_R \% = 200s_{\text{meas}} / \bar{x}$$

where $\bar{x}$ is the estimated mean concentration of the analyte in the site. For Pb at this site the relative random uncertainty estimated from the 27 duplicate samples collected from all the sampling designs is 49.8%. The interpretation of this relative random uncertainty assumes that it does not change as a function of concentration.

Analytical bias (B_a), estimated independently by the analysis of the six certified reference materials with a wide range of metal concentration can be added into this estimate of uncertainty. When this component is added the equation for uncertainty becomes:

$$U = U_R + B_a$$

The uncertainty interval of any concentration c (e.g. the estimated mean for the site) becomes asymmetric because of the negative sign of the estimated analytical bias. The upper limit can therefore be calculated from:

$$c + U = c (1 + U_R \% / 100 - B_a \% / 100)$$

and the lower limit from:-

$$c - U = c (1 - U_R \% / 100 - B_a \% / 100)$$

For the mean concentration of Pb across the site, calculated from all the sampling designs ($c = 6233 \mu\text{g g}^{-1}$, $U_R = 49.8\%$, $B_a = -3.0\%$), the upper limit of uncertainty is therefore given by:-

$$c + U = 6233 (1 + 49.8/100 + 3/100) = 9524 \mu\text{g g}^{-1}$$

The lower limit of uncertainty is given by:-

$$c - U = 6233 (1 - 49.8/100 + 3/100) = 3316 \mu\text{g g}^{-1}$$

These equations can be refined if either the bias or the precision varies as a function of concentration. The main problem with this method is that it does not allow for bias originating from the sampling process. This may be caused by the variability between different sampling protocols designed to fulfil the same objective (i.e. estimation of the mean Pb concentration of the site) or from different samplers applying the same protocol.

4.3.5 Estimation of Measurement Uncertainty from the Initial Survey in Derbyshire: Method 2

It is possible in principle to estimate the bias between different sampling protocols, applied by the same sampler on one sampling target. However, at this particular site the bias between the mean values of Pb concentration estimated by each one of the four protocols was found not to be statistically significant for the number of samples taken (see section 4.3.1).

In general systematic errors due to sampling are more difficult to detect than those from chemical analysis. The lack of a reference value for the mean metal concentrations in soil at this site and the effect of the magnitude of random errors are both limiting factors for the estimation of sampling bias. Specifically the great standard deviation of the estimated mean from the different sampling protocols results in a great spread of both the Pb and Cu distributions that largely obscures the existence of bias between the estimated means. Lower values of sampling precision or greater amounts of geochemical variation would have made the detection of sampling bias easier. However, in situations with high geochemical variability of the analytes high precision and trueness of the measurements do not have significant implication on the interpretation of the data.

An estimate of uncertainty in the estimated mean Pb concentration for the whole site, calculated across the four separate sampling protocols as $200s_{4P} / x_{4P}$ (where, x_{4P} is the mean of the means of the four protocols and s_{4P} is the standard deviation) is 60.4%. This shows a much smaller uncertainty for the estimated site means than for concentration estimates at individual locations, as would be expected.

4.4 Conclusions from the Initial Survey at the Derbyshire Site

It is possible to estimate uncertainty in field sampling when different sampling protocols are implemented on an area of contaminated land. The natural heterogeneity of the elemental concentration affects the magnitude of precision and makes the sampling variance the dominant factor in the estimation of measurement uncertainty. This is also a limiting factor in the size of systematic differences that can be detected between the different sampling protocols.

The size of the sampling bias that can be detected at any site therefore varies according to the geochemical variance of the element. This has been demonstrated for Pb and Cu at the particular site for which the minimum sampling bias that is demonstrable is $726.5 \mu\text{g g}^{-1}$ for Pb and $0.8 \mu\text{g g}^{-1}$ for Cu.

No significant difference was found between the mean concentrations estimated by the different sampling designs for Pb. The optimal sampling scheme for this objective therefore appears to be the regular grid because it is the fastest to set up and least prone to location errors.

Duplicate sampling is a useful method for the estimation of sampling precision at individual locations for various sampling protocols. The improvements in sampling precision gain by using a five-fold composite sample at each regular grid point is significant for Pb at this site (5.2% to 2.7% of variance). This improvement judged to be not worthwhile in this instance however, when the original sampling variance was well below the limit of 20% of the total.

The results from sampling using different sampling protocols gave useful information on the levels and the extent of contamination. The particular site studied was judged to be suitable for inter-organisational sampling trials such as a sampling proficiency test and a collaborative sampling trial. With these methods, an estimate of variability between different samplers (i.e. sampling bias) could be made.

The selection of a site to become a reference sampling target should use the criteria of having relatively low geochemical variance, which should minimise the sampling bias that can be detected between the concentrations estimated by different samplers. This sampling target in Derbyshire may have limited usefulness as a reference sampling target, although it may be typical of such heterogeneously contaminated sites.

CHAPTER FIVE. FIRST COLLABORATIVE TRIAL IN SAMPLING UNDERTAKEN AT THE DERBYSHIRE SMELTING SITE

5.1 Introduction

The collaborative trial is traditionally the final stage in determining the performance characteristics of an analytical method. An application to environmental sampling was made in this research to test the feasibility of the collaborative trial in sampling. The objectives of this trial were:

1. to validate the sampling protocol used for the collection of soil samples
2. to *investigate a method to* provide a realistic estimate of measurement uncertainty in contaminated land investigations. This uncertainty is related to the estimation of the mean metal concentration in soil across the site without taking into account the spatial variability of the contamination.

The ~~later~~ sub-objective, although somewhat unrealistic in this context was selected for simplicity of interpretation.

The collaborative trial in sampling (CTS) can be viewed as an application of the “top-down” approach for the estimation of measurement uncertainty (see Chapter Two, section 2.5.2), where multiple samplers use a single sampling protocol. It should be mentioned that the term “sampling” is defined here to cover all of the processes up to and including the preparation of the laboratory sample. The process of “analysis” is defined here to begin with the weighing out of the test portion.

Nine samplers took independently duplicate samples from a contaminated site according to a common sampling protocol ('W' design) (MAFF, 1979). The protocol employed was selected as a frequently used protocol by soil samplers. The eighteen composite samples so produced were analysed in duplicate under randomised repeatability conditions, i.e. in a single run in one laboratory (Fig. 5.1).

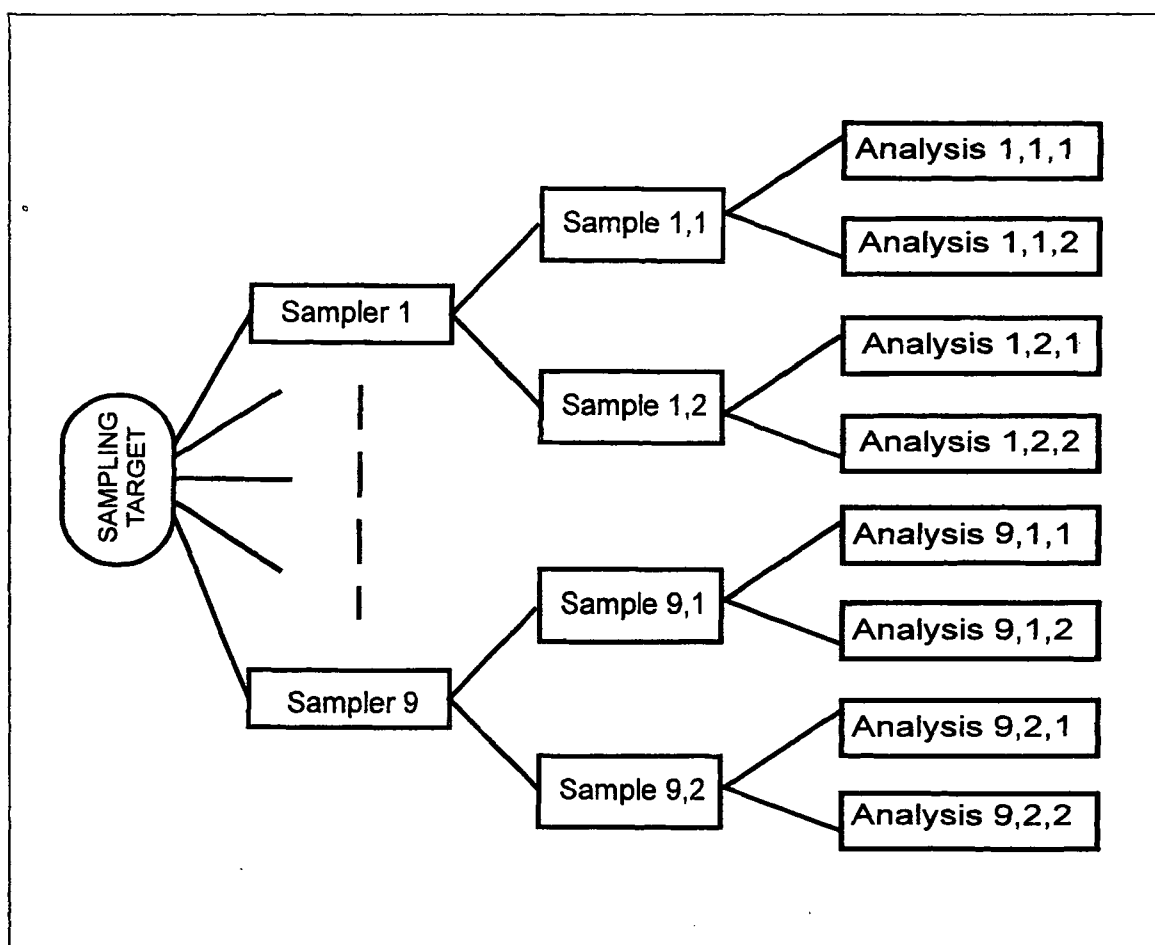


Figure 5.1. Schematic experimental design of the collaborative trial in sampling.

Nested ANOVA was applied to the results to provide estimates of the analytical repeatability standard deviation ($s_{r(a)} = s_0$), the sampling repeatability standard deviation ($s_{r(s)} = s_1$) and the between-sampler standard deviation (s_2). The sampling reproducibility standard deviation is estimated as $s_{R(s)} = \sqrt{(s_1^2 + s_2^2)}$. The precision $s_{r(s)}$ quantifies the variation between the samples collected from a single target by one sampler using the same protocol and equipment within a short period of time, while $s_{R(s)}$ quantifies the variation between samples collected from a single target by different samplers, using different equipment but according to the same protocol (these terms have been discussed in Chapter Two, section 2.5.2). Chemical analysis under 'repeatability conditions' (ISO, 1993) is necessary to avoid confounding analytical variations with sampling variations.

5.2 Experimental

5.2.1 Sample Collection and Preparation

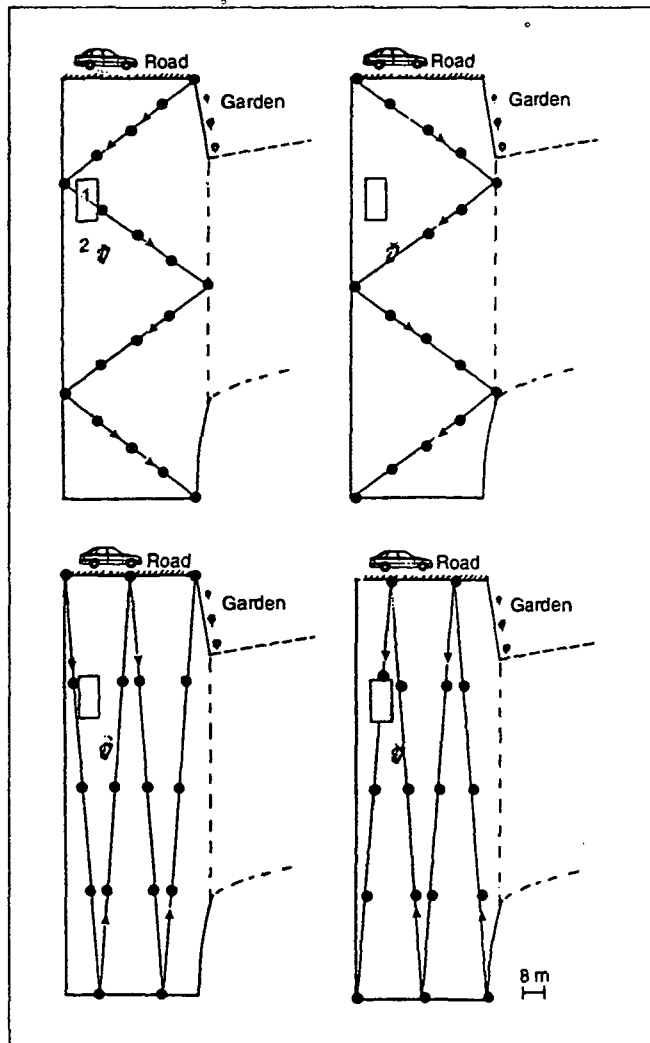


Figure 5.2 Map of the sampling target for the first collaborative trial in sampling in Derbyshire, showing the four possible “W” patterns for collecting the sample increments.

A sub-rectangular field of about 60x150 m was used as the sampling target for the collaborative trial (Fig. 5.2). This was a part of the field used in the pilot study, described in Chapter Four. Nine organisations from commerce, government and academia sent samplers to the site sequentially over the course of one week in August 1994. This was done to preserve the statistical independence of the samplers. An explanation of the aims and the intentions of the project were given to the samplers one month before the sampling period.

The nine participants of this trial were: Paul Nathanail of *Wimpey Environmental*; Judith Nathanail of *Delta^{Simons} Environmental*; Malcolm J. Brown of the *British Geological Survey*; Matt Hill of the *Department of Environmental Sciences, University of Bradford*; Joy Rae of the *Postgraduate Institute for Sedimentology, University of Reading*; Peter Abrahams of the *Department of Geography, University College of Wales*; Janet Cotter-

Howells of the *Department of Geology University of Manchester*; and Aradhana Mehra of the *School of Environmental and Applied Sciences, University of Derby*.

The exact sampling protocol and the equipment required (see Chapter Three, section 3.2.2.1) were given to the participants immediately before commencing sampling. No sampler observed any other during the sampling trial. The composite samples were prepared for chemical analysis in Imperial College. The full procedure is described in detail in Chapter Three, section 3.3.2.

5.2.2 Analytical Quality Control on the Data

Internal quality control for the analytical determinations was provided by the inclusion in the analytical run of six appropriate certified reference materials. The analytical quality control data are represented in Table 5.1:

Table 5.1 Results of the analysis of control materials for the first collaborative trial in sampling in Derbyshire. Concentration and bias values in $\mu\text{g g}^{-1}$.

CRM	BCR141	BCR142	BCR143	NBS2709	NBS2710	NBS2711
Pb						
n	3	3	3	3	3	3
$\bar{x}$	42.22	43.66	1312.59	31.05	5339.55	1179.62
s	11.75	4	0.95	15.02	0.5	2.89
Accepted value	29.4	37.8	1333	18.9	5532	1162
Bias	12.82	5.86	-20.41	12.15	-192.45	17.6
%Bias	43.6	16	-2	64	-4	2
Cu						
n	3	3	3	3	3	3
$\bar{x}$	30.6	25.64	227.8	33.38	2776.88	108.68
s	6.41	3.8	0.28	1.58	0.93	1.95
Accepted value	32.6	27.5	236.5	34.6	2950	114
Bias	-2	-1.9	-8.7	-1.2	-173.1	-5.32
%Bias	-6	-7	-4	-4	-6	-5

Small amounts of analytical bias were evident for both Pb and Cu in the reference materials. Relative high percentages of bias (40-60%) were detected at low concentrations of Pb ($30\text{-}40 \mu\text{g g}^{-1}$). Instrumental memory effects from the very high concentrations in the samples might have been the cause of this bias. In the concentration range of the samples

for Pb ($4000\text{--}10000\text{ }\mu\text{g g}^{-1}$) the relative bias detected for this element was less than 5%. For Cu, the percentages for relative bias are less than 10% for all the reference materials used. However, because all of the samples were analysed at random within one run the detected bias does not affect the interpretation of the trial results.

5.3 Results from the First Collaborative Trial in Sampling

5.3.1 Statistical Interpretation of the Data

The interpretation of the results focused on two elements with very different distributions within the sampling target: Pb and Cu. The statistical characteristics of the distributions of these elements have been described in Chapter Four (section 4.3.1). The concentrations for these elements determined in the laboratory samples from the CTS are presented in Figure 5.3 and in Table 5.2. The statistics are given in Table 5.3.

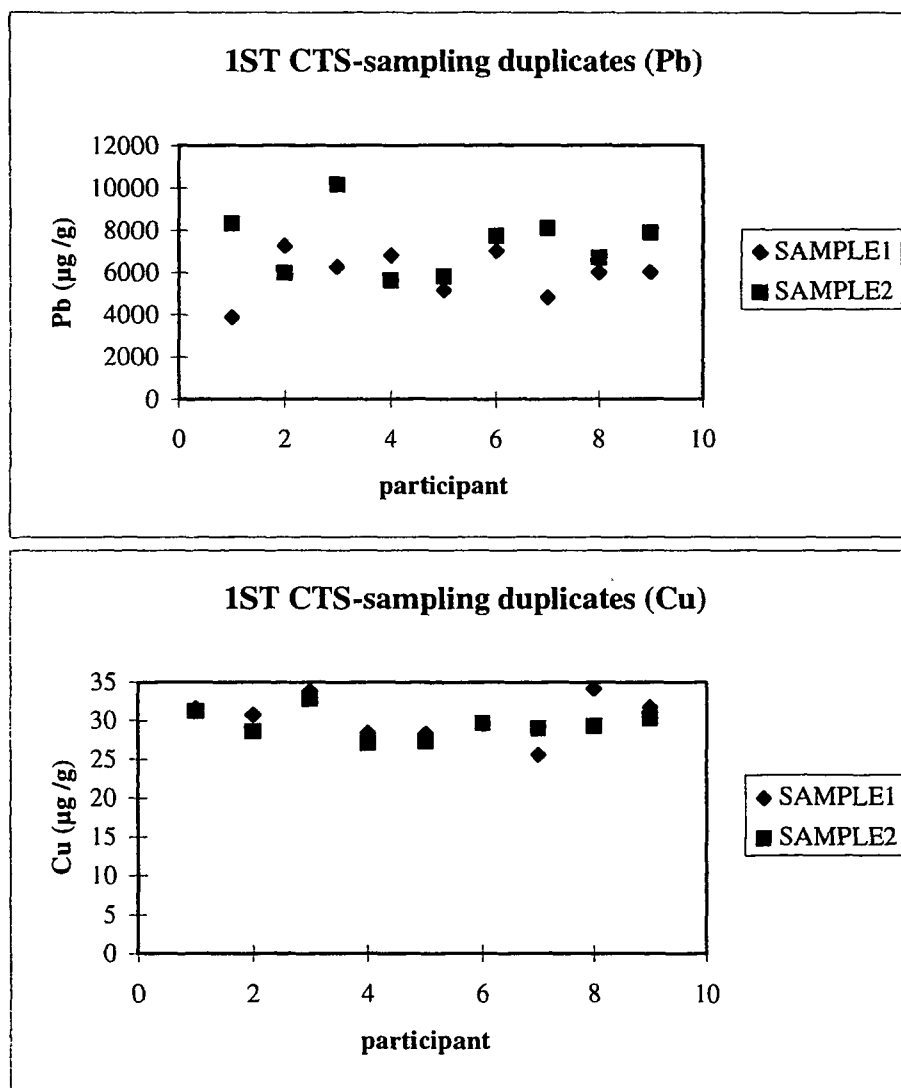


Figure 5.3. Graphic representation of the results of the first collaborative trial in sampling. Each point shows the mean of the duplicated analytical determinations.

Table 5.2. Elemental concentration values from the first Collaborative Trial in Sampling in Derbyshire. Values in $\mu\text{g g}^{-1}$.

Sampler	Sample	Analysis	Pb	Cu
1	1	1	3771	32.1
		2	3964	31.1
	2	1	8271	32.9
		2	8359	29.5
2	1	1	7366	31.3
		2	7152	30.2
	2	1	6079	28.4
		2	5890	28.8
3	1	1	6113	32.8
		2	6369	34.9
	2	1	9826	32.8
		2	10451	32.9
4	1	1	6707	28.2
		2	6867	28.7
	2	1	5647	26.4
		2	5582	27.8
5	1	1	5218	28.9
		2	5102	27.7
	2	1	5953	27.7
		2	5642	26.9
6	1	1	7016	29.9
		2	7015	29.3
	2	1	7978	30.6
		2	7455	28.8
7	1	1	4779	26.1
		2	4869	25.1
	2	1	8166	28.6
		2	8015	29.4
8	1	1	5871	35.3
		2	6114	33.0
	2	1	6842	30.0
		2	6541	28.6
9	1	1	6043	32.5
		2	5977	31.0
	2	1	7915	31.3
		2	7851	29.3

Table 5.3: Statistics from the first collaborative trial in sampling for Cu and Pb ($\mu\text{g g}^{-1}$). (s_0 : repeatability standard deviation; $s_1 = s_{r(s)}$: sampling repeatability standard deviation; s_2 : between-sampler standard deviation; $s_{R(s)}$: sampling reproducibility standard deviation; $\text{RSD}_0, \text{RSD}_{r(s)}, \text{RSD}_{R(s)}$: relative standard deviations).

	Pb		Cu	
<i>Statistic</i>	<i>Classical</i>	<i>Robust</i>	<i>Classical</i>	<i>Robust</i>
Grand mean	6618	6575	30	30
s_0	245	165	1.1	1.1
$s_1 = s_{r(s)}$	1683	1822	1.4	1.2
s_2	0	0	1.7	2.1
$s_{R(s)}$	1683	1822	2.2	2.4
RSD_0	3.7%	2.5%	3.7%	3.7%
$\text{RSD}_{r(s)}$	25.4%	27.7%	4.7%	4.0%
$\text{RSD}_{R(s)}$	25.4%	27.7%	7.3%	8.0%

A question that was considered was whether to use the robust or the classical estimates of the statistics. The robust method reduces the influence of outlying results which would otherwise have a disproportionate influence on the standard deviations. The elimination of the influence of outliers is widely regarded as essential in analytical collaborative trials (Horwitz, 1988). However, outlying samples would be a normal expectation in the sampling analogous of such trials because of the great heterogeneity of the sampling target. Therefore the inclusion of outlying results in the calculation of the standard deviations should be considered if they reflect differences between samples rather than differences between analyses. For the statistical interpretation of the results of this trial the classical estimates were used.

The data for Cu show that the element is evenly distributed throughout the site. The sampling standard deviations are small and comparable with the analytical error. The sampling reproducibility ($s_{R(s)} = 2.2 \mu\text{g g}^{-1}$) which expresses differences between the samplers is higher than the sampling repeatability ($s_{r(s)} = 1.4 \mu\text{g g}^{-1}$) which represents variations in the performance of a single sampler. This is similar to the analogous situation encountered in collaborative trials in chemical analysis, where inter-laboratory variations tend to be greater than those within a laboratory. The sampling relative standard deviations are reasonably small (3.7% analytical RSD, 4.7% sampling repeatability RSD) and the sampling protocol could therefore be regarded as 'fit-for-purpose' using a criterion of 20%.

This is not the case for Pb for which the results reflect extreme variations in the concentrations. The analytical RSD of 3.7% gives good precision, however differences

between duplicate samples dominate giving sampling repeatability RSD of 25.4%. The fitness-for-purpose of the sampling protocol is therefore questionable for Pb using the criterion of 20%.

5.3.2 Estimation of Measurement Uncertainty Using the CTS Data: Method 3

The collaborative trial in sampling can be viewed as an application of the “top-down” approach for the estimation of measurement uncertainty. The experimental design of this trial, utilising ANOVA, allows the calculation of the uncertainty at three different levels: analytical, single sampler and multiple samplers as described below.

For this particular site, it is apparent from the data that there is a large degree of variation in the estimated mean for Pb ranging from 4000 $\mu\text{g g}^{-1}$ to 10000 $\mu\text{g g}^{-1}$. This variation gives rise to high standard deviations and consequently large expanded uncertainty (Table 5.4).

Table 5.4. Standard deviations and expanded uncertainties for Pb and Cu in soil, assuming coverage factor $k=2$). Data from the first CTS in Derbyshire.

Element	Pb			Cu		
Grand mean $\mu\text{g g}^{-1}$	6618			30		
Uncertainty	$s \mu\text{g g}^{-1}$	$U \mu\text{g g}^{-1}$	U%	$s \mu\text{g g}^{-1}$	$U \mu\text{g g}^{-1}$	U%
analytical	245	490	7.4	1.1	2.2	7.3
single sampler	1683	3366	50.9	1.4	2.8	9.3
multiple samplers	1683	3366	50.9	2.2	4.4	14.7

Expanded uncertainties (see Chapter Two, section 2.5.1) were calculated based on the duplicate analyses and duplicate samples by multiplying the adequate standard deviations by a coverage factor of 2 ($U = 2s$). Each one of the uncertainty levels includes the uncertainty from the lower level. The within-sampler variance, for example, contributes extra uncertainty to the uncertainty estimated by the analytical duplicates. Hence the single sampler uncertainty comprises the uncertainty calculated from the two duplicate samples each one of them analysed in duplicate.

An interesting point is that the between sampler variance does not contribute any extra uncertainty for Pb. This is considered to arise mainly from the heterogeneity of the sampling target in respect to the element's concentration. In general the dominant factor of the uncertainty for Pb is the within-sampler precision. Any between-sampler variations are obscured because of the high within-sampler standard deviation.

On the contrary for Cu, the percentage of the expanded uncertainty is doubled when different samplers are applying the same sampling protocol. This is probably because Cu is present at background levels in soil and shows a more homogeneous distribution across the site. The sampling statistics provide an estimate of the sampling uncertainty of $\pm 14.7\%$ the estimated mean value for Cu.

Finally, it should be noted that this method for estimating measurement uncertainty, by using the results from the CTS (Method 3) did not allow for the estimation of the extra uncertainty introduced when samplers use different sampling protocols. For the quantification of this uncertainty the Sampling Proficiency Test, discussed in the following chapter, can be used.

5.3.3 Comparison of Sampling and Analytical Collaborative Trials

The sampling equivalent of the analytical collaborative trial has been tested for the first time at the contaminated site in Derbyshire. The general experimental design for this CTS was based on the principles described in the protocol for design, conduct and interpretation of collaborative studies in analysis (Horwitz, 1988). However, the adoption of the existing methodology that was applied in the CTS demanded several changes. Some similarities and differences between the two types of collaborative trials are discussed below.

Some preliminary work, described in the protocol for analytical collaborative trials, is required for having information on several characteristics of the method to be tested, before conducting the trial. In analogy a pilot study of the sampling target is valuable before conducting a CTS in order to examine the levels of contamination and the distribution characteristics of the elemental concentration in the site.

A common analytical / sampling protocol is given to the participants in both trials. The method must be clearly and unambiguously written (Horwitz, 1988) for the analytical

collaborative trial. However, in the first sampling analogue of the trial, the sampling protocol was not totally unambiguous (e.g. the use of tape for measuring the distances was optional) in order to observe how different samplers interpret the protocol. Calibration procedures are also one of the main factors affecting the interlaboratory analytical variations, however there is no direct analogy of such procedures in the sampling analogue of the trial.

In the protocol for analytical collaborative trials it is specified that at least five materials must be used. In the first CTS only one sampling target (the sampling analogue of the test material) was used. Ideally the CTS should be tested in several distinct targets with different levels of contamination.

For the statistical interpretation of the data, the same procedure described in the protocol for analytical collaborative trials was followed in the sampling version, i.e. by utilising one-way ANOVA on the duplicate measurements. However, the treatment of outlying results is an interesting point for discussion. The elimination of the influence of outliers by applying several statistical outlier tests (Horwitz, 1988) is an essential part of analytical collaborative trials. An outlying result in the CTS however, would indicate the degree of heterogeneity of the sampling target. It was preferred therefore in this first CTS not to exclude any outlying results. Moreover, for the statistical interpretation of the data and the calculation of the precisions, classical estimates of the statistics (that include outlying results) were used.

The main difference between the analytical and sampling collaborative trials seems to be the effect of the heterogeneity ^{in metal concentrations} of the sample target in CTS.

In analytical collaborative trials the reproducibility standard deviation is not affected by the heterogeneity of the test material, since this is carefully homogenised prior to distribution. On the contrary, it has been shown that in the sampling analogue, between-sampler variations may be obscured because of the great magnitude of within-sampler variances. The cause of this later point is the high degree of heterogeneity of the sampling target.

5.4 Conclusions of the First Collaborative Trial in Sampling

The feasibility and usefulness of CTS has been demonstrated with this experiment. By considering two analytes (Pb and Cu) with very different distribution characteristics within the sampling target, very useful information was gained regarding the objectives of the study, i.e. the validation of the sampling protocol and the estimation of measurement uncertainty on the estimated mean metal concentration.

The performance of the sampling protocol has been evaluated and compared with criteria based on fitness-for-purpose considerations. The “W” design used in this trial proved to be fit-for-purpose for Cu but suspect for Pb, using a fitness-for-purpose criterion of 20%. The main factor affecting the suitability of the sampling protocol is the degree of heterogeneity of the sampling target. It should be noted that even for analytes with fairly homogeneous distribution in the sampling target, Cu in this instance, the sampling protocol has been applied at only one level, i.e. close to background concentrations of Cu in soil. Ideally the collaborative trial in sampling should be repeated on several sampling targets with different analyte concentrations. In this way, the sampling protocol could be validated at different concentration levels.

A large degree of variation on the estimated mean was observed for Pb, showing the significant role of soil variability on the outcome of analytical measurements on site investigations. Estimates of the uncertainty associated with the sampling protocol could be made. In the instance of Pb the large sampling precision dominates the total uncertainty. The main reason is the great variability in Pb concentration at the sample target.

Within-sampler precision is as big as between-sampler precision and consequently the between-sampler variance does not contribute any extra uncertainty for Pb.

On the contrary there is a clear effect of the within-sampler and between-sampler variances on the estimation of the measurement uncertainty for Cu. Both of these variances contribute ^{significantly} to the total measurement uncertainty of the mean concentration for the element. The magnitude and the proportion of their contribution were quantified using the data from the CTS.

In general, between-sampler variations may be obscured depending on the degree of heterogeneity of the sampling target. This is a limiting factor in the analogy with such

trials in the chemical analysis. In an analytical trial the organiser is careful to homogenise the trial materials, so that the contribution of heterogeneity to the reproducibility standard deviation is negligible. Moreover the absence in sampling of the process of calibration, one of the main contributors to interlaboratory analytical variations, is a second limiting factor in the analogy between sampling and analytical collaborative trials.

The feasibility and usefulness of the CTS have been demonstrated but only in one sampling target. Repetition of the trial on initially one and eventually several distinct targets was considered essential to validate the results of this first CTS. Specifically, different types of sampling targets in respect to levels and sources of contamination should be considered. Also consideration of the spatial resolution of the contamination and its effect on the measurement uncertainty is a vital but more demanding aspect of the problem. Finally the CTS is not allowing for the estimation of the extra uncertainty introduced when samplers use different sampling protocols. This uncertainty can be quantified by using a Sampling Proficiency Test, discussed in the following chapter.

CHAPTER SIX. FIRST SAMPLING PROFICIENCY TEST UNDERTAKEN AT THE DERBYSHIRE SMELTING SITE

6.1 Introduction

The objective of this part of the study was to assess the applicability of the proficiency test to the process of sampling (Thompson and Ramsey, 1995) as opposed to that of chemical analysis. "Sampling" is defined for this purpose to include all the procedures in the field and in the laboratory, up to and including the preparation of a laboratory sample. The sampling proficiency test (SPT) conducted for this study was designed to assess the general applicability of this method to sampling and also to estimate the uncertainty generated in the sampling of contaminated land when different samplers sample a contaminated site using different sampling protocols.

The particular objective given to each sampling organisation was to design and implement a protocol for the estimation of the mean concentration of two metals in the top soil at a site. By analogy with the proficiency test in chemical analysis, participating samplers took independent samples using a method of their choice, analysed the samples in their own laboratories using methods of their own selection, and reported the results to the SPT organisers. The participants were asked to report only one result for each analyte, which would therefore include errors arising from the processes of both sampling and analysis. In this way the interpretation of the reported results considers the sum of the sampling and analytical bias in one term, defined here as the operational bias.

In order to separate the bias introduced by sampling from that of analysis, a reference material (HRM 31) was prepared. The matrix composition of the reference material was matched to the samples (in mineralogy and analyte concentration) by using a composite material taken previously from the same site. After tests for stability and between- and within- bottle homogeneity, HRM 31 was distributed to the participants, to be analysed in the same run as the sample taken from the site.

The sampling analogue of the z-score that is proposed by the International Harmonised Protocol for Proficiency Testing (Thompson and Wood, 1993) was calculated using appropriate values of assigned value and target standard deviation which were based on the initial survey of the same site discussed in Chapter Four.

6.2 Experimental

6.2.1 Sampling Target and Sampling Procedures

The sampling target used for this study was the same sub-rectangular field (60x170 m) of contaminated land near Wirksworth, Derbyshire that was also used for the collaborative trial (discussed in Chapter Five).

The nine organisations that participated in this study, collected soil samples from the field over a period of five days in August 1994. No sampler observed any other during the sampling. The instructions for the proficiency test were given to the participants one month before the sampling period. The instructions required the participants to apply the protocol of their own choice for sampling the top soil, without defining a particular depth. This was done to give a realistic picture of how soil samplers usually interpret the term 'top soil'. The sampling schemes used by the nine organisations are presented in Fig. 6.1. The participants were asked to analyse the samples in duplicate, using the analytical method of their choice and to report the values for Pb and Cu concentration (on a dry weight basis) to the SPT organisers. These two elements were selected on the basis of ease of analysis, considering also their properties of distribution at this site as discussed in Chapter Four. The participants were also asked to report duplicate analytical determinations of HRM 31, which was to be analysed in the same analytical run as the soil samples.

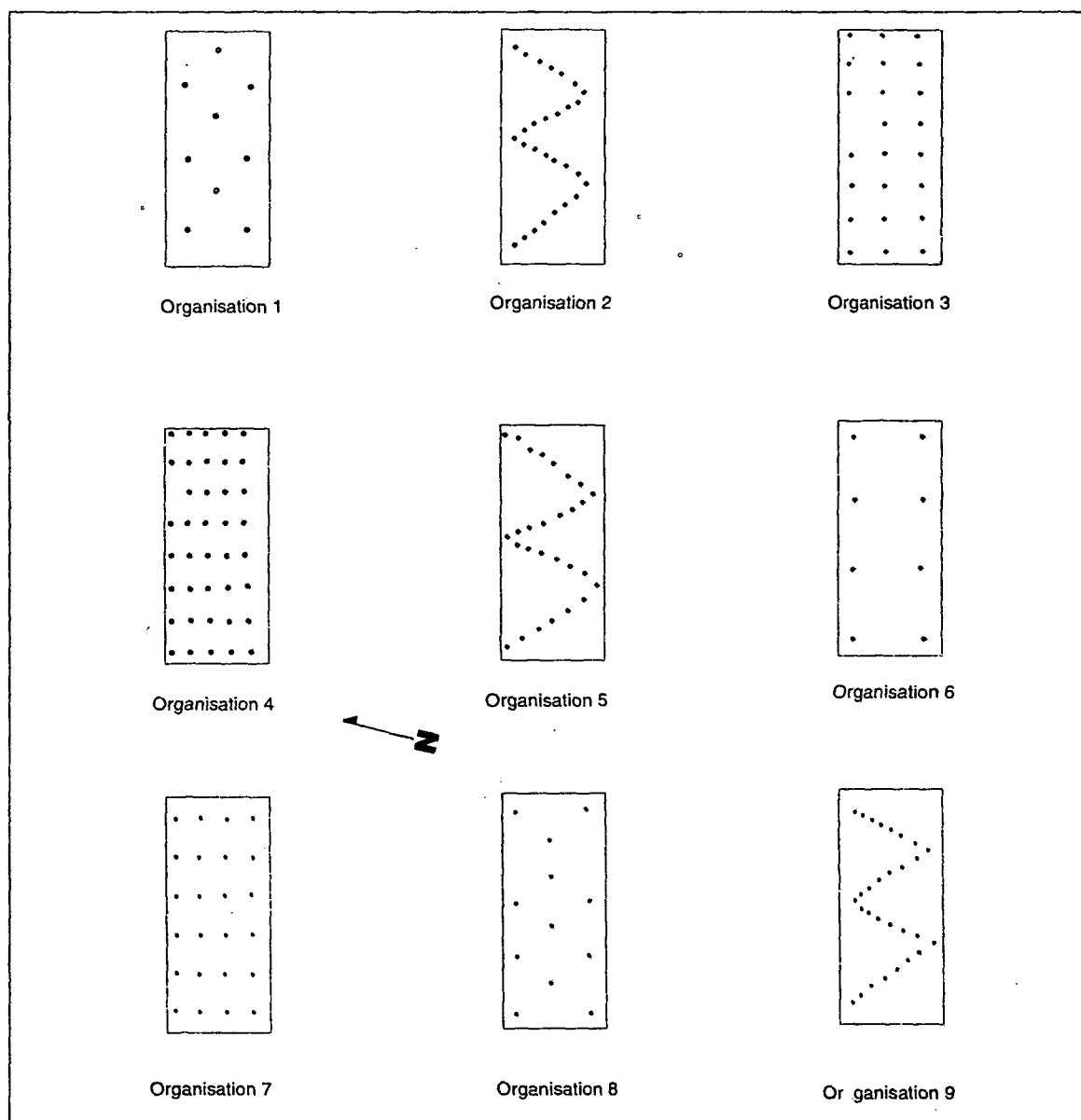


Figure 6.1: Graphic presentation of the nine different sampling designs used by the participants to sample top soil in the rectangular field in Derbyshire for the first sampling proficiency test.

The sampling and analytical procedures used by the participants are presented in Table 6.1. It should be noted here that some of the features of the experiment, i.e. the collection of the data in Table 6.1 and the analytical duplication are not usually employed in analytical proficiency testing and were introduced essentially as diagnostic tools for this experiment.

Table 6.1. Sampling and analytical procedures used by the nine organisations for the estimation of the average Pb and Cu concentration in soil.

Participant	No. of sample increments	Sampling equipment	Sampling technique	Analytical technique
1	9	Trowel	Collect sample increments in separate jars at each sampling point-composite at the end to make one sample. Depth, 0-12 cm	ICP-AES
2	28	Cylindrical auger	Take sample increments 5 m apart. Depth, 0-15 cm	ICP-AES
3	23	Auger	Sample increments 20 m apart. Collect separately-mix in lab to make a composite sample. Depth, 0-15 cm	ICP-AES
4	39	Cylindrical auger	Collect three-fold composite samples at each sampling point. Depth, 0-10 cm	AAS
5	28	Auger	Sample increments 5 m apart. Depth, 0-20 cm	AA for Pb, ICP-AES for Cu
6	8	Auger	Sample increments 50 m apart. Collect three-fold composite samples on a 1 m line at each sampling point. Depth, 0-15 cm	XRF
7	24	Auger	Depth, 0-15 cm	AAS
8	12	Trowel	Remove the top 2 cm of soil, collect sample to 12 cm depth	ICP-AES
9	30	Auger	Collect single auger sample increments 6 m apart. Depth, 0-15 cm	AAS

6.2.2 Preparation of HRM 31

In order to separate the bias introduced by sampling from that of analysis, a reference material (HRM 31) was prepared. The matrix composition of the reference material was matched to the samples by using a composite material taken previously from the site.

Specifically, for the preparation of the reference material all the single auger samples collected in the initial survey on this site (119 in total) were used. The samples were first ground in an agate Tema mill for 3 min in order to be sure that the grain size

was reduced to < 100 μm . Portions of 15 g of each ground sample were mixed together to produce a bulk sample of approximately 2 kg. This bulk sample was homogenised in a rotating bottle. Homogenisation was followed by splitting into 16 bottles using a rotating splitter. The 16 subsamples so produced were dried overnight in the oven at 105 °C and 0.100 g of each subsample was weighted out to produce the test portions for chemical analysis using ICP-AES.

The assigned values for the elemental concentrations of the elements of interest, namely Pb and Cu, in HRM 31 were established by replicate analysis of the material within a run containing six appropriate certified reference materials: BCR141, BCR142, BCR143, NBS2709, NBS2710 and NBS2711. The analytical data of the analysis of these certified reference materials for Pb and Cu are presented in Table 6.2.

Table 6.2. Elemental concentrations and certified values from the analysis of the six CRMs. Concentrations in $\mu\text{g g}^{-1}$. Data from the preparation of HRM 31.

CRM	Pb	Pb certified	Cu	Cu certified
BCR141	41.4	29.4	33.1	32.6
BCR142	38.5	37.8	25.3	27.5
BCR143	1299.9	1333	230.9	236.5
NBS2709	20.4	18.9	31.5	34.6
NBS2710	5245.8	5532	2738.5	2950
NBS2711	1117.5	1162	107.8	114

A statistically significant rotational bias of -4.6% (with 95% confidence interval) was recognised for Pb, by using the MLFR technique (Ripley and Thompson, 1987). The grand mean of 6575.9 $\mu\text{g g}^{-1}$ of the analytical measurements for Pb in HRM 31 was then corrected to give the assigned value of 6895.1 $\mu\text{g g}^{-1}$ of Pb in this reference material. Similarly a rotational bias of -4.8% was determined for Cu and the grand mean of 28.6 $\mu\text{g g}^{-1}$ was corrected to 30 $\mu\text{g g}^{-1}$. The calculations of the accepted values for Pb and Cu in HRM 31 are given below according to Table 6.3.

Table 6.3. MLFR data for Pb and Cu showing statistically significant rotational bias in the HRM 31 samples.

	Pb	Cu		Pb	Cu
Intercept	3.670	-1.138	Slope	0.9537	0.9521
Standard Error on intercept	2.645	0.5392	Standard Error on slope	0.005718	0.01365
% Translational bias			% Rotational bias		
bias	3.670	-1.138	bias	-4.63	-4.79
t_{calc}	1.387	2.110	t_{calc}	8.1	3.5
t_{tab}	2.571	2.571	t_{tab}	2.571	2.571
Statistical Significance	No	No	Statistical Significance	Yes	Yes

Students t value, taken from table for 95% confidence interval and 5 degrees of freedom.

Pb HRM 31 accepted value = $\bar{x}_{\text{Pb}} / \text{slope} = 6895 \mu\text{g g}^{-1}$

Cu HRM 31 accepted value = $\bar{x}_{\text{Cu}} / \text{slope} = 30 \mu\text{g g}^{-1}$

6.3 Results from the First Sampling Proficiency Test

The participants reported their estimates of Pb and Cu concentrations in the soil samples and the HRM 31 to the SPT organisers. A number was allocated to each participating organisation and a full set of results was returned to each of them, keeping the anonymity of the other participants (Table 6.4).

Table 6.4. Reported analytical results for the first sampling proficiency test at the Derbyshire site, (concentrations in $\mu\text{g g}^{-1}$).

Pb						
Pb in soil samples: assigned value, 6296 grand mean, 6171				Pb in HRM 31: assigned value, 6895 grand mean, 7909		
Organisation	Pb A1	Pb A2	Mean Pb	Pb HRM31 A1	Pb HRM31 A2	Mean Pb HRM31
1	4630	4559	4595	6242	6412	6327
2	4850	4450	4650	5500	5550	5525
3	6611	6452	6532	6242	6412	6327
4	5199	5171	5185	5945	5949	5947
5	6012	5860	5936	7972	8149	8060
6	11560	11582	11571	19995	19807	19901
7	6089	5847	5968	6423	6307	6365
8	4285	4120	4203	6242	6412	6327
9	7300	6500	6900	6400	6400	6400
Cu						
Cu soil samples: assigned value, 28 grand mean, 27				Cu in HRM 31: assigned value, 30 grand mean, 26		
Organisation	Cu A1	Cu A2	Mean Cu	Cu HRM31 A1	Cu HRM31 A2	Mean Cu HRM31
1	29.42	28.53	29	27.86	28.02	28
2	<10	<10	<10	<10	<10	<10
3	31.13	30.52	31	27.86	28.02	28
4	27.87	25.83	27	27.33	30.00	29
5	25.70	27.30	27	26.20	26.90	27
6	29.00	29.00	29	32.00	32.00	32
7	*	*	*	*	*	*
8	34.49	26.80	31	27.86	28.02	28
9	31	30	31	27.00	27.00	27
* Value not reported.						

The interpretation of the reported results comprised the assessment of the agreement between the participants, as well as the agreement with the assigned value for this site, using z-scores (discussed in the next paragraph). Also the separation of the analytical bias of the reported results was made by using the results of the analyses of HRM 31. Uncertainty estimates were also made from the results of this trial.

6.3.1 The z-score Assessment

The estimation of bias between participants that is proposed by the harmonised protocol (Thompson and Wood, 1993) uses z-scores. This, requires the definition of an assigned value of concentration, X , and a target standard deviation, s . The X value for this field was derived from the pilot study of this and an adjacent field, discussed in Chapter Four, in which a high density sampling scheme was employed. The assigned values were calculated to be $6296 \mu\text{g g}^{-1}$ for Pb and $28 \mu\text{g g}^{-1}$ for Cu. These values of the mean concentrations of the metals in top soil of the field were the best estimates of the 'true' values since the 'consensus of experts' suggested by the harmonised protocol was unavailable to be used as the assigned value. The target standard deviation, σ , was arbitrarily set at 20% of the assigned value based on a broad estimate of fitness for purpose.

The raw data were converted to the z-scores as recommended in the International Harmonised Protocol. The z-score is given by:

$$z = (x - X) / s$$

where, x is the participants reported result, X the assigned value and s the target value for standard deviation. A satisfactory value of z between -2 and +2 would be obtained if the result was within $\pm 20\%$ of the assigned value. The diagrams for the z-scores are shown in Figure 6.2.

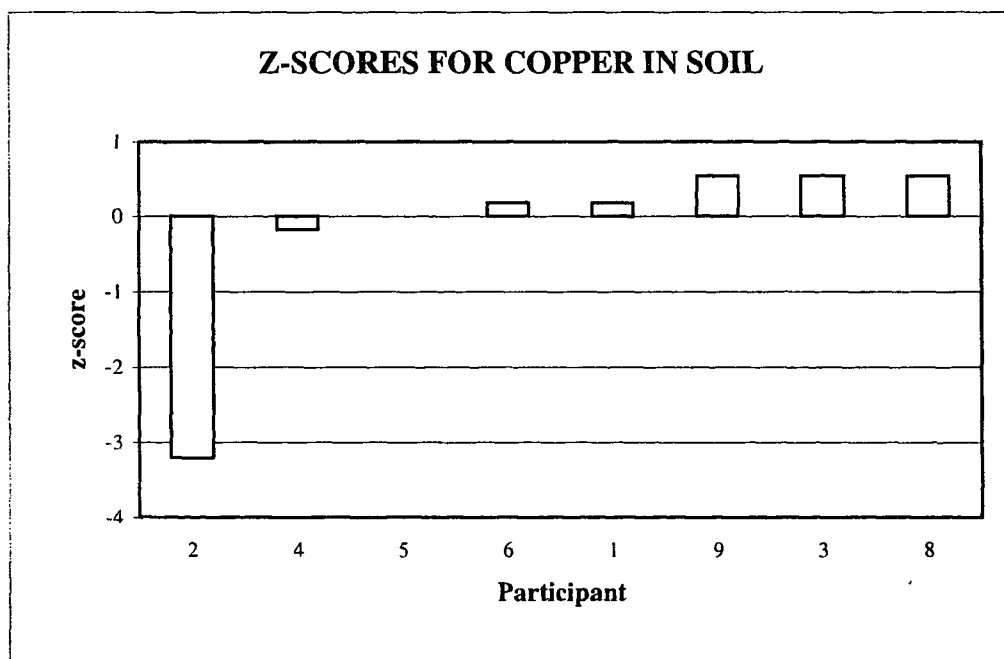
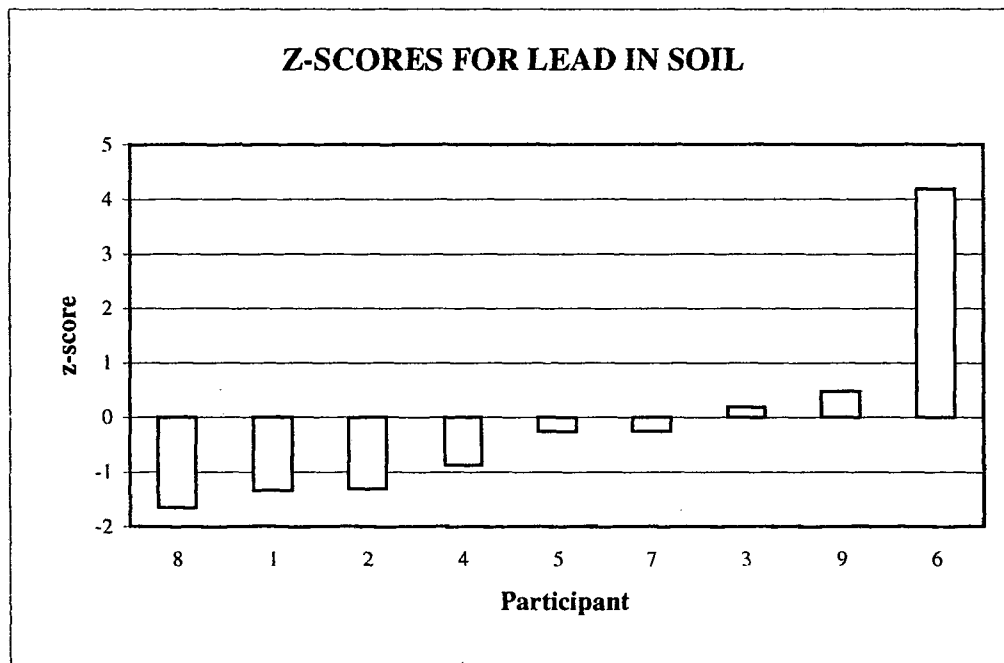


Figure 6.2. z-scores for Pb and Cu in soil from the first sampling proficiency test, showing unsatisfactory performance for participants 6 and 2 for Pb and Cu respectively.

For Pb, participant number 6 has a score that falls outside the range ± 3 ($z_6 = 4.2$), so the performance of this organisation is considered to be unsatisfactory. For Cu participant number 2 with $z_2 = -3.2$ has the most unsatisfactory performance.

The estimates of concentration reported by the participants were also assessed for agreement with the assigned values for the field and the bias was calculated (Table 6.5). Four of the participants reported results significantly different from the assigned mean concentration for Pb and three of the participants reported results significantly different for Cu using the *t*-test with 95% confidence interval (see Appendix 4).

Table 6.5. Bias of the reported results from the assigned value and from the grand mean for Pb and Cu. Data from the first SPT in Derbyshire. * denotes bias statistically significant with 95% confidence interval.

Organisation	Reported Pb ($\mu\text{g g}^{-1}$)	% Bias from assigned value of 6296 $\mu\text{g g}^{-1}$	% Bias from grand mean of 6171 $\mu\text{g g}^{-1}$
1	4595	-27*	-26*
2	4650	-26*	-25*
3	6532	4	6
4	5185	-18	-16
5	5936	-5	4
6	11571	84*	88*
7	5968	-5	-3
8	4203	-33*	-32*
9	6900	10	12
Organisation	Reported Cu ($\mu\text{g g}^{-1}$)	% Bias from assigned value of 28 $\mu\text{g g}^{-1}$	% Bias from grand mean of 27 $\mu\text{g g}^{-1}$
1	29	4	7
2	<10	-64*	-63*
3	31	11*	15
4	27	-4	0
5	27	-4	0
6	29	4	7
7	-	-	-
8	31	11*	15
9	31	11*	15

The agreement of the results between organisations was assessed by calculating the difference between the reported results of each laboratory and the grand mean (Table 6.5). The relative standard deviation for Pb is 36.1%. For the estimates of Pb concentration six organisations reported results over 10% different from the grand mean of $6171 \mu\text{g g}^{-1}$. For Cu, the spread of the reported values was smaller with a relative standard deviation of 26%. Only one organisation estimated the Cu concentration as significantly lower than the grand mean of $27 \mu\text{g g}^{-1}$.

6.3.2 Separation of Sources of Bias

The unsatisfactory performance of the organisations detected from the z-scores as well as the great spread of the values, suggest deficiencies in the measurement practice of the participants but cannot detect the actual source of errors, as being due to either the sampling or the analytical practice. In order to separate the errors originating from chemical analysis an interpretation of the HRM 31 results was made. The reported results for HRM 31 are in satisfactory general agreement with the assigned values, except for organisations number 6 for Pb and organisation number 2 for Cu, which reported values significantly different from the assigned values (Table 6.6). These results indicate that the poor performance of these organisations occurs partially because of analytical errors.

Table 6.6: Bias of HRM 31 reported results from the assigned values of $6895 \mu\text{g g}^{-1}$ Pb and $30 \mu\text{g g}^{-1}$ Cu. Data from the first SPT in Derbyshire.

Organisation	Pb in HRM 31 ($\mu\text{g g}^{-1}$)	% Bias Pb in HRM 31	Cu in HRM 31 ($\mu\text{g g}^{-1}$)	% Bias Cu in HRM 31
1	6327	-8	28	-7
2	5440	-21	<10	-67
3	6327	-8	28	-7
4	5947	-14	29	-3
5	8060	-17	27	-10
6	19901	189	32	7
7	6365	-8	-	-
8	6327	-8	28	-7
9	6400	-7	27	-10

The reported values (x) for the metal concentrations at this site are related to the true mean concentration of the analytes (μ) according to the equation:

$$x = \mu + \beta_s + \beta_a$$

where, β_s is the bias of the sampling method and β_a the bias of the analytical method.

If the analytical bias of the method is assumed to be of a proportional nature, then it is possible to 'correct' the reported concentration for the analytical component of the bias (Table 6.7). The analytical bias β_a for each participant is estimated by comparison of the measured value (x_R) with the assigned value for HRM 31 (X_R):

$$\beta_a = [(x_R - X_R) / X_R] \times 100 = [(x_R / X_R) - 1] \times 100$$

The reported concentration for the soil sample (x) can then be corrected for analytical bias to give:

$$x_c = x \times X_R / x_R = x \times [100 / (\beta_a + 100)]$$

Table 6.7. Lead and Cu results after the correction for analytical bias using the HRM 31 results. Data from the first SPT in Derbyshire.

Organisation	x Pb in soil ($\mu\text{g g}^{-1}$) reported	x_c Pb in soil ($\mu\text{g g}^{-1}$) after correction	x Cu in soil ($\mu\text{g g}^{-1}$) reported	x_c Cu in soil ($\mu\text{g g}^{-1}$) after correction
1	4595	5008	29	31
2	4650	5868	<10	30
3	6532	7118	31	33
4	5185	6012	27	28
5	5936	5078	27	30
6	11571	4009	29	27
7	5968	6465	-	-
8	4203	4580	31	33
9	6900	7434	31	34

The effect of sampling bias for Pb, in relationship to the analytical bias for this element can be more clearly demonstrated after removing the one outlying result of participant number six. This value has been recognised as an outlier using Dixon's Q-test (Dixon, 1950) (see Appendix 4). When participant number six is included the spread of the corrected measurements (RSD 20.2%) is significantly less than for the uncorrected values (RSD 36.1%) for Pb (Table 6.8- outlier included). This difference of 15.9% is an estimate of the error due to the chemical analysis. When this value is removed, the difference between the uncorrected and the corrected relative standard deviation changes from 17.93% to 17.26% (Table 6.8- outlier removed). The error due to chemical analysis is represented in this instance by the difference of 0.69%. The non-significant change in the spread of the values when the analytical error is removed, implies that the dominant source of the error, for all but one participants, is the sampling procedure.

Table 6.8: Statistical characteristics of reported and corrected Pb in soil with the outlying result of organisation six included and removed. Data from the first SPT in Derbyshire.

	Outlier included		Outlier removed	
	Pb reported	Pb corrected	Pb reported	Pb corrected
n	9	9	8	8
Mean ($\mu\text{g g}^{-1}$)	6169	5730	5494	5945
St Dev. ($\mu\text{g g}^{-1}$)	2225	1156	985	1026
% RSD	36.1	20.2	17.9	17.3
% Bias	-2	-9	-13	-6

6.3.3 Estimation of Measurement Uncertainty Using the SPT Data: Method 4

The analytical measurements for this trial were each completed by the participating organisation, rather than centrally as for the CTS. This design was employed so that each organisation had responsibility for all aspects of their measurements.

The resulting scatter in the estimates of the mean concentration of Pb gives an estimate of overall extended uncertainty $U=72\%$. This was calculated as:

$$U = 200s / \bar{x}$$

where, $\bar{x} = 6169 \mu\text{g g}^{-1}$, the mean of the reported results for Pb; $s = 2225 \mu\text{g g}^{-1}$ the standard deviation of the reported results.

For Cu, which is considered to be more homogeneously distributed across the site, the estimate of extended uncertainty from the SPT is 52%. Extended uncertainties calculated after the correction of the reported results for analytical bias (by using the HRM 31 measurements) are 40% for Pb and 16% for Cu.

Interestingly with this method for estimating measurement uncertainty (Method 4), any systematic error in the performance of the participants has been considered to become random error and can thereby be expressed as uncertainty in terms of the ISO definition when the results from the proficiency test are examined as a whole. In this way, each one of the participants is aware of his performance but at the same time estimates of the uncertainty due to sampling can be made for a given contaminated site. However, it should be noted that unlike analytical proficiency tests, such estimates of uncertainty have no general applicability and are site specific.

6.3.4. Effect of the Heterogeneity of the Sampling Target

From the interpretation of the results of the first proficiency test in sampling, it was evident that the heterogeneity of the site in respect to Pb concentration in soil had a major effect on the scatter of the reported data. Generally, variations of the results might occur due to several factors, such as contamination or loss of analyte occasioned by the sampler. It was noticed for example, that samplers operating early in the week of the experiment had difficulties to collect full auger samples because of the dry condition of the soil. This was not so for samplers operating after the rain that occurred in the middle of the week. However, nearly all of the variation between samples probably stems from the heterogeneity, in this instance. Heterogeneity has a direct relation to the fitness-for-purpose of the sampling protocols used by the samplers. Unsatisfactory results reported by the participants may be due to the use of sampling protocols that had been validated in less extreme circumstances.

In this particular trial, no systematic relationship between the used sampling protocol and unsatisfactory performance was observed. For example, organisation number five that used a 'W' design protocol, collecting 28 sampling increments, performed better than organisation number four that collected 39 sampling increments on a regular grid. This suggests that the dominant factor for the discrepancy of the results is the heterogeneous nature of the sampling target.

Several possible ways to overcome this problem in sampling proficiency testing need consideration. The simplest solution would be to make the target value for standard deviation descriptive rather than prescriptive. The z-scores could be calculated by using the robust mean and standard deviation of the participants results. Outlying results would be considered unsatisfactory. In this way only the consistency between the participants would be tested without taking into account any fitness-for-purpose criterion related to the sampling protocol used. However, the fitness-for-purpose of a sampling protocol depends directly on the extent to which the sampling can overcome the heterogeneity of the target.

An other possibility would be to check that the heterogeneity did not exceed a predetermined upper limit before the site was selected for the test. This could be achieved by a reconnaissance sampling survey of the site by the organisers and estimation of the magnitude of heterogeneity in analyte concentrations.

The collection of sampling duplicates by the participants of the trial is also a way to measure the heterogeneity of a site, by performing analysis of variance to calculate the within- and between-sampler variation. The within sampler variation would give an estimate of the heterogeneity of the sampling target. This would be subsequently taken into account in the calculation of z-scores. The value of the estimated within-sampler variance could be used in some way to act as warning that undue heterogeneity was affecting the results. The performance of a collaborative trial could be also used for the same purpose.

A complete solution to this problem would depend on practical experience with repeated sampling proficiency tests. A clear way to proceed would be to select for proficiency testing targets that were heterogeneous to a "typical" extent. Specifically for

proficiency testing on contaminated land sites, this approach would mean the use of sampling targets of similar sources and extents of contamination.

6.3.5. Comparison of Sampling and Analytical Proficiency Testing

It was shown with this experiment that sampling proficiency tests are feasible and are partially analogous to analytical proficiency tests. The major principles of analytical proficiency testing could be applied in the sampling analogous situation. However there are different factors that affect the sampling version that limit this analogy. Similarities and differences of the two types of proficiency tests are discussed below.

There is a clear analogy between the test material in the analytical proficiency tests and the sampling target in the sampling version. Test materials are prepared by the organisers and distributed to the participants after being tested for sufficient homogeneity and stability (Thompson and Wood, 1993). In the sampling version of the test the organisers are responsible of choosing the sampling target e.g. an area of land or a pile of bulk material. Preliminary sampling and analysis performed on the sampling target by the organisers is vital for the determination of its qualities and the subsequent set of the objectives given to the participants.

Stability and homogeneity of the test materials do not have a direct analogue in the sampling version of the trial, however the sampling target has to fulfil several requirements. The target would need to be sufficiently large to be effectively unchanged by sampling. The participants have to visit the target separately to maintain statistical independence between the collected samples. The stable condition of the target is also required during the access period. It should not be used for any other purpose and could be restored to its original use after sampling.

An other apparent analogy is between the analytical and the sampling protocols used in proficiency tests. In analytical proficiency testing there is no prescription of the method to be used. The participants should operate under conditions routinely used for their analyses (Thompson and Wood, 1993). In analogy there is not a fixed sampling protocol to be used by the participants of a sampling proficiency test. They should collect their samples by methods of their choice. Information on the history and the physical

characteristics of the sampling target could be supplied by the organisers but great attention would be needed so as not to bias the participants' decision on the sampling protocol they intend to use.

The duration of the trials does not appear to be in direct analogy either. Analytical proficiency schemes comprise sequential rounds of distribution of the samples. The frequency of the distribution of test sample depends upon a number of factors and usually falls between once every two weeks and once every four months (Thompson and Wood, 1993). In the sampling analogue the trial could be completed within a period of several days or weeks depending on the number of the participants, bearing in mind that the samplers need to travel to the location of the sampling target and to collect the samples independently. The size of the sampling target also plays a significant role because it affects the time required for the collection of the samples. However, possible seasonal variation in analyte characteristics suggests that the sampling period should be as brief as possible, especially for sampling targets susceptible to temporal variation, e.g. water sampling.

The interpretation and the statistical analysis of the results in analytical proficiency tests relies on the assessment of the differences between the reported results and the assigned value. The formation of a z-score is suggested for comparing the bias estimate with a target value for standard deviation that forms the criterion of performance (Thompson and Wood, 1993). The same approach can be used for the interpretation of the results from sampling proficiency tests. However, there are differences in the procedures of setting the assigned value and the target standard deviation in the z-scores. In the first sampling proficiency test the assigned value was derived from previous dense sampling of the area and the target standard deviation was set arbitrarily to be 20% of the assigned value, based on fitness-for-purpose considerations. Practical experience with proficiency testing in sampling is required before a standard methodology for setting the factors of z-scoring is suggested.

An interesting point is that in sampling proficiency tests the performance of the participants in sampling is assessed by analysis. However errors are generated independently in both sampling and analysis. The simplest approach would be to regard the whole process of sampling and analysis as the subject of the proficiency test. This is an

exact analogue of analytical proficiency testing as errors might arise from different sources during the preparation of the test portion or during instrumental measurement (Thompson and Ramsey, 1995). However, the interesting question of which of the processes has the dominant contribution in the measured uncertainty cannot be answered with this approach. In the first sampling proficiency test one common sample (HRM 31) was distributed to the participants to be analysed together with the collected samples. In this way the errors from analysis could be recognised and separated from *sampling errors*. Alternatively all the collected samples could be analysed under repeatability conditions, in a single run, in one laboratory to eliminate analytical errors.

6.4 Conclusions of the First Sampling Proficiency Test

Proficiency testing has been shown to be a concept that is usefully applicable to the practice of sampling. Inter-organisational errors in sampling can be estimated and the performance of the participants can also be assessed. Subsequently any need for improvement of the sampling procedures used can be identified.

Assigned values in normal circumstances will have to be based on the consensus of the participants in each round. It seems unlikely that the result of an expert organisation would be generally acceptable as an assigned value. Organisations exceeding acceptable limits for both Cu and Pb were detected, although in this instance the problems for one organisation were demonstrably analytical rather than due to sampling.

The detection and separation of the sources of error is also possible. The distribution and the analysis of a common sample although not a basic requirement, proved to be very useful for the identification and quantification of the analytical error. In this way, errors due to sampling, the other major source of errors, can be separated.

The effect of heterogeneity of the sampling target has been recognised. Its effects limit the analogy with proficiency testing in analysis and also makes the results of such trials on contaminated land to be site specific. However, since its effect on the results is recognised it is a matter of continuing practical experience on sampling proficiency tests before it could be taken into account when scoring the performance of the participating organisations.

Sampling proficiency testing can also be used as a method for the quantification of measurement uncertainty in sampling. In spite^{of} the high cost of this method it is possible to gain valuable information. Ideally, rounds of sampling proficiency tests have to be repeated on a regular basis so as to validate the methodology and alert the interested bodies on the effect of sampling error in contaminated land investigations.

CHAPTER SEVEN. UNCERTAINTY FROM ITERATIVE SAMPLING DESIGNS USING PORTABLE X-RAY FLUORESCENCE

7.1 Introduction

The evaluation of measurement uncertainty when the analysis of the samples is done in-situ is presented in this chapter. To achieve this, measurements were carried out by using a field portable EDXRF (SPECTRACE TN9000) in the same field in Derbyshire, in collaboration with Dr. Phil Potts of the Open University. The experiment was completed during a two day visit to the sampling target in March 1995. Supplementary information was collected during a second visit at the area in October 1995.

The experiment is related to the objectives of the project as it can be viewed as a means of implementing an alternative sampling protocol, utilising iterative sampling and on-site estimation of the contamination. The uncertainty generated by using the portable XRF (P-XRF) methodology, especially important in the interpretation of measurements by this technique, has been studied. The use on this specific sampling target has been considered appropriate because of the plethora of existing data from previous work in the area. Also the P-XRF methodology provides the means for the study of heterogeneity in metal concentrations at various scales, already recognised in this area of contaminated land. Hence the methodology proved to be a very useful tool for the study of metal variability down to the centimetre scale. The specific objectives of this experiment were:

1. The evaluation of the performance of the technique for in-situ measurements on contaminated land, using an iterative sampling strategy.
2. The estimation of the magnitude of measurement uncertainty for in-situ analysis of contaminated soil.
3. The assessment of the ability of the equipment to estimate geochemical variability in different spatial scales.
4. The rational optimisation of the sampling and analytical protocol using the P-XRF for the investigation of contaminated land, in both the 'in-situ' and the laboratory mode.

5. The identification of needs for further work, which will enable the ultimate limits of performance of the P-XRF to be approached.

The first two objectives were accomplished by comparison of the field P-XRF analyses against laboratory XRF measurements and ICP-AES analysis results, for different analytes and specifically lead and zinc. The third objective was accomplished by taking the measurements at different spatial scales, starting on a wide grid and zooming in over the areas with the highest estimated metal concentrations. The optimisation of the method is based on the ability of P-XRF for spatial mapping at three different scales, for locating contamination hot-spots and for gaining flexibility in the design of sampling protocols. The general usefulness of the technique was assessed in respect of its ability to locate 'hot spots' of contamination, but with special consideration of the general quality and therefore uncertainty of the measurements.

7.2 Background Information on Portable XRF Methodology

In-situ measurements of soils at contaminated sites have the potential of giving both a rapid assessment of the concentration of the contaminant and also first hand information on its spatial distribution and the degree of heterogeneity in an undisturbed position. The disadvantage of in-situ measurement is often the high degree of uncertainty associated with the measurement, caused by factors such as the heterogeneity of the contaminant concentration. If however this uncertainty can be quantified realistically, then these measurements can be fit for many purposes of environmental interpretation.

The collection of top soil samples (e.g. 0-15 cm) using a hand auger and the multielemental analysis of solutions made from the ground samples using a laboratory based technique such as ICP-AES is a well established methodology (Thompson and Walsh, 1989). The estimated values of contaminant concentration from such investigations may vary due to sampling error when different sampling designs are used in the field as described in Chapter Four.

XRF is also a well established methodology for the assessment of contaminated land, usually analysing samples in the form of compressed pellets. Ease of sample preparation as well as analytical trueness and high precision are reported advantages of this methodology (Potts et al., 1995).

Field portable instruments have additional advantages such as the achievement of immediate in-situ analyses of the samples and the possibility of non-destructive analysis of objects e.g. works of art. Their performance for the investigation of contaminated land have recently been evaluated by several workers but after removal of the soils to a field laboratory. Specifically portable XRF devices have been used for the immediate delineation of chromium source contamination 'hot spots' (Kuharic et al., 1993) and the spatial distribution of Pb concentration in residential soils (Kuharic et al., 1993). Portable XRF methodology has been also used to provide data for remedial activities at sites contaminated with Pb and As (Swift, 1995) and the determination of Pb in urban soil and dust samples (Vincent and Boyer, 1995).

Comparisons of the estimates of elemental concentrations made using portable XRF with estimates from other analytical methods such as ICP-AES and AAS have revealed data of acceptable quality when the portable methodology is used properly. However a slight bias between the XRF results and the other methods has been reported (Puls et al., 1994; Kuharic et al., 1993; Swift, 1995; Vincent and Boyer, 1995; Kuharic and Cole, 1995; Pyle et al., 1995). A significant problem of the portable XRF methodology, addressed in most of the above studies is the use of appropriate calibration samples.

7.3 Experimental Design

7.3.1 Instrumentation

Field measurements were undertaken with a Tracor Northern Spectrace TN9000 portable XRF instrument on hire from Thermo Unicam (Cambridge, UK). The instrument comprised a hand held analyser unit and a portable spectrum acquisition and data processing unit. The analyser unit incorporated three radioactive isotope excitation sources: ^{55}Fe , ^{109}Cd and ^{241}Am and a mercuric iodide X-Ray detector. To undertake an analysis in the field, the analyser unit was placed against the sample surface which was excited with each of the three sources in turn (Figure 7.1).

For the analysis of samples (e.g., powdered soil) in the laboratory, the analyser unit was mounted in a lab stand with the analyser window pointing vertically upwards and

covered by a sample lid, the position of which (open or closed) was interlocked to the operation of the instrument. Samples were then conveniently contained in conventional XRF sample cups, placed in position over the window, and analysed once the lid was closed. Whatever the form of sample presentation, spectra were accumulated in the spectrum acquisition and data processing unit and at the end of the pre-set count time, fluorescence intensities were measured using a spectrum deconvolution procedure based on region-of-interest integration. Data were then corrected for matrix effects and quantified using a fundamental parameter procedure. Quantitative results were displayed for immediate appraisal and stored for subsequent downloading to an external microcomputer via an RS232 interface.

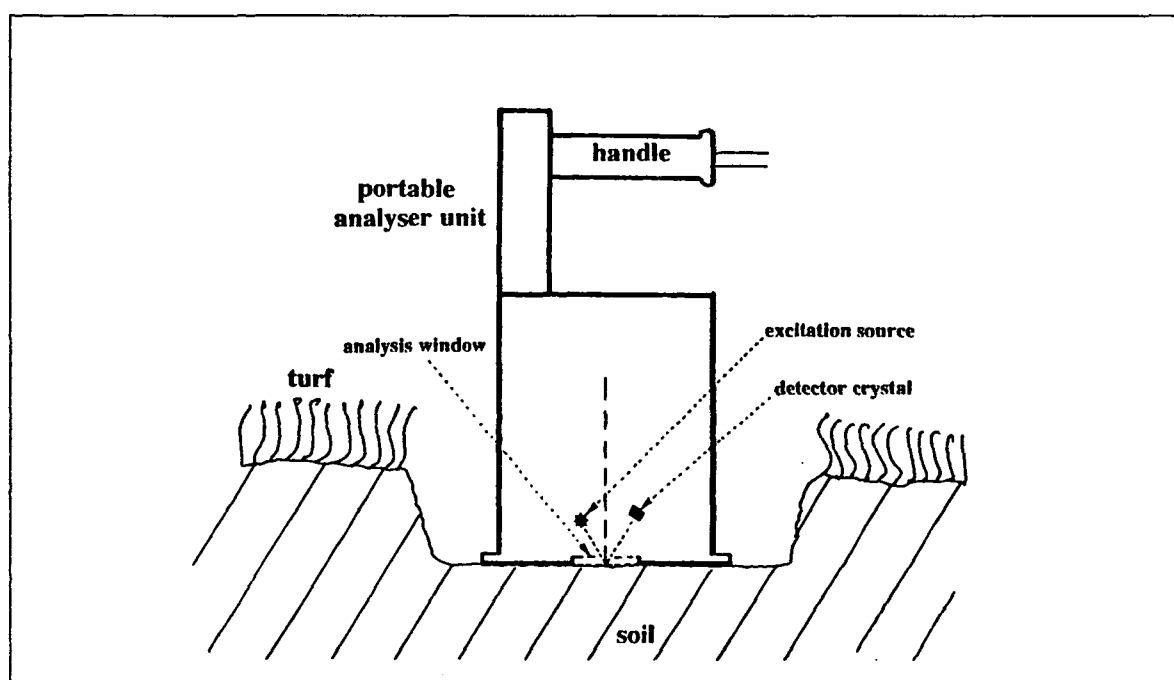


Figure 7.1. Schematic representation of P-XRF instrumentation for in-situ analysis.

7.3.2 Method Development and Selection of Sampling and Analytical Protocol

The development of the analytical protocol using the portable XRF method was based on the performance of the technique in the laboratory. Previous use of the instrument in laboratory conditions for the analysis of silicate rocks prepared as compressed powder pellet samples, showed impressive performance characteristics (Potts et al., 1995). Repeatability precision, using a 200 s count time per source was in the range

of 0.45-1.8% for the major elements and 2-5% for trace elements. A high degree of linearity was also achieved in the relationship between analysed and expected values for 70 international reference materials. Detection limits for the two elements of interest were found to be $39 \mu\text{g g}^{-1}$ for Pb and $63 \mu\text{g g}^{-1}$ for Zn, both representative of 200 s count times.

In the particular field used in this study, previous work by auger sampling and ICP-AES analysis has revealed metal concentrations in 0-15 cm top soil indicative of a heavily contaminated site. Lead concentration had a mean concentration of $6233 \mu\text{g g}^{-1}$ (see Chapter Four) and Zn $614 \mu\text{g g}^{-1}$ over the sampling target. The spatial distributions of these elements were also well known from this pilot study, so that this site was considered suitable for the evaluation of the portable XRF instrument. A proficiency test and a collaborative sampling trial (discussed in Chapters Five, Six) had also been set up in a part of the same field. Useful information about the sampling variance from these trials was also used in this study.

An iterative sampling strategy was employed in which the areas of highest concentration were located by in-situ P-XRF analysis, using a 30x30 m grid (Fig.7.2). A second, more dense grid (10x10 m) was then employed on the same day, focusing on these hot spots of contamination. Finally measurements were taken on a 1m^2 square area in two sites in the field to observe the small scale lateral variation of the metals in soil. The measurements for each survey were taken on the nodes of the grid, after the removal of a turf approximately 15 cm x 15 cm x 3 cm deep with a spade. The removal of the turf was considered necessary to achieve a reasonably flat and unvegetated soil surface for analysis, keeping in mind the critical penetration depth of the X-Ray beam for Pb analysis of about 2 mm.

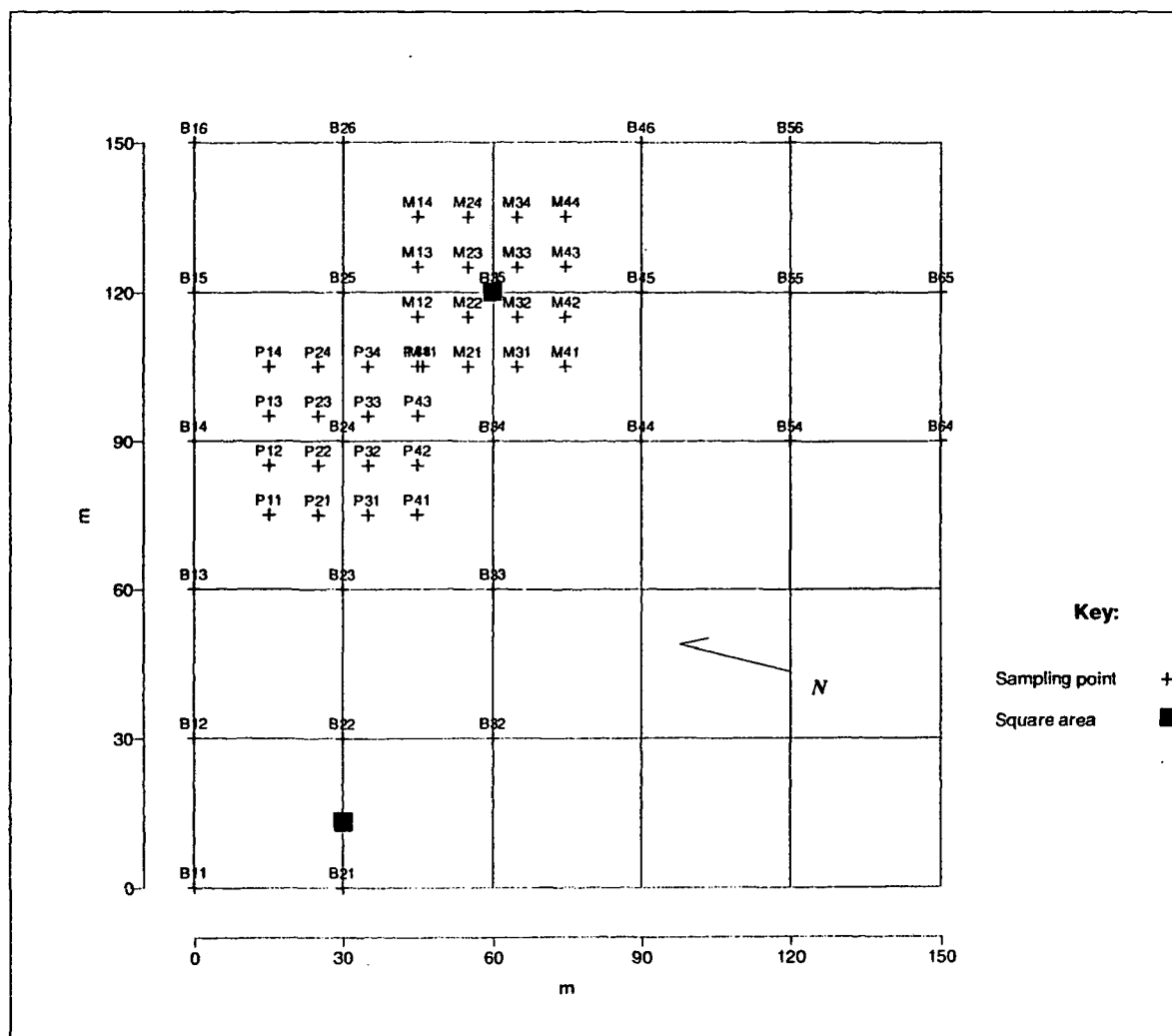


Figure 7.2. Sampling design for the P-XRF measurements and collection of auger samples in the Derbyshire site. 30 m, 10 m, 1 m grids.

Duplicate measurements were taken at every sampling point at a distance of 2 m away from the initial sampling point, for the estimation of measurement precision that includes sampling error. This distance represents the possible location error of the sampling points on the field given the surveying technology employed (30 m tape triangulated). Auger samples 0-15 cm deep, 2.5 cm diameter were also taken in each of the measuring sites for XRF and ICP-AES analysis in the laboratory. The time required at each of the sampling locations was approximately 3-5 minutes allowing 24 sampling locations to be covered in half working day.

For this particular experiment the radioisotope sources in the XRF used to excite fluorescence were ^{109}Cd measuring for 50 s, ^{55}Fe for 20 s and ^{241}Am for 5 s. These low count times did not achieve the best detection limits possible but did allow more

measurements to be made in the time available and hence a survey with higher spatial resolution. The subsequent relatively high detection limits (i.e. Pb 39 $\mu\text{g g}^{-1}$, Zn 63 $\mu\text{g g}^{-1}$) did not affect the results significantly because of the high concentration of the contaminants in the soil at this particular sampling target.

The samples collected for laboratory analysis were dried at 65°C, disaggregated in a pestle and mortar and the <2 mm fraction was ground (<100 μm) using an agate ring grinder. Analytical test portions of these samples were analysed as loose powder with the portable XRF in laboratory mode, using a sample cup fitted with a 6 μm thick polyester film. The measurements were taken for 100 s, 40 s, and 20 s using Cd, Fe and Am radioisotope sources respectively, simulating the conditions in a field laboratory rather than for the in-situ measurements. The same samples were analysed using ICP-AES after treatment with a mixture of nitric and perchloric acids to solubilize the analytes which were presented for analyses in dilute hydrochloric acid solutions (Thompson and Walsh, 1989).

For the analytical quality control, four reference materials (HRM1, HRM2, NBS2710, HRM31) in a form of compressed pellets were run periodically between the measurements in the field. In the laboratory XRF analysis five reference materials were used (HRM1, HRM2, HRM31, GXR-2, NBS2710) and for the ICP-AES analysis seven reference materials (BCR 141, BCR 142, BCR 143, NBS 2709, NBS 2710, NBS 2711, HRM 31). No statistically significant analytical bias was observed for any of the reference materials used. Analytical duplicate measurements were taken on each sampling duplicate in the laboratory, for ten of the sampling sites, in order to estimate the sampling and analytical precision.

7.4 Results

The interpretation of the results, corresponding to the objectives of the study, was focused on the following points discussed below: i) the estimation of the magnitude of the sampling and analytical variances as proportions of the total on the three data sets, i.e. field P-XRF, laboratory P-XRF and ICP-AES, ii) the estimation of the relative bias between the P-XRF methodology and the two other analytical methods, iii) the

determination of the origin of the bias, and iv) the evaluation of the ability of P-XRF for spatial mapping of the contamination.

7.4.1 Analytical and Sampling variances

The proportion of the total variance caused by the measurements for Pb by all three techniques (Fig. 7.3) is within the suggested limit of 20% (Ramsey et al., 1992) (raw data in Appendix 3). For the field P-XRF it was not possible to separate the analytical and sampling components of the measurement variance (Fig. 7.3a) since no analytical duplicates were available, but it is clear that the sampling is the dominant contributor for both laboratory P-XRF and ICP-AES (Fig. 7.3b & 7.3c). The similarity between the percentage of measurement variance for both the field and laboratory P-XRF also suggests that this would also be the case for the measurements by field P-XRF.

Zinc shows similar percentages of measurement variance within the 20% limit of the total (Fig. 7.4). The sampling variance is the dominant factor of the measurement variance with an exemption of the laboratory P-XRF measurements for which the robust ANOVA results give analytical variance greater than the sampling. This is not the case however for the classical ANOVA results where the proportion of the sampling variance is greater.

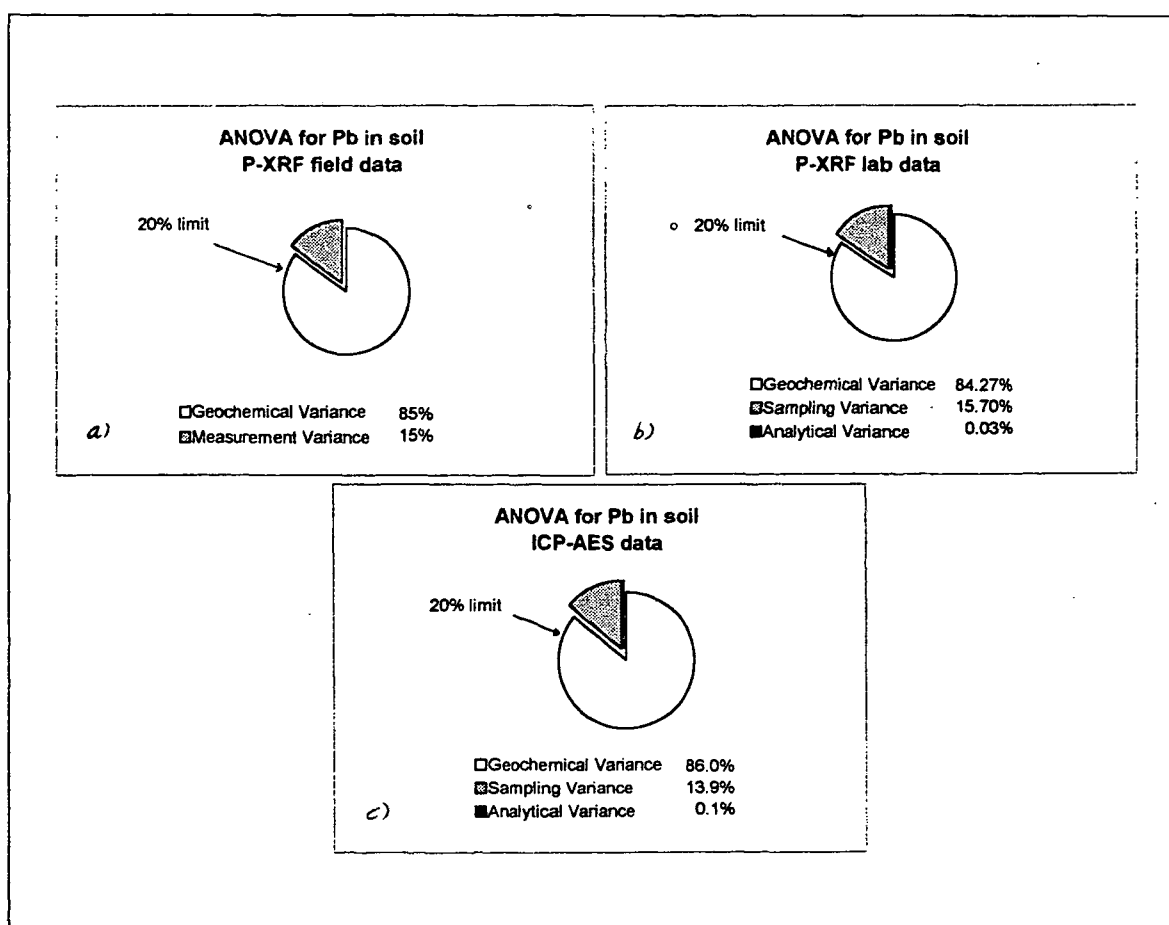


Figure 7.3. ANOVA pie charts for Pb in soil, showing acceptable proportions of measurement variances over the total for all three analytical methods. (a) P-XRF field data (n=24), (b) P-XRF lab data (n=10), (c) ICP-AES data (n=10). (n = number of duplicate samples).

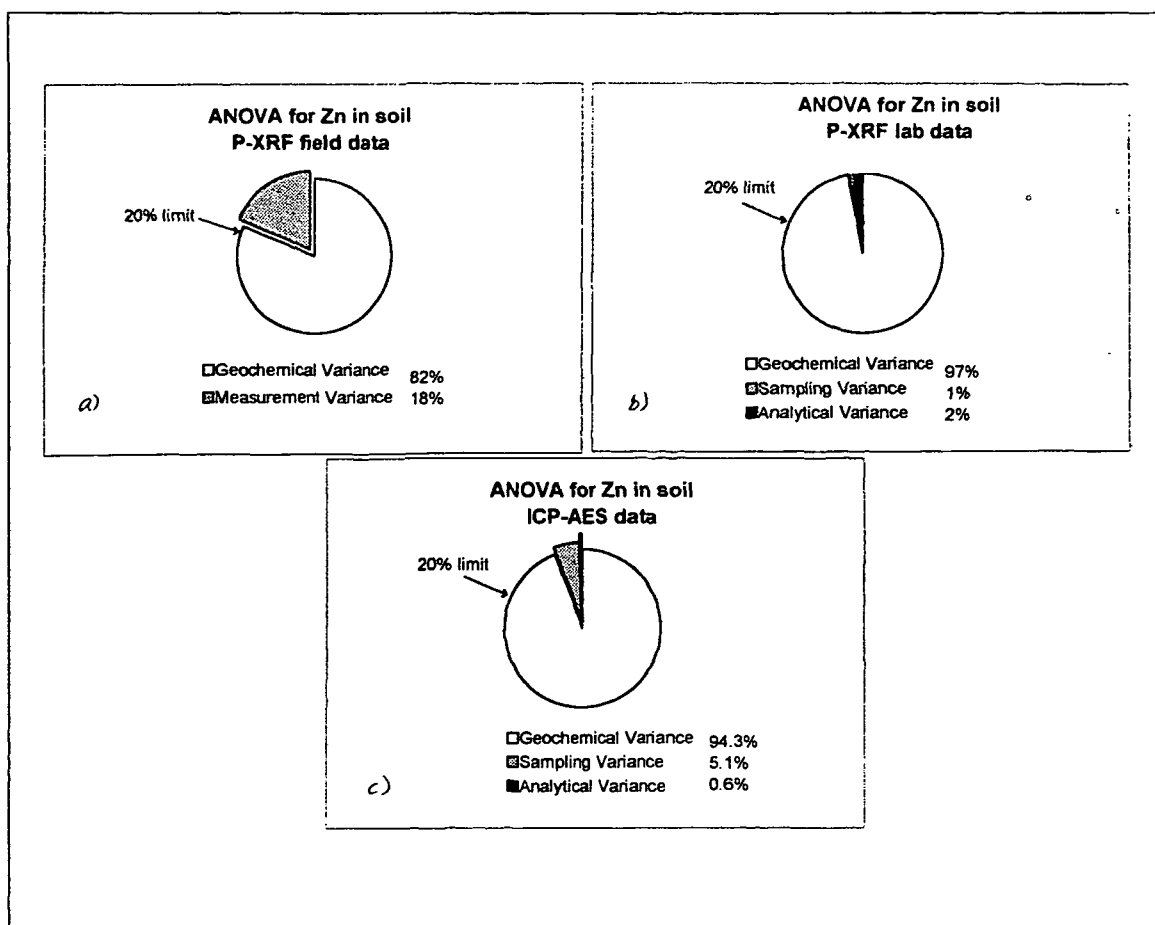


Figure 7.4. ANOVA pie charts for Zn in soil, showing acceptable proportions of measurement variances over the total for all three analytical methods. (a) P-XRF field data (n=24), (b) P-XRF lab data (n=10), (c) ICP-AES data (n=10). (n = number of duplicate samples).

7.4.2 Bias of P-XRF Measurements

The bias at the typical analyte concentrations was estimated using one common certified reference material for all three methods of analysis (Table 7.1).

Table 7.1. Analytical bias estimated with CRM NBS2710 by the three analytical methods (field P-XRF, lab P-XRF, ICP-AES).

CRM-NBS2710	Pb	Zn
Accepted value ($\mu\text{g g}^{-1}$)	5532	6952
Measured field P-XRF ($\mu\text{g g}^{-1}$) (n=10)	4935	6232
% Bias	-10.8	-10.4
Measured lab P-XRF ($\mu\text{g g}^{-1}$) (n=4)	4898	6278
%Bias	-11.5	-9.7
Measured ICP-AES ($\mu\text{g g}^{-1}$) (n=2)	5228	6403
% Bias	-5.5	-7.9

For Pb the bias of -5.5% for ICP-AES probably originates in under-correction of matrix effects. For P-XRF the bias of -11% for both field and in-situ measurements probably reflects a slight mismatch between the real and composition of the material assumed by the calibration. There are similar levels of bias for the Zn measurements. The use of a dry, finely ground, pelleted reference material to estimate the bias for the in-situ measurements of moist, uncompacted and not ground soils with an undulating surface is obviously severely limited. Nominally however, these levels of bias are probably acceptable and 'fit-for-purpose' in the context of this work, but these limitations of bias estimation will be discussed below.

Bias can also be estimated between the three measurement techniques rather than against an accepted reference value. A simplistic comparison between the measurements made for Pb by field P-XRF and those by ICP-AES show a rotational 'bias' of $-57\% \pm 6\%$ (1s) in the former, by simple linear regression (Fig. 7.5a). For Zn rotational bias of $-63\% \pm 6\%$ was observed by the same method (Fig. 7.5b).

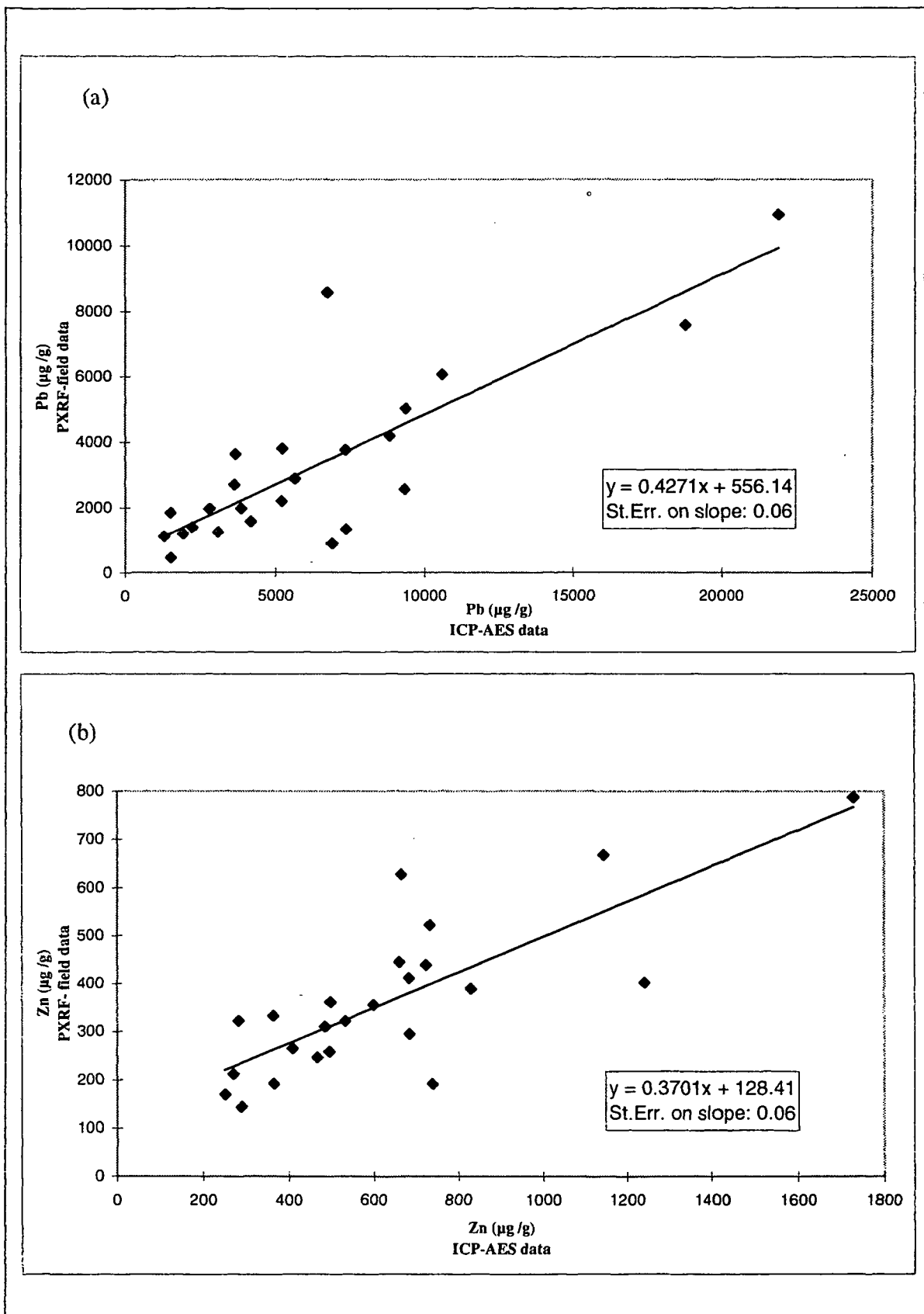


Figure 7.5. (a) Simple linear regression of Pb field P-XRF measurements against Pb ICP-AES measurements, (b) Simple linear regression for Zn.

This 'bias' however, includes a measure of sampling bias as well as purely analytical bias, because the samples used were not identical. The field P-XRF analysed a depth of few millimetres at the top surface of the soil, whereas the sample for ICP-AES came from the top 150mm. Furthermore the field P-XRF measurements were made on wet, uncompacted soils, whereas the ICP-AES measurements were on dried, homogenised and dissolved powders. A bias between the two techniques is to be expected therefore, but it is interesting to investigate the components of this bias.

When the bias for Pb is compared between field P-XRF and the laboratory P-XRF a very similar rotational 'bias' is apparent, of $-52\% \pm 7\%$ (1s). This suggests that the bias is not produced by a difference in the performance of the analytical technique, which is identical here, but due to one of these other factors of sampling and sample preparation.

The use of simple least squares linear regression has been shown to have limitations in the detection of bias between analytical methods (Ripley and Thompson, 1987), but such limitations are considered not to have affected this broad conclusion.

7.4.3 Origin of Bias for the Field P-XRF Measurements

One hypothesis for the apparent low 'bias' of the field P-XRF measurements would be if the top few millimetres of the soil were systematically lower in Pb than the 150 mm sampled for both the laboratory P-XRF and the ICP-AES. To test this hypothesis a 62mm core of top of soil was taken, and progressively cut to remove eleven slices of around 6mm thickness. Each slice was analysed, both in-situ by P-XRF (Fig. 7.6) and also as a dried powder prepared from the slice by ICP-AES.

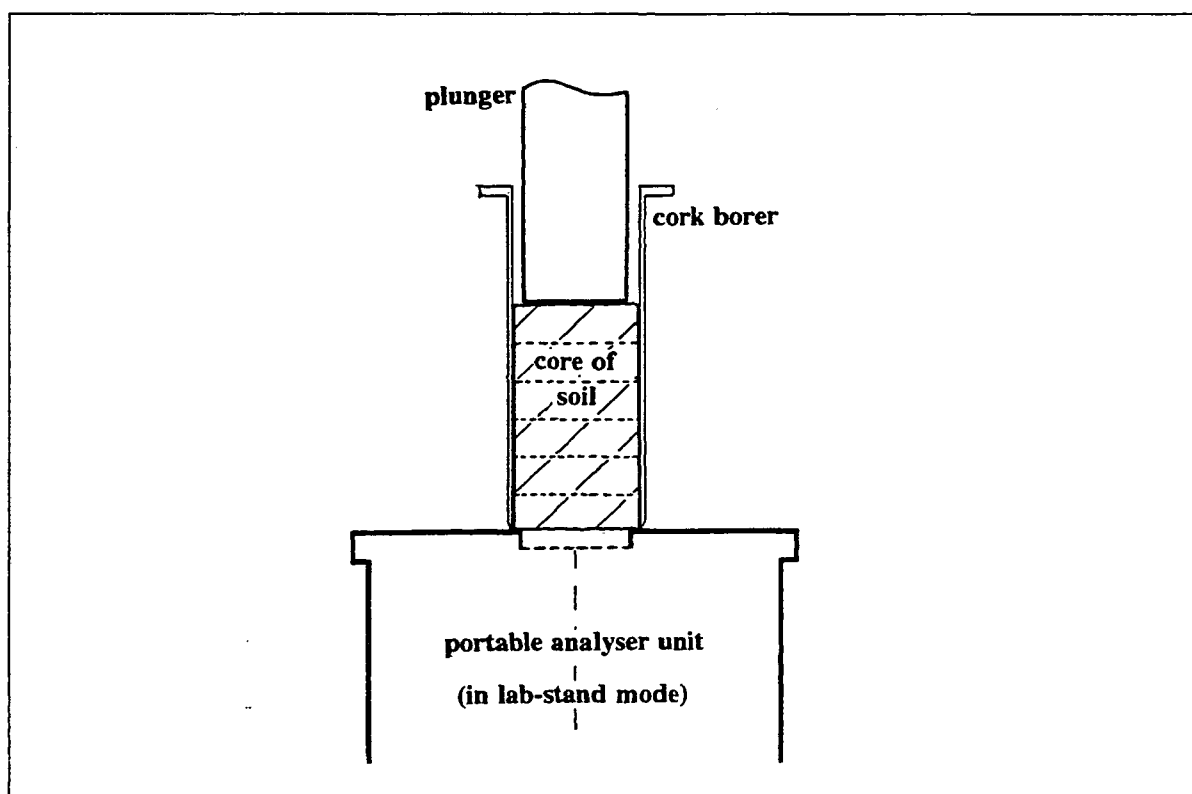


Figure 7.6. Schematic representation of the 'core experiment' set-up.

The location in the field for taking the core was selected to be close to the area with the highest measured metal concentrations. Measurements of the pellets of the reference materials (NBS 2710, HRM 31, HRM 7, GXR2) were taken before measuring the metal content in the core. Also a blank measurement was taken on the empty tube before the start of the experiment to ensure that the measured metal concentrations reflect the real metal content in the soil. The measuring times were 100 s for ^{109}Cd , 20 s for ^{55}Fe and 10 s for ^{241}Am .

The Pb profiles produced by this 'sliced core experiment' show little systematic spatial trend in this core (Fig. 7.7a), and no particular depletion in the top few millimetres, as would be required to confirm this hypothesis. A second core, taken 2m away from the first one, showed a different vertical variation in Pb concentration, but again no significant depletion at the surface (Fig. 7.7b). No systematic spatial trend also characterises the distribution of Zn in the cores (Fig. 7.7c,d).

Two other potential causes for the low 'bias' of the field P-XRF measurements are the moisture content of the soil and the rough surface of the soil used for excitation. These factors were corrected to assess their relative contribution to the bias. The moisture content of each slice of core experiment was determined and ranged from 6.83% to 19.09% m/m. When a correction was made for moisture to the field P-XRF measurements an initial bias of -19.9% against ICP-AES determinations for this particular experiment was reduced to -7.2%. When a Rayleigh scatter correction was also made to allow for the roughness of the soil surface, the bias was effectively eliminated, giving a non-significant bias of +0.8 % over the nineteen slices of both cores.

It would appear therefore that the cause of the apparent 'bias' of the field P-XRF measurements, at least for these two cores, is entirely due to the combined effects of soil moisture and surface roughness. In practice, the Rayleigh scatter correction can be applied routinely with minimal extra measurement time, but the determination of moisture correction will need to be either a retrospective correction, or require the deployment of a field probe.

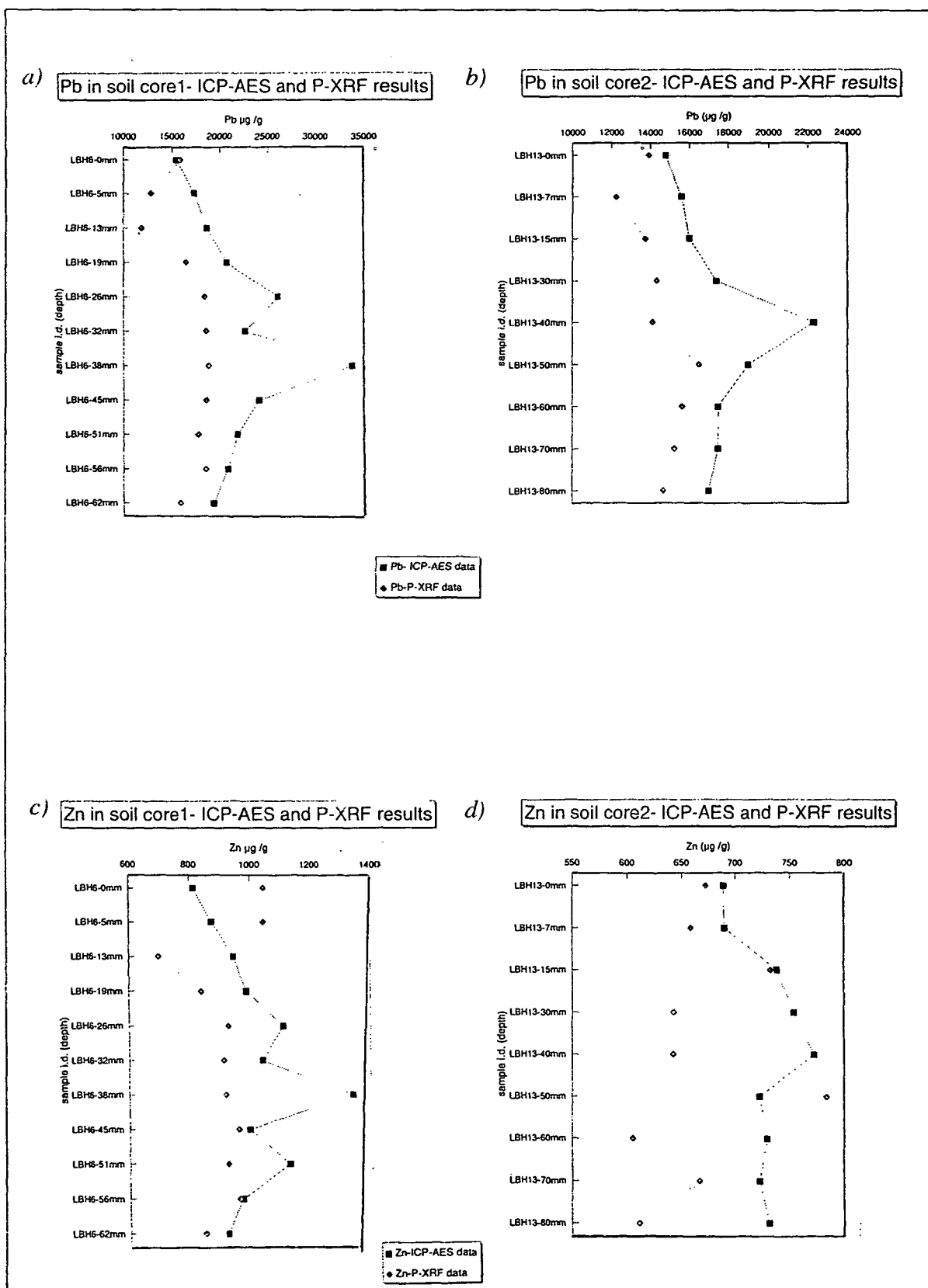


Figure 7.7. P-XRF core experiment: Pb in soil as determined in the cores by the two analytical methods. (a) core 1, (b) core 2. Zn in soil as determined in the cores; (c) core 1, (d) core 2.

7.4.4 Mapping of Contamination Using P-XRF

Real time spatial mapping of metal contamination across the sampling target is one possible significant advantage of using in-situ P-XRF. The measurements from the initial 30m grid indicated two 'hot spots' with Pb concentration of over $8000\mu\text{g g}^{-1}$ (Fig. 7.8). A further strength of the technique is that 'hot spots' can be investigated in more detail during the same site visit. In this case two 10m grids were set up over the candidate 'hot spots' from the initial 30m grid survey (Fig. 7.8). Interestingly the existence the eastern hot spot was substantiated by a cluster of similarly high Pb concentrations. The western 'hot spot', however, was not substantiated, with most of that 10m grid being lower than $6000\mu\text{g g}^{-1}$. Given the high measurement variance found at this site the extended measurement uncertainty was estimated to be $\pm 55\%$ of the mean by using the in-situ duplicate determinations. It would be expected therefore that occasional isolated high values will occur by chance, without being symptomatic of being taken from a large scale 'hot spot'.

The use of field P-XRF allows such 'spurious hot spots' to be eliminated at the time of the survey. Any bias in the measurements would not directly affect the investigation of this spatial coherence of the hot spots, but only affect their intensity and extent. Fitness-for-purpose criteria should be considered in order to determine the significance of any bias in the measurements. In this particular instance the estimated bias of -57% between the field P-XRF measurements and those made in the laboratory, does not affect the spatial delineation of the hot spots since the observed metal concentrations are well over the limits for contaminated land.

The spatial mapping on the smallest scale (i.e. $< 1\text{m}$) using the field P-XRF was used to investigate the heterogeneity of Pb on what is called the 'within sampling location' scale, in the bigger surveys. The square metre surrounding the eastern hot spot was surveyed using 16 samples, four taken at each of the distances from the centre of 50cm, 25cm, 12cm and 5cm (Fig. 7.9).

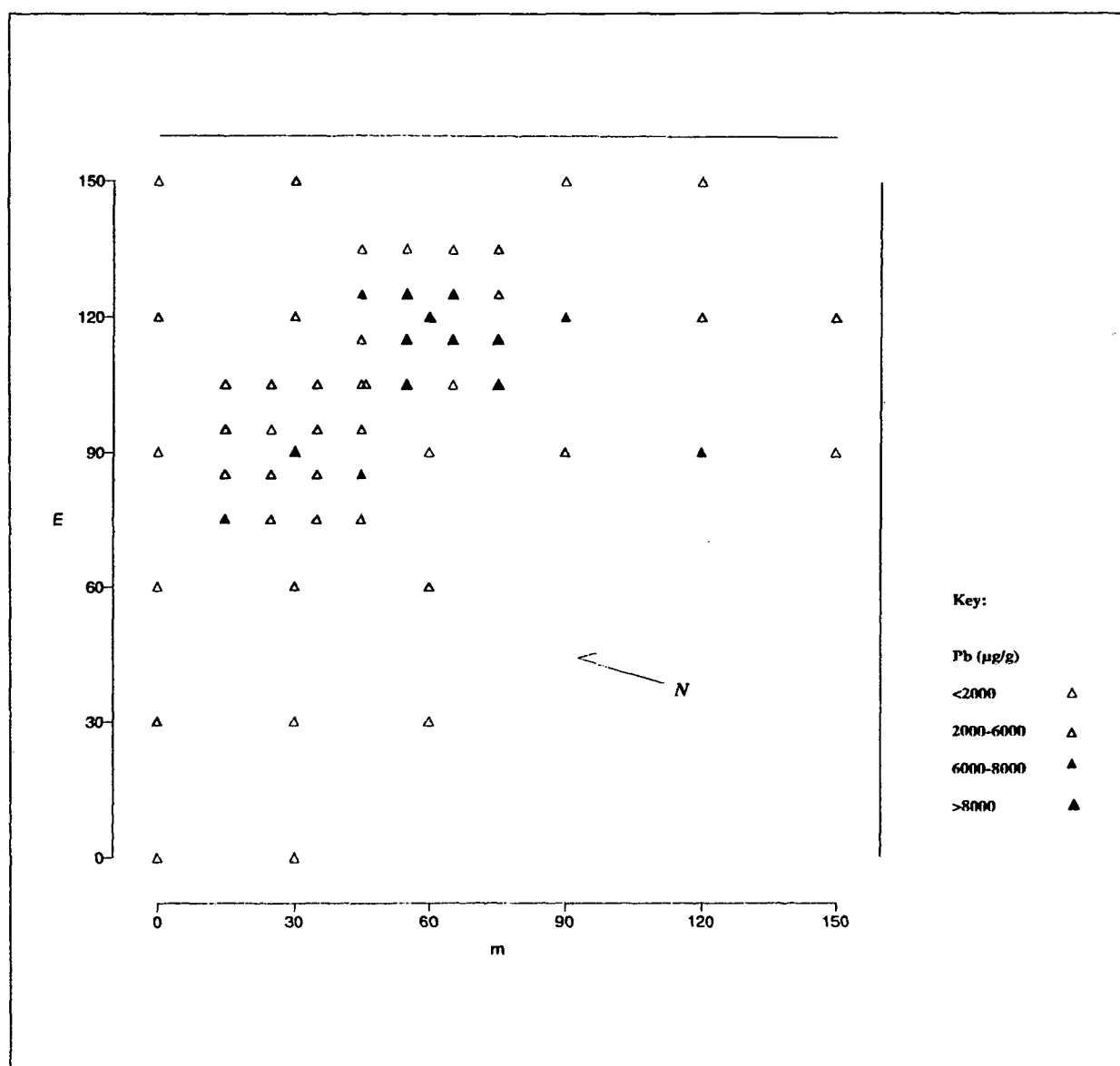


Figure 7.8. Spatial mapping of Pb concentration in soil determined by P-XRF, showing the iterative sampling design with initial 30 m grid followed by the 10 m grid centred on candidate 'hot spots' ($\text{Pb} > 8000 \mu\text{g g}^{-1}$). The eastern 'hot spot' is confirmed by a cluster of high values whereas the western 'hot spot' is not confirmed.

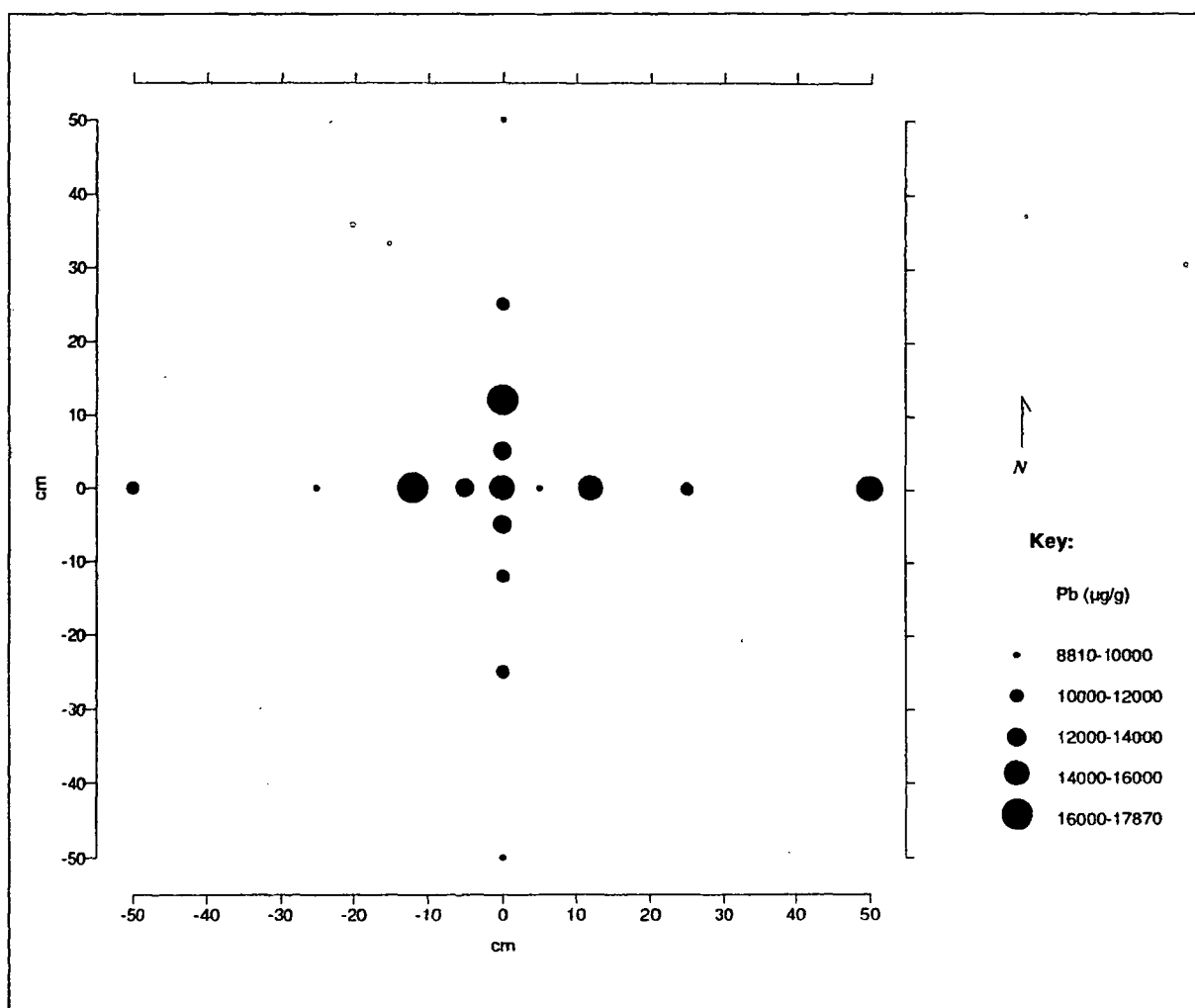


Figure 7.9. Concentration of Pb in soil in the 1 m² area over the contamination hot spot on the eastern side of the field, showing no systematic trend in the highest values.

The pattern of Pb concentrations measured does not conform to a single smooth maximum. Although high concentrations ($>14000 \mu\text{g g}^{-1}$ Pb) dominate the central area, there is one at the eastern edge, and also a low value ($8810 \mu\text{g g}^{-1}$) close to the centre. If the variance of the Pb concentration is estimated from the four replicates at each distance then there is not systematic change with distance (RSD = 28%, 24%, 32%, 18% for the four distances mentioned above). Furthermore, this variance within 1m is not significantly different from that measured at the 10m or 30m scales, when expressed in units relative to the mean value at the corresponding scale (RSD). Similar results were observed for Zn in the same square meter (RSD = 10%, 7%, 7%, 18%). It would be interesting to investigate the existence of a fractal dimension for the variance of the metals so that observations on a small scale could be used for the determination of variability and thus degree of

heterogeneity in a larger scale over the sampling target. However, this idea needs substantiating through further experimentation not attempted here.

A second square meter was selected at the west side of the field, downslope, where the metal concentrations in soil give the lowest values. In this square there is also no systematic change in the variance with scale neither for Pb or Zn (Fig. 7.10).

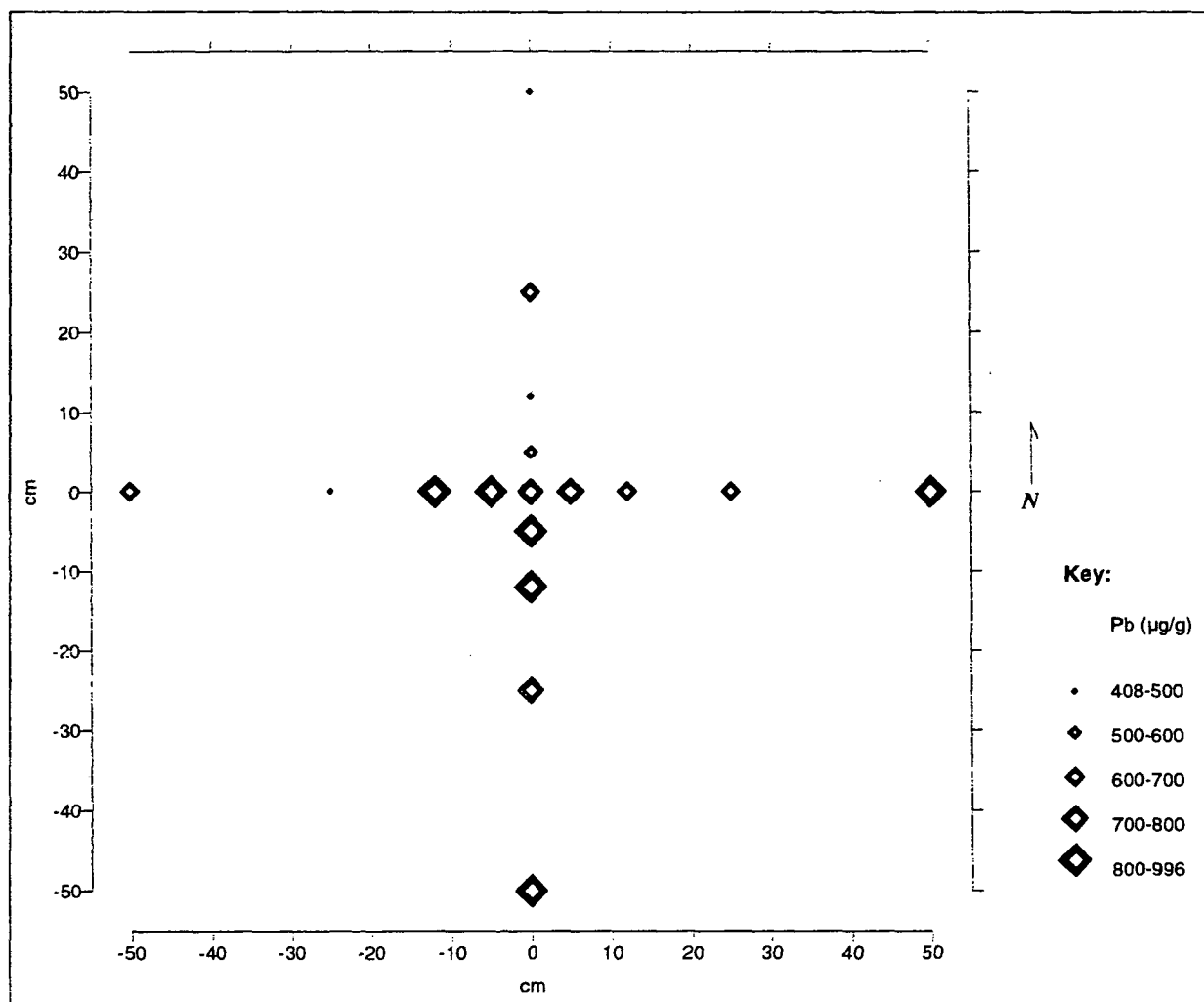


Figure 7.10. Concentration of Pb in soil in the 1 m² area on the western side of the field, where the lowest metal concentrations were observed, showing no systematic trend in the highest values.

7.5 Advantages and Disadvantages of the P-XRF Methodology

The potential of P-XRF for the in-situ investigation of metal contaminated land has been evaluated, showing the strengths and disadvantages of the methodology. The possibility of immediate estimates of metal contamination of soil by in-situ measurement

is the most important advantage of the method over previous studies which have used P-XRF with the soil removed to an on-site laboratory. Large quantities of data can be obtained in a timely and cost-efficient manner during a single visit to the sampling target.

Detection limits were low enough to measure elevated levels of contamination with sufficient precision for several metals of interest (e.g. Pb, Zn, and potentially Cu, Ni, As, Ba) in this particular field under study. The fact that the contaminants were present in high concentrations enhanced the abilities of the method but, since the P-XRF detection limits are well below the regulatory limits for most metals on contaminated land, the methodology is considered to be very useful for such investigations.

The rapid analysis time for a suite of several metals (e.g. 2 minutes) is an other advantage. However, the measurement time may need to be extended for sampling targets where elemental concentrations are lower than in this present study. Because of its ability for rapid analysis, P-XRF methodology facilitates the use of iterative sampling designs. For example a follow-up survey can be implemented to investigate any apparent 'hot spots', during a single site visit.

Satisfactory analytical quality was observed in the in-situ analysis, giving good analytical precision (e.g. RSD=1%). However the overall measurement precision is determined by the degree of heterogeneity of the sample site, effectively increasing the uncertainty of the analytical measurements. Another important advantage of the P-XRF methodology is the potential for the study of the heterogeneity of the sampling target at whatever scale this is considered. It provides a tool for the investigation of in-situ heterogeneity of the metal concentration across the site, down to the centimetre scale.

Disadvantages of the method include the very small sample size ($\sim 1 \text{ cm}^3$) which although useful for mapping heterogeneity, gives potentially poor representative. The high measurement variance that is a consequence, may be a small enough fraction of the total variance to be acceptable for a specified purpose, as is the case in this study. Also the small depth of analysis (1.5 mm) could be considered as a potential problem. This could be overcome by excavation at the site to allow in-situ analysis of samples at greater depth such as the analysis of drill cores.

A bias in the estimates of Pb and Zn concentrations was quantified in this study by comparison of the measurements with laboratory analysis of the corresponding samples using ICP-AES and the P-XRF in laboratory mode. The problem of apparent bias was

overcome in this instance by making corrections for soil moisture and surface roughness of the soil surface analysed. In general, recognition of the sources of bias is very important for the evaluation and improvement of the P-XRF methodology.

Detection limits are not low enough to quantify some elements (e.g. Cd) at typical background concentrations in soil. Also a slightly longer survey time was required on the site compared with that required for the simple taking of soil samples. The method gains credit however in the total time spend for analysis, since the time for sample preparation is minimised.

Finally, disadvantages related to the logistics of the use of the method include the relatively high capital cost of equipment (£ 50,000 at 1996 prices). Also the time and expense required for compliance with safety regulations involved with use of radioactive source is a potential problem that will vary between different countries.

When the P-XRF is used in laboratory mode with larger, dry and homogenised samples, many of these disadvantages are overcome, but at the expense of immediate measurement output and lack of information on in-situ variation in concentration.

7.6 Conclusions on the Use of P-XRF Methodology

The performance of the portable XRF Spectrace TN9000 used as an in-situ sampling and analytical technique has been evaluated. The technique used for the estimation of Pb and Zn concentrations in soil has proved to be a useful and powerful tool capable of giving precise and very rapid results.

The method was compared with established methods for the assessment of contaminated land, such as ICP-AES analysis of solutions derived from the soil samples and laboratory analysis of powdered samples using XRF. A rotational bias of -57% was calculated between the in-situ P-XRF and the ICP-AES measurements for Pb. The magnitude of bias between the in-situ P-XRF measurements and the P-XRF results in laboratory mode was also estimated to be -52% for the same element. These differences in the performance of the analytical technique were shown to be due largely to factors related to the nature of the samples. Specifically the soil moisture and the surface roughness of the samples are the two factors affecting the quality of the results. When a

correction was made for these factors at two locations within the field, the bias for Pb measurements was effectively reduced to a statistically insignificant level.

In spite the bias of the method, in-situ P-XRF measurements were capable of rapidly locating contamination hot spots in the field. Furthermore the detected contamination hot spots on the initial 30 m grid could be investigated during the same working day by setting up a more dense sampling grid around the areas with the highest metal concentrations. Bias in uncorrected field P-XRF results did not affect the spatial coherence of the hot spots but only affect their intensity and extend. Consequently the effect of the bias, if uncorrected, would be vital in the instance of metal concentrations being close to a regulatory limit.

A very important characteristic of the in-situ P-XRF methodology is the ability to investigate the small scale variability of metal content in soil. This is particularly useful in sampling targets showing high degree of heterogeneity in respect to the metal distributions. The degree of heterogeneity can easily be quantified by estimating the variability in different spatial scales within the field.

The limitations of the methodology at this particular site include the small effective sample size ($\sim 1 \text{ cm}^3$) and the small depth of analysis, that might result to poor representativity, the bias on the estimated concentration values and the relatively high cost of the equipment. The first two limitations can be overcome by taking measurements on a finer grid so as to observe the variability at a smaller scale, and by excavation at the site to increase the depth of the analysis. The recognition of the cause of the bias and the application of corrections of the soil moisture and surface roughness effect on the results was needed to eliminate the differences detected between the on-site and off-site analyses.

Further work is needed for the improvement of the performance of the method. The use of appropriate reference materials could be useful for the elimination of bias of the in-situ measurements. However it would be a difficult task to maintain materials simulating the soil in its natural state, i.e. containing enough moisture and with a matching grain size distribution. Comparisons of the concentration estimates with those from other established analytical methods are very important in the initial stages of development of the in-situ analytical method. Finally, good planning of the survey is needed before using the P-XRF for in-situ analysis so as to be able to gain as much

information as possible, on estimates of both concentration and uncertainties, to enable a realistic interpretation of the extend of contamination at the site.

CHAPTER EIGHT. ESTIMATION OF SAMPLING UNCERTAINTY AT A COMPLETED LANDFILL SITE IN WEST LONDON

8.1 Introduction

The suggested methodology for measuring sampling uncertainty was tested for a second time on a different contaminated site in West London. The objective of this part of the project was to validate the suggested methodology for the estimation of sampling uncertainty of the mean metal concentration in soil across a site. The properties of this new site were different from the characteristics of the old Pb smelting site in Derbyshire in respect to the levels and the source of contamination. Lower levels of heavy metal concentrations, closer to the regulatory limits, demonstrated the importance of measuring sampling uncertainty for the classification of contaminated land more clearly. The different sources of contamination at this landfill site had an effect on the degree of heterogeneity of the analyte concentrations, and therefore provided a way to test the general applicability of the methodology.

A reconnaissance survey followed by repetition of the inter-organisational sampling trials was set up in an area of Hounslow Heath, West London, UK. The results of these experiments are presented in the following paragraphs of this chapter.

8.2 Reconnaissance Survey

8.2.1 Site Description

The study area is a part of Hounslow Heath in West London (Fig. 8.1). The sampling area for the pilot study has dimensions of 270x270 m, a flat topography with high coarse grass vegetation and some trees. The geology of the site is generally simple. Hounslow Heath lies in the London Basin and comprises fluvial deposits of gravels and sands designated on the ^{geological} map as the Third River Terrace (^{British} Geological Survey, 1981)

overlying London Clay. The surficial heathland in the area is characterised by soils which are sandy, well drained, acidic and possibly podzolised to some extent (Yee, 1995).

Hounslow Heath has a long history of contamination. Initially, this was from military activity since the Heath was used for military training through the First World War and perhaps until the end of Second World War. Areas of the Heath have also been used as an aerodrome and rifle ranges. Sand and gravel excavation works on the heath began in the early 1950's and continued through the 1960's. The large gravel pits that were left behind after the workings were subsequently backfilled with waste materials. Records of the Greater London Council survey of sand and gravel workings indicated that backfilling activities were complete by the end of 1973 (Yee, 1995). The nature of the waste materials is not known, however it is believed they include domestic refuse material.

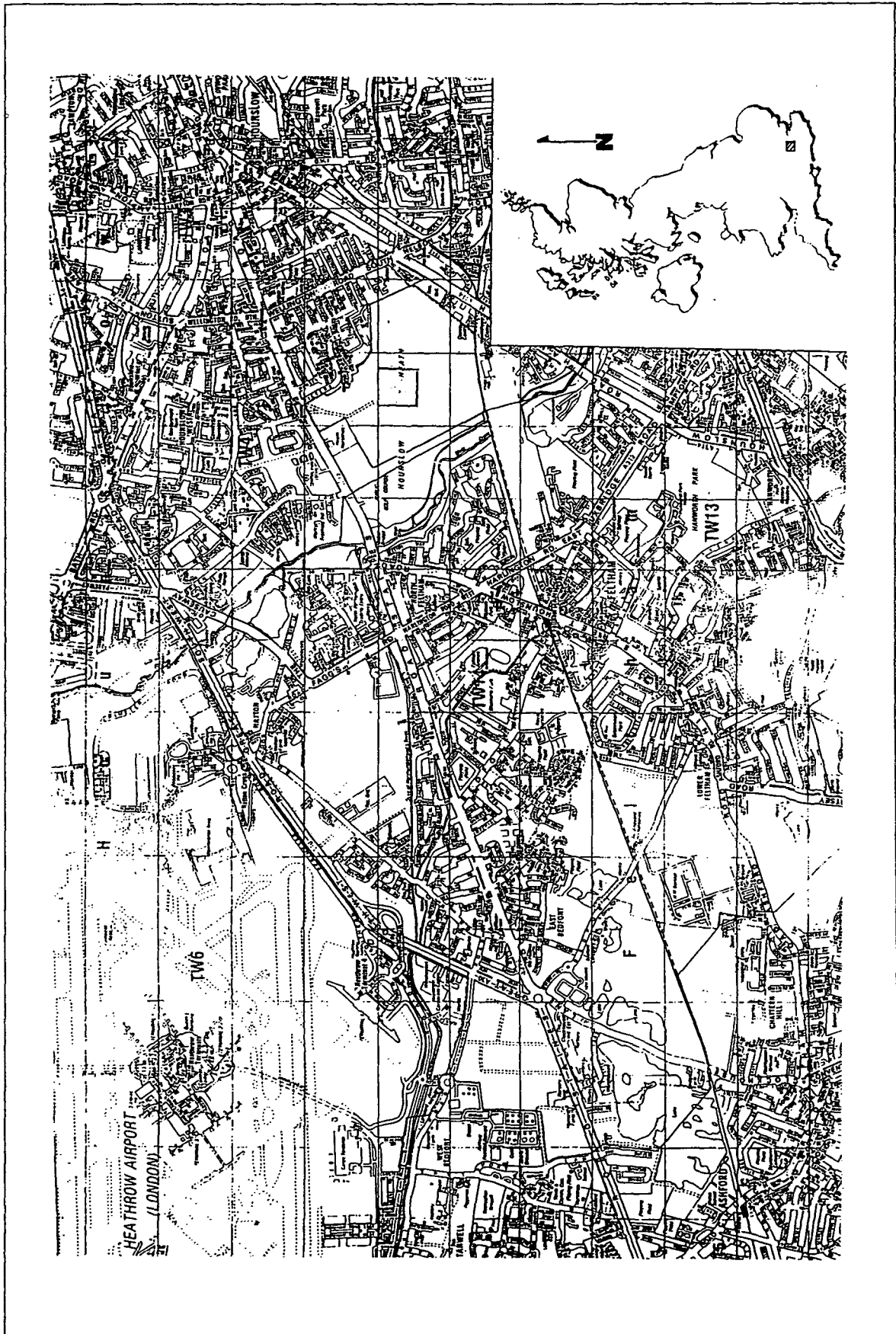


Figure 8.1. Geographical location of the study area in West London. ©

8.2.2 Sampling Strategy for the Pilot Study in Hounslow Heath

The objectives of the initial survey in Hounslow Heath were to determine the levels and the extent of contamination and to estimate the degree of heterogeneity of the soil in respect to metal concentrations. These features were considered essential for planning the inter-organisational trial in the site.

The sampling area for the pilot study was a 270x270 m square, located in the old landfill site in the heath. This area was selected taking into account the results of a previous survey that covered the whole heath area with a 100 m resolution regular grid (Yee, 1995). Areas with relatively high metal concentrations that could be distinguished on the 100x100 m grid were considered interesting for setting up the inter-organisational trial. After consideration of logistical factors, such as ease of proximity to the sampling area, topography and type of vegetation, the specific 270x270 m square was selected to be the sampling target.

A 30 m regular grid was set up in this area of 270x270 m, in August 1995 (Fig. 8.2) and top-soil samples (0-15 cm) were collected on the nodes of the grid using a hand auger. Each sampling point was located with a marking line made into a right angle isosceles triangle with a side of 30 m. The origin of the grid was selected to coincide with a sampling point of the 100 x 100 m grid. Its location was determined using a compass (120° N and 75 m away from a wooden post) and a metal tube was inserted into the ground for future reference.

Ten duplicate samples were collected for the estimation of sampling precision. The duplicate samples were taken 2 m away from the initial sampling location in a random direction (Ramsey, 1993). The distance between the initial sampling point and the location of the duplicate sample has been selected considering the location error for the used surveying technique.

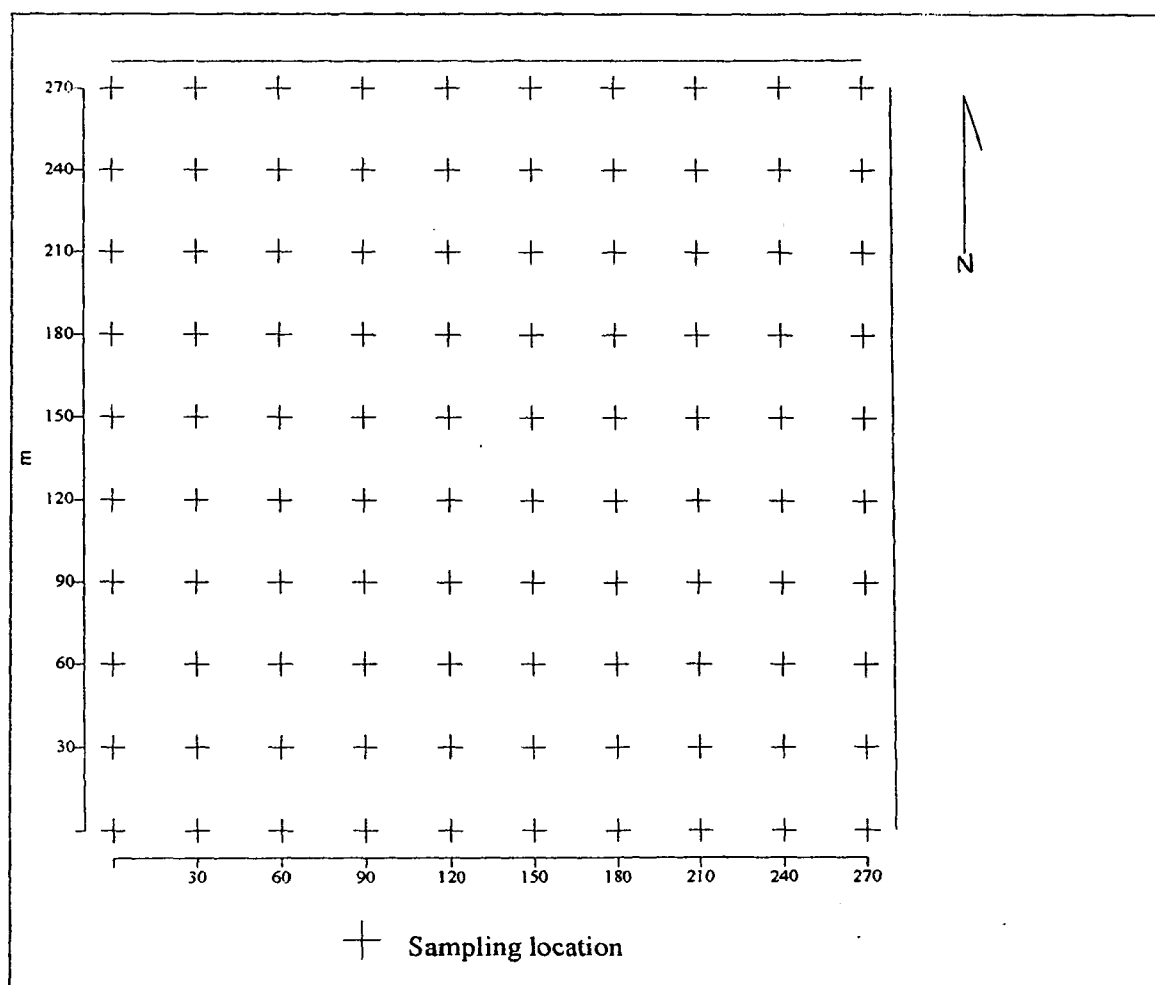


Figure 8.2. 30 x30 m grid for the pilot study in Hounslow Heath.

Test portions of all soil samples (0.100 g) were prepared for chemical analysis by ICP-AES as described in Chapter Three (section 3.3.2) for a suite of twenty-four elements (Thompson and Walsh, 1989). The raw data for all elements are given in Appendix 2 . Analytical duplicates were prepared for each sampling duplicate, for the estimation of sampling and analytical precision (Ramsey, 1993). Reagent blanks and six certified soil reference materials were analysed at random positions between the sample test portions for the estimation of analytical bias (Ramsey et al., 1987).

8.2.3 Statistical Characteristics of Pb and Zn Distributions

The interpretation of the analytical measurements focused on the study of two elements, Pb and Zn. These were selected as examples of heavy metal contaminants in the site suitable for target analytes in the subsequent inter-organisational trial.

8.2.3.1 Estimation of Random Errors from the Initial Survey at the Hounslow Heath Site

For the estimation of sampling and analytical random errors from this pilot study in Hounslow Heath, analysis of variance was applied on the data from the ten duplicate samples (full listing of the ANOVA results in Appendix 3). The robust mean Pb and Zn estimates as well as the measurement uncertainty on the data are presented in Table 8.1.

Table 8.1. Estimates of extended uncertainty ($k = 2$) for Pb and Zn in soil from the initial survey of part of Hounslow Heath.

Sampling Design	Mean Robust $\mu\text{g g}^{-1}$	s_{total} Robust $\mu\text{g g}^{-1}$	s_{meas} Robust $\mu\text{g g}^{-1}$	$U = 2s_{\text{meas}}$	$U\% = 200s_{\text{meas}}/\bar{x}$
Pb					
Regular Grid	297.3	218	124.3	248.6	83.6%
Zn					
Regular Grid	250.9	145	63.2	126.4	50.4%

The very high level of uncertainty (83.6%) reflects the extreme small-scale spatial variation of Pb contamination on the site. For Zn, the level of uncertainty is lower (50.4%) but still of a high percentage. However, such high levels of measurement uncertainty might be expected when the source of contamination is waste material from an old landfill site.

The pie charts for Pb and Zn, showing the proportions of the analytical, sampling and geochemical variances in the total variance are presented in Fig. 8.3. The calculations of the proportions were done using the robust estimates from the ANOVA.

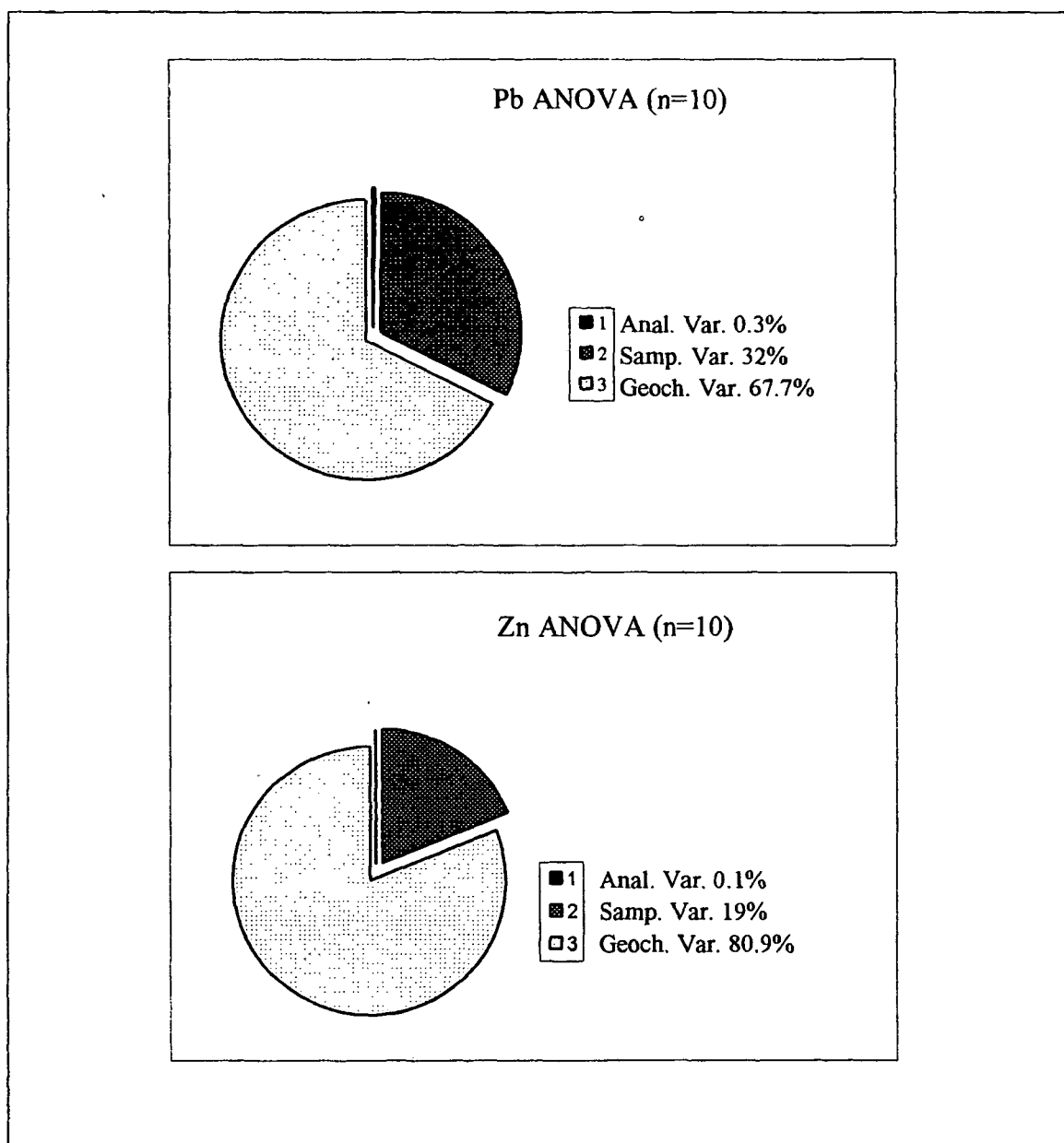


Figure 8.3. ANOVA pie charts for Pb and Zn in soil from the ten duplicate samples collected in the pilot study in Hounslow Heath.

The pie chart for Pb shows a low analytical variance of 0.3% and a great proportion of sampling variance (32%) over the total variance. This reflects the highly heterogeneous nature of the site in respect to Pb concentration in the soil. Also the relatively low geochemical variance of Pb (67.7%) contributes to the raise of sampling variability in the site. Any attempt at spatial interpretation of the Pb data at this site, with this particular sampling protocol would not be justified due to this high percentage of the sampling variance (over the 20% limit). However, the mean concentration results can be interpreted within their stated uncertainties. The estimate of random variability

for the sampling protocol used, also affects the fitness-for-purpose criterion for any subsequent sampling in this site. Given the high sampling variance for Pb, a denser sampling protocol is needed to give a clear picture of the element's distribution in the soil.

For Zn the proportions of analytical and sampling variance are within the 20% acceptable limit with percentages of 0.1% and 19% respectively. The sampling protocol could be used for spatial interpretation of the results in this instance.

8.2.3.2 Estimation of Systematic Errors from Analysis from the Initial Survey at the Hounslow Heath Site

The estimates of the systematic analytical errors for Pb and Zn were calculated as analytical bias by using six certified reference materials, namely BCR141, BCR142, BCR143, NBS2709, NBS2710 and NBS2711. The results of their analysis and the derived analytical bias are given in Table 8.2.

The levels of relative analytical bias are generally low (<10%) for both Pb and Zn. Statistical significant bias was only detected for two of the reference materials for Zn.

Table 8.2. Analytical bias for Pb and Zn concentration estimated from six CRMs ($\bar{x}$, mean concentration; s , standard deviation; and c , certified values in $\mu\text{g g}^{-1}$ unless otherwise stated). * denotes bias statistically significant with 95% confidence interval.

CRM ($n=2$)	$\bar{x}$	s	$s / \bar{x}$ (%)	c	Bias	Bias (%)
Pb						
NBS2709	24.9	2.1	8.4	18.9	6.0	32
NBS2710	5298.8	165.9	3.1	5532.0	-233.2	-4
NBS2711	1125.8	8.9	0.8	1162.0	-36.2	-3
BCR141	39.9	1.2	3.0	29.4	10.5	36
BCR142	39.6	4.5	11.4	37.8	1.8	5
BCR143	1279.5	44.4	3.5	1333.0	-53.5	-4
Zn						
NBS2709	99.8	0.2	0.2	106.0	-6.2*	-6*
NBS2710	6467.1	206.9	3.2	6952.0	-484.9	-7
NBS2711	342.1	4.1	1.2	350.4	-8.3	-2
BCR141	71.6	1.1	1.5	81.3	-9.7*	-12*
BCR142	86.2	2.8	3.3	92.4	-6.2	-7
BCR143	1253.2	36.0	2.9	1272.0	-18.8	-1

8.2.3.3 Statistical Characteristics of Pb and Zn Distributions

After examining the quality of the data the statistical interpretation of the analytical measurements was made.

The measured concentration of Pb is not very high with a mean of $292.1 \mu\text{g g}^{-1}$ and a maximum of $3590 \mu\text{g g}^{-1}$. The frequency distributions for Pb and Zn in the soil show a positive skew (Fig 8.4 (a), (b)).

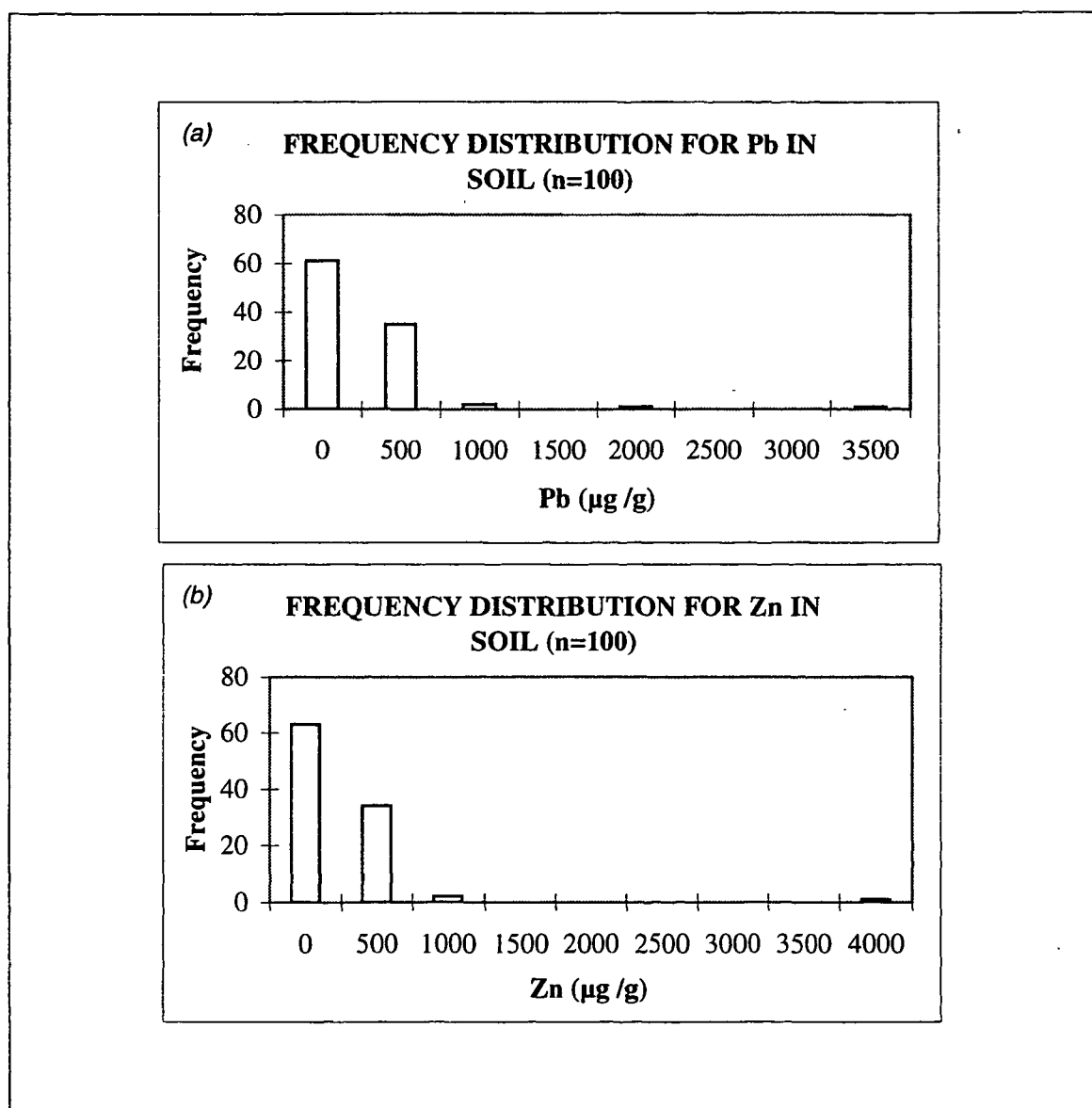


Figure 8.4. Frequency distributions for Pb (a) and Zn (b) from the samples of the pilot study in Hounslow Heath.

The mean concentration of Zn in the site was estimated to be $283.6 \mu\text{g g}^{-1}$, with a maximum of $3930 \mu\text{g g}^{-1}$. The descriptive statistics for Pb and Zn concentrations are summarised in Table 8.3:

Table 8.3 Descriptive statistics for Pb and Zn in soil in the initial survey of Hounslow Heath.

<i>Statistic</i>	<i>Pb</i>	<i>Zn</i>
Mean $\mu\text{g g}^{-1}$	292	284
Standard Error	40.3	40.1
Median $\mu\text{g g}^{-1}$	198	212
Mode $\mu\text{g g}^{-1}$	327	193
Standard Deviation $\mu\text{g g}^{-1}$	403.1	400.5
Sample Variance	162465.2	160418.1
Kurtosis	47.4	70.9
Skewness	6.3	7.9
Range	3534.5	3871.8
Minimum $\mu\text{g g}^{-1}$	55.5	58.2
Maximum $\mu\text{g g}^{-1}$	3590	3930

Dot maps showing the concentrations of Pb and Zn at each sampling location were produced (Fig. 8.5 (a), (b)). The concentrations to be plotted were classified taking into account not only the measured concentration, C , but also the measurement uncertainty estimates, U , calculated earlier (see Table 8.1). Four categories were established based on the overlap of the uncertainty with a single threshold level (T):

- 'Uncontaminated', where all the range of the uncertainty is lower than the threshold value ($C + U < T$),
- 'Possibly contaminated' where the upper side of the range of the uncertainty covers the threshold value ($C < T < C+U$),
- 'Probably contaminated', where the measured concentration exceeds the threshold value but the lower side of the range of uncertainty is below the threshold value ($C-U < T < C$),
- 'Contaminated', where the range of uncertainty is over the threshold value ($T < C+U$).

The concentration equal to the upper limit of the category 'Uncontaminated' ($C = T - U$, equivalent to $C + U = T$) is $82 \mu\text{g g}^{-1}$ and the lower limit of 'Contaminated' ($C = T + U$, equivalent to $C - U = T$) is $918 \mu\text{g g}^{-1}$. These concentration limit values are

calculated for 95% confidence, but changes to the coverage factor in the calculation of U% (see section 8.2.3.1) could be used to specify a different level for the acceptable level of misclassification.

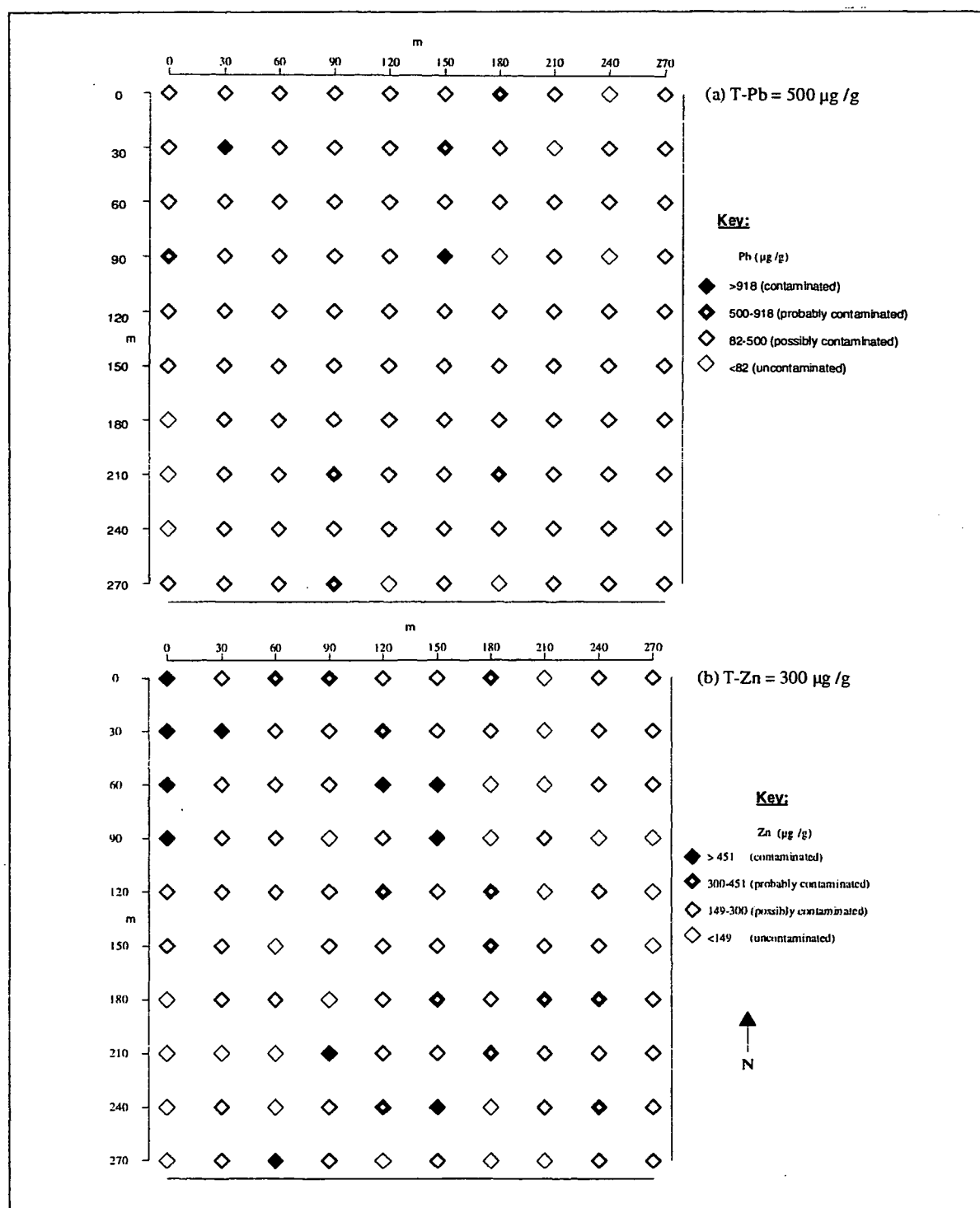


Figure 8.5. Dot maps of Pb (a) and Zn (b) concentrations in soil from the pilot study in Hounslow Heath, classified using the probabilistic approach that allows for measurement uncertainty.

From the maps for Pb and Zn it is apparent that the highest concentrations are located at the NW side of the area. This was a useful point taken into account for the selection of the area for the inter-organisational trial.

8.2.4 Conclusions from the pilot study in Hounslow Heath

From the initial survey in Hounslow Heath very useful information for the planning of the inter-organisational trial was gained. The levels of contamination and the degree of heterogeneity of the soil were also determined.

The study focused on two elements, Pb and Zn as examples of contaminants in the site. The concentration of Pb in soil was not very high with a mean of $292.1 \mu\text{g g}^{-1}$. However, high values of this element's concentration were observed at isolated sampling locations. The variability of Pb in soil was quite high, giving rise to high levels of measurement uncertainty of 83.6% for the 95% confidence interval.

Concentration of Zn was elevated with a mean of $283.6 \mu\text{g g}^{-1}$ which is close to the phytotoxic limit for this element in soil according to the ICRCL classification for soils ($300 \mu\text{g g}^{-1}$). The measurement uncertainty for this element was estimated to be 50.4%, a lower value compared to Pb but still relatively high.

The elevated values of measurement uncertainty are mainly due to the high degree of heterogeneity of the soil. This was expected since the sampling area was a part of the old landfill site. The sampling protocol chosen was unable to overcome this heterogeneity and hence the spatial interpretation of the data could not be justified.

For the inter-organisational trial that followed this pilot study a smaller area had to be selected, considering the time available for each participant to sample the site (one working day). The 100 x 100 m square on the NW corner of the site was judged to be the most appropriate area for the sampling trial because of the presence of the most interesting values of the studied elements. Also the origin of the grid which was set on the NW corner of the sampling area could be used as a reference point for any subsequent sampling at the site.

8.3 Second Inter-Organisational Sampling Trial on Contaminated Land, at Hounslow Heath

8.3.1 Introduction

The second sampling trial on contaminated land took place in a 100x100 m square on Hounslow Heath, part of the sampling area used in the pilot study. This sampling trial comprised two parts with different purposes: a collaborative trial in sampling (CTS) and a sampling proficiency test (SPT), similar to the trial at the Derbyshire site.

Nine organisations (five university departments, one local authority, two government laboratories and one commercial organisation) sent samplers to the site over the course of four months, from November 1995 until March 1996. The samplers were given information on the sampling trial beforehand and the exact protocol for the collaborative trial^{in sampling} immediately before commencing sampling. No sampler observed any other during the sampling exercise. The nine participants of this trial were: Ariadni Argyraki from *Imperial College, University of London*, Matt Hill from *University of Bradford*, Aradhana Mehra from *Derby University*, Joy Ræ from *Reading University*, Michael Thompson from *Birkbeck College, University of London*, Chris Dawes from *Walsall MBC*, Julian Trick from *British Geological Survey*, Jeffrey Jones from *British Gas*, and David Barry from *W.S. Atkins Environmental*.

All the collected samples were prepared for chemical analysis at Imperial College. Each aggregate sample was dried at 65 °C and broken down to individual grains with a pestle and mortar. The fraction passing a 2 mm plastic sieve was subsampled using a splitter to produce a 50 g sample. This sample was ground in a Tema mill to <100 µm to comprise the laboratory sample.

The test portions (0.1 g) were treated with a mixture of nitric and perchloric acids to solubilize the analytes which were presented for analysis in dilute hydrochloric acid solution (Thompson and Walsh, 1989). Twenty four elements were determined in the solutions by ICP-AES. Internal analytical quality control was provided by the inclusion of reagent blanks, six appropriate certified reference materials and duplicated analyses of samples.

The study of the results of the trial focused on four elements: Pb, Zn, Cd and Co. The first three of these elements showed high concentrations in soil, as expected because of the previous use of the land as a landfill site. Cobalt on the other hand had concentrations on the Heath that were within the normal range in soils (0.5 - 65 $\mu\text{g g}^{-1}$) (Alloway, 1995) and could be used for comparisons with the contaminating elements in the interpretation of the results.

8.3.1.1 Analytical Quality Control of the Data

For the quality control of the analytical determinations in the second sampling trial, six appropriate certified reference materials were used. The analytical quality control data are represented in Table 8.4.

Table 8.4. Analytical bias estimated from six CRMs for the sampling trial in Hounslow Heath. (ACC.V.=accepted value, t_{tab} = Student's t for 95 % confidence, st.si.= statistical significance).

Pb						
CRM	BCR141	BCR142	BCR143	NBS2709	NBS2710	NBS2711
n	2	2	2	2	2	2
$\bar{x}$ ($\mu\text{g g}^{-1}$)	29.58	34.63	1253.77	20.92	5071.95	1124.43
s ($\mu\text{g g}^{-1}$)	1.25	4.96	16.93	2	97.38	18.9
% RSD	4.2	14.3	1.35	9.56	1.92	1.68
ACC.V. ($\mu\text{g g}^{-1}$)	29.4	37.8	1333	18.9	5532	1162
BIAS ($\mu\text{g g}^{-1}$)	0.18	-3.17	-79.23	2.02	-460	-37.57
%BIAS	0.62	-8.38	-5.94	10.69	-8.32	-3.23
t_{calc}	0.203	0.904	6.618	1.428	6.68	2.811
t_{tab}	63.657	63.657	63.657	63.657	63.657	63.657
st.si.	N	N	N	N	N	N
Zn						
CRM	BCR141	BCR142	BCR143	NBS2709	NBS2710	NBS2711
n	2	2	2	2	2	2
$\bar{x}$ ($\mu\text{g g}^{-1}$)	70.04	84.27	1227.07	99.88	6143.39	340.18
s ($\mu\text{g g}^{-1}$)	1.36	3.78	19.69	0.85	101.83	6.80
% RSD	1.94	4.49	1.60	0.85	1.65	1.99
ACC.V. ($\mu\text{g g}^{-1}$)	81.3	92.4	1272	106	6952	350.4
BIAS ($\mu\text{g g}^{-1}$)	-11.25	-8.12	-44.93	-6.11	-808.60	-10.21
%BIAS	-13.84	-8.79	-3.53	-5.76	-11.63	-2.91
t_{calc}	-11.662	-3.034	-3.225	-10.112	-11.229	-2.125
t_{tab}	63.657	63.657	63.657	63.657	63.657	63.657
st.si.	N	N	N	N	N	N
Cd						
CRM	BCR141	BCR142	BCR143	NBS2709	NBS2710	NBS2711
n	2	2	2	2	2	2
$\bar{x}$ ($\mu\text{g g}^{-1}$)	1.26	1.23	31.82	1.24	20.77	41.50
s ($\mu\text{g g}^{-1}$)	0.03	0.21	0.52	0.27	0.15	1.68
% RSD	2.52	17.60	1.64	21.72	0.76	4.06
ACC.V. ($\mu\text{g g}^{-1}$)	0.36	0.25	31.1	0.38	21.8	41.7
BIAS ($\mu\text{g g}^{-1}$)	0.90	0.98	0.72	0.86	-1.02	-0.19
%BIAS	250.66	395.78	2.34	228.51	-4.70	-0.47
t_{calc}	40.03	6.41	1.96	4.52	-9.07	-0.16
t_{tab}	63.657	63.657	63.657	63.657	63.657	63.657
st.si.	N	N	N	N	N	N
Co						
CRM	BCR141	BCR142	BCR143	NBS2709	NBS2710	NBS2711
n	2	2	2	2	2	2
$\bar{x}$ ($\mu\text{g g}^{-1}$)	7.97	7.98	11.33	11.54	7.77	8.62
s ($\mu\text{g g}^{-1}$)	0.51	0.65	0.33	0.10	0.26	0.35
% RSD	6.49	8.25	2.92	0.92	3.39	4.10
ACC.V. ($\mu\text{g g}^{-1}$)	9.2	7.9	11.8	13.4	10*	10*
BIAS ($\mu\text{g g}^{-1}$)	-1.22	0.08	-0.46	-1.85	-2.22	-1.37
%BIAS	-13.36	1.08	-3.94	-13.85	-22.21	-13.70
t_{calc}	-3.359	0.1846	-1.986	-24.635	-11.881	-5.477
t_{tab}	63.657	63.657	63.657	63.657	63.657	63.657
st.si.	N	N	N	N	N	N

The relative analytical bias detected was less than 10% for most of the RMs for all the elements except Cd. The large bias in this latter case is probably due to the low concentrations compared with the detection limit for this element. In any case the fact that all the samples were analysed at random within one run makes any bias detected irrelevant to the objectives of this study.

8.3.2 Second Collaborative Trial in Sampling

The collaborative trial is a means of estimating the sampling uncertainty in the instance of different samplers sampling the same sampling target using the same sampling protocol. The within-sampler (repeatability) precision and the between-sampler (reproducibility) precision on metal concentrations are calculated based on an experimental design utilising analysis of variance (ANOVA).

The common sampling protocol for the nine participants used on Hounslow Heath was a regular grid, taking samples every 20 m (see Chapter Three, section 3.2.2.1). This protocol was applied in duplicate to produce two composite soil samples from the field. Subsequently the samples were analysed in duplicate in a single run and the resultant data were analysed statistically using nested ANOVA.

8.3.2.1 Statistical Interpretation of the Data from the Second CTS

The elemental concentrations determined in the laboratory samples are recorded in Table 8.5 and presented in Fig. 8.6. The statistics for the second CTS are given in Table 8.6.

Table 8.5. Estimated concentration values from the second collaborative trial in sampling. Concentrations in $\mu\text{g g}^{-1}$

Sampler	Sample	Analysis	Co	Pb	Zn	Cd
1	1	1	13.01	229.85	254.53	3.67
		2	15.58	224.28	265.66	3.40
	2	1	14.41	329.02	442.00	4.75
		2	14.64	340.09	443.43	5.52
2	1	1	13.01	1073.54	421.01	2.31
		2	13.44	1017.07	397.59	1.96
	2	1	14.30	313.60	442.52	5.54
		2	14.24	330.39	443.77	6.73
3	1	1	13.71	284.07	404.59	3.54
		2	14.12	359.05	418.38	3.95
	2	1	15.99	310.45	1040.72	2.23
		2	15.20	301.79	965.95	2.41
4	1	1	14.48	332.66	370.98	4.37
		2	15.60	305.75	400.02	4.26
	2	1	13.45	234.26	287.10	4.01
		2	12.75	232.72	292.22	4.08
5	1	1	15.62	308.25	403.22	6.85
		2	14.39	297.61	406.16	6.03
	2	1	15.50	482.41	436.74	5.36
		2	14.29	557.92	420.50	4.95
6	1	1	13.19	310.78	299.87	3.68
		2	12.52	262.28	300.11	3.68
	2	1	12.82	324.72	386.22	5.30
		2	13.79	326.33	399.75	5.44
7	1	1	12.90	285.36	310.79	3.24
		2	13.16	276.38	309.74	3.18
	2	1	13.39	268.80	283.18	3.20
		2	13.43	287.14	273.61	3.18
8	1	1	12.64	273.11	258.27	3.39
		2	12.80	470.74	271.04	3.07
	2	1	12.70	259.86	347.64	6.65
		2	12.66	257.10	362.41	6.09
9	1	1	14.38	413.06	434.63	6.20
		2	14.55	607.74	616.42	8.04
	2	1	13.80	283.17	354.48	7.22
		2	13.38	309.29	348.83	7.51

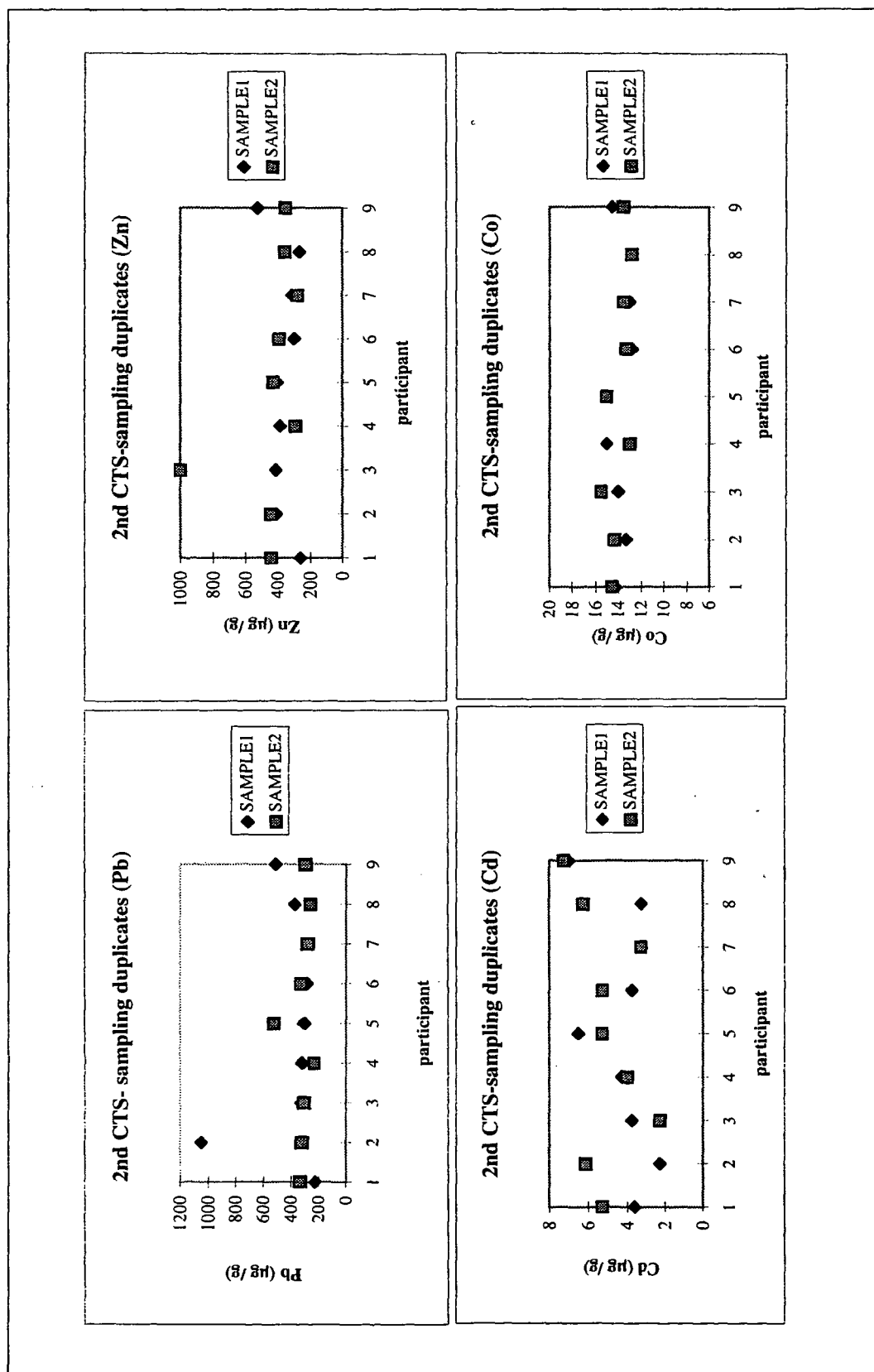


Figure 8.6. Graphical presentation of the results from the second CTS in Hounslow Heath. Each point shows the mean of the duplicated analytical determinations.

Table 8.6. Statistics from the second collaborative trial in sampling in Hounslow Heath.

	Co		Pb		Zn		Cd	
<i>Statistic</i>	<i>Classical</i>	<i>Robust</i>	<i>Classical</i>	<i>Robust</i>	<i>Classical</i>	<i>Robust</i>	<i>Classical</i>	<i>Robust</i>
Grand mean $\mu\text{g g}^{-1}$	13.9	13.9	363	336	403	378	4.6	4.5
s_0 $\mu\text{g g}^{-1}$	0.6	0.6	52	28	34	11	0.5	0.3
$s_1 = s_{r(s)}$ $\mu\text{g g}^{-1}$	0.5	0.6	186	114	155	98	1.4	1.3
s_2 $\mu\text{g g}^{-1}$	0.6	0.7	0	0	58	22	0.8	0.7
$s_{R(s)}$ $\mu\text{g g}^{-1}$	0.78	0.92	186.00	114.00	165.50	100.44	1.61	1.48
%RSD ₀	4.32	4.32	14.33	8.33	8.44	2.91	10.87	6.67
%RSD _{r(s)}	3.60	4.32	51.24	33.93	38.46	25.93	30.43	28.89
%RSD _{R(s)}	5.62	6.63	51.24	33.93	41.07	26.57	35.05	32.81

The data for Co suggest that this element is evenly distributed across the site. The sampling standard deviations are comparable with the analytical error. The differences between samplers are greater than the variations in the performance of a single sampler. This is described statistically with the sampling reproducibility standard deviation, $s_{R(s)} = 0.78 \mu\text{g g}^{-1}$, being greater than the sampling repeatability standard deviation, $s_{r(s)} = 0.5$

$\mu\text{g g}^{-1}$ (using the classical estimates as described in Chapter Five, section 5.3.1). The sampling relative standard deviations are reasonably small (4.3% within and 5.62 % between samplers). The sampling protocol could be regarded as fit for purpose.

On the other hand the results for Pb reflect extreme variations in its concentration. The analytical precision is reasonably good (RSD₀ = 14 %) but variations between duplicate samples are so great (RSD_{r(s)} = 51 %) that obscure the difference between the samplers. In this instance the sampling protocol is not fit-for-purpose using a criterion of 20 % due to the great heterogeneity of the site in respect to the Pb concentration.

The Zn and Cd data reflect boundary conditions with very small differences between the between-sampler and within-sampler variations. The heterogeneity of the field in respect to the concentrations of these elements is just on the limit for judging the sampling protocol as fit-for-purpose by this criterion.

8.3.2.2 Estimation of Measurement Uncertainty Using the Data from the Second CTS

The data from the collaborative trial were used for the estimation of measurement uncertainty according to the equation $\% U = 200 s / \text{grand mean}$ (Table 8.7).

Table 8.7. Standard deviations and expanded uncertainties for Pb, Zn, Cd and Co in soil, assuming coverage factor $k=2$. Data from the second CTS.

Element				Pb			Zn		
Grand mean				336			378		
$\mu\text{g g}^{-1}$									
Uncertainty	s	$\mu\text{g g}^{-1}$	U	$\mu\text{g g}^{-1}$	U%		s	$\mu\text{g g}^{-1}$	U
analytical	28		56		16		11		22
single sampler	114		228		68		98		196
multiple sampler	114		228		68		100		200
Element				Cd			Co		
Grand mean				4.5			13.9		
$\mu\text{g g}^{-1}$									
Uncertainty	s	$\mu\text{g g}^{-1}$	U	$\mu\text{g g}^{-1}$	U%		s	$\mu\text{g g}^{-1}$	U
analytical	0.3		0.6		13		0.6		1.2
single sampler	0.7		1.4		31		0.6		1.2
multiple sampler	1.5		3		67		0.9		1.8

Expanded uncertainties were calculated on the basis of the duplicate analyses and duplicate samples by multiplying the adequate standard deviations by a coverage factor of 2. The robust estimates from ANOVA were used in this instance to reduce the disproportional effect on the standard deviations caused by great differences between analyses. From the calculated components of uncertainty for Pb it is apparent that the dominant factor of the uncertainty for this element is the within-sampler variance. This means that any between-sampler variations are obscured because of the high within-sampler standard deviation. The main reason for this is probably the high degree of heterogeneity in the soil.

For all the other analytes the between-sampler variance exceeds the value for the within-sampler variance. These results are comparable with the results from analytical collaborative trials. However, the levels of uncertainty are still relatively high for Zn (U

= 53 %) and Cd (U = 67 %). The lowest measurement uncertainty was calculated for Co (U = 13 %). It must be noted that for this element the within-sampler uncertainty does not contribute any extra percentage to the analytical uncertainty.

8.3.3 Second Proficiency Test in Sampling

The Second proficiency test took place almost simultaneously with the collaborative trial. Each of the participants collected the soil samples on the same day as the collaborative trial experiment but just before the execution of the common sampling protocol. The particular objective given to the participants was to design and implement a sampling protocol for the estimation of the mean metal concentration in the top soil at the site. The analysis of all the collected samples from this part of the trial was done in Imperial College, in the same batch for the elimination of analytical bias. The interpretation of the results focused on the same four elements with the results from the collaborative trial (Pb, Zn, Cd, Co).

8.3.3.1 Sampling Target and Sampling Procedures

The second SPT took place on the same sampling target as the collaborative trial. The different sampling protocols used by the participants in the sampling proficiency test on Hounslow Heath are presented in Fig. 8.7. Different types of augers were used to collect soil and the depth of sampling varied from 0-10 cm to 0-20 cm.

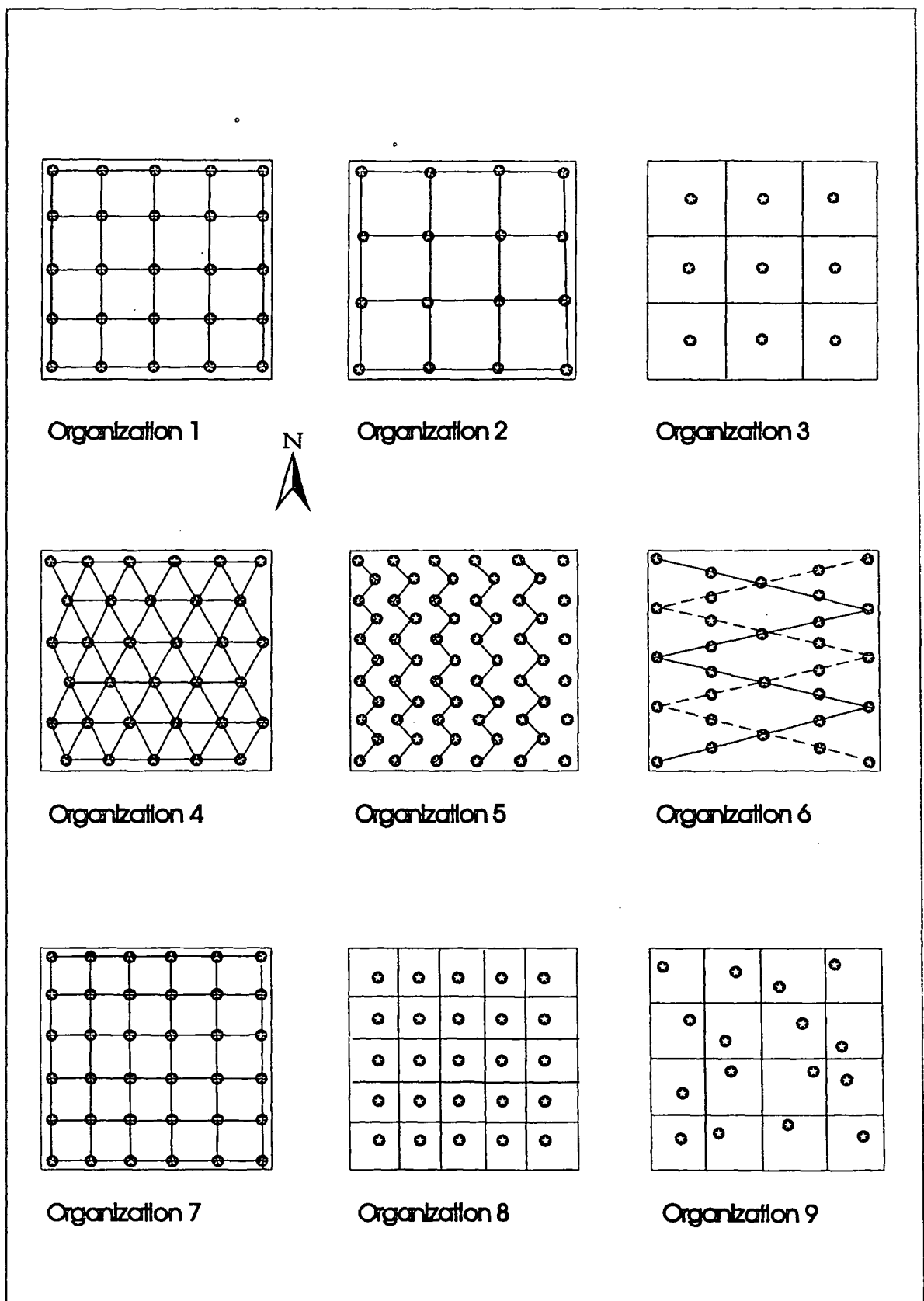


Figure 8.7. Graphic presentation of the nine different sampling designs used by the participants to sample top soil in the square area for the second proficiency test.

Details on the sampling procedures used by the participants are presented in Table 8.8.

Table 8.8. Sampling procedures used by the nine organisations for the second proficiency test.

Participant	No. of sample increments	Sampling equipment	Sampling technique
1	25	Cylindrical auger	Remove top turf vegetation. Single auger increments. Depth, 0-11.5 cm
2	16	Auger-5 cm diameter	Mix 2 auger samples 1m away at each sampling location. Depth, 0-15 cm
3	9	Bulb planter	Collect top 0-6 cm of soil. Cut off top 2 cm.
4	33	Auger-2.5 cm diameter	Collect single auger samples at each sampling location. Depth, 0-15 cm
5	61	Cylindrical auger	Remove top vegetation with trowel. Collect single auger sample increments. Depth, 0-15 cm
6	34	Auger 2.5 cm diameter	Collect single auger samples on the legs of 2 'W'. Depth 0-10 cm
7	36	Auger 2.5 cm diameter	Collect single auger samples at each sampling location. Depth, 0-15 cm
8	25	Trowel	Dig out top vegetation with spade. Collect sample increments with trowel. Depth, 0-20 cm
9	16	Auger 2.5 cm diameter	Remove top vegetation. Collect sample increments in separate pots- mix in lab to make a composite sample. Depth, 0-15 cm

8.3.3.2 Statistical Interpretation of the Results from the Second SPT

The analytical results are presented graphically in Fig. 8.8. The values from the analysis, the bias from the grand mean and the calculated z-scores are presented in Table 8.9.

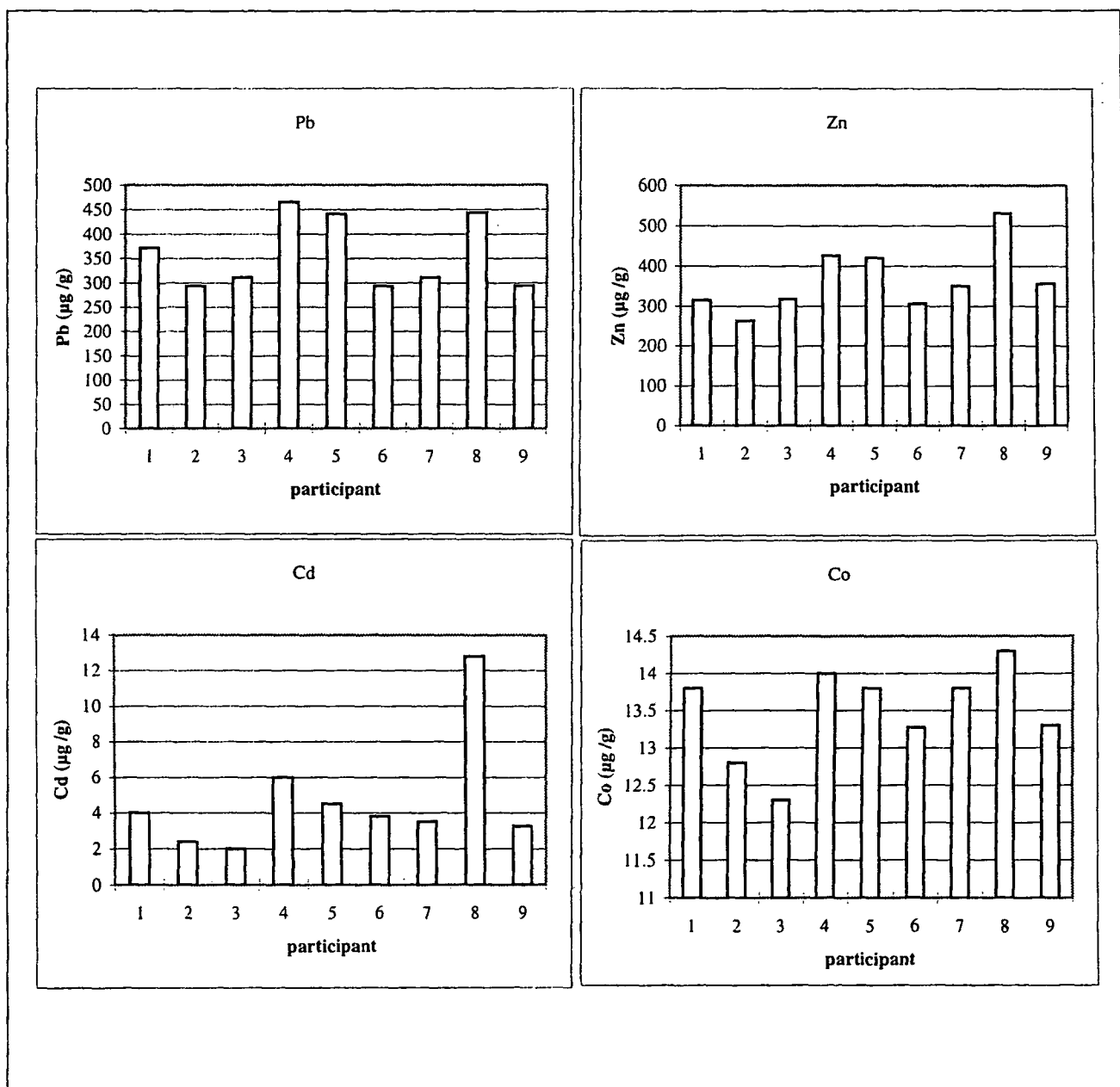


Figure 8.8. Graphic representation of the analytical results from the second SPT.

Table 8.9. Analytical results from the second SPT.

Element Pb				Zn		
Participant	Concentration $\mu\text{g g}^{-1}$	% bias	z-score	Concentration $\mu\text{g g}^{-1}$	% bias	z-score
1	372	3.9	0.20	314	-14.0	-0.70
2	294	-17.9	-0.89	263	-27.9	-1.40
3	311	-13.1	-0.66	317	-13.2	-0.66
4	465	29.9	1.49	426	16.7	0.84
5	441	23.2	1.16	420	15.1	0.75
6	293	-18.2	-0.91	306	-16.2	-0.81
7	311	-13.1	-0.66	350	-4.1	-0.21
8	444	24.0	1.20	531	45.5	2.28
9	294	-17.9	-0.89	356	-2.5	-0.12
Mean	358			365		
St.Dev.	73			82		
Target	71.7			72.9		
St.Dev.						

Element Cd				Co		
Participant	Concentration $\mu\text{g g}^{-1}$	% bias	z-score	Concentration $\mu\text{g g}^{-1}$	% bias	z-score
1	4	-18.4	-0.74	13.8	2.2	0.11
2	2.4	-51.0	-2.45	12.8	-5.2	-0.26
3	2	-59.2	-2.87	12.3	-8.9	-0.44
4	6	22.4	1.38	14	3.7	0.19
5	4.5	-8.2	-0.21	13.8	2.2	0.11
6	3.8	-22.4	-0.96	13.3	-1.5	-0.07
7	3.5	-28.6	-1.28	13.8	2.2	0.11
8	12.8	161.2	8.62	14.3	5.9	0.30
9	3.3	-32.7	-1.49	13.3	-1.5	-0.07
Mean	4.7			13.5		
St.Dev.	3.3			0.6		
Target	0.94			2.69		
St.Dev.						

Good agreement between the participants can be observed for Co with maximum bias for participant number 3 of -8.9% from the grand mean. For Pb all the concentration estimates are below the ICRCL trigger concentration for gardens and allotments of $500\mu\text{g g}^{-1}$. The results for Zn and Cd range from lower to higher than the regulatory limits ($300\mu\text{g g}^{-1}$ and $5\mu\text{g g}^{-1}$ respectively). For Zn only participant number 2 gave a result lower than the phytotoxic concentration limit of $300\mu\text{g g}^{-1}$. For Cd only participant number 8 gave a concentration estimate higher than the regulatory limit of $5\mu\text{g g}^{-1}$.

The estimation of bias between participants that is proposed by the harmonised protocol uses z-scores (Thompson and Wood, 1993). This, requires the definition of an assigned value of concentration, X , and a target standard deviation, s . The assigned values were calculated to be $13.5 \mu\text{g g}^{-1}$ for Co, $358 \mu\text{g g}^{-1}$ for Pb, $365 \mu\text{g g}^{-1}$ for Zn and $4.7 \mu\text{g g}^{-1}$ for Cd. These values are the grand means of the metal concentrations in soil from all the participants. There were the best estimates of the 'true' values since the 'consensus of experts' suggested by the harmonised protocol was unavailable to be used as the assigned value. The target standard deviation, s , was arbitrarily set at 20% of the assigned value based on a broad estimate of fitness for purpose (Table 8.9).

The raw data were converted to z-scores recommended in the International Harmonised Protocol. The z-score is given by $z = (x - X)/s$, where x is the participants reported result, X the assigned value and s the target value for standard deviation. The diagrams for the z-scores are shown in Fig. 8.9. For most of the elements z-scores are within the ± 2 range except for participant number 8 with high z-scores for Cd and Zn. However, it should be noted that a z-score outside the ± 2 range does not necessarily mean bad performance because of the high heterogeneity conditions in this particular site.

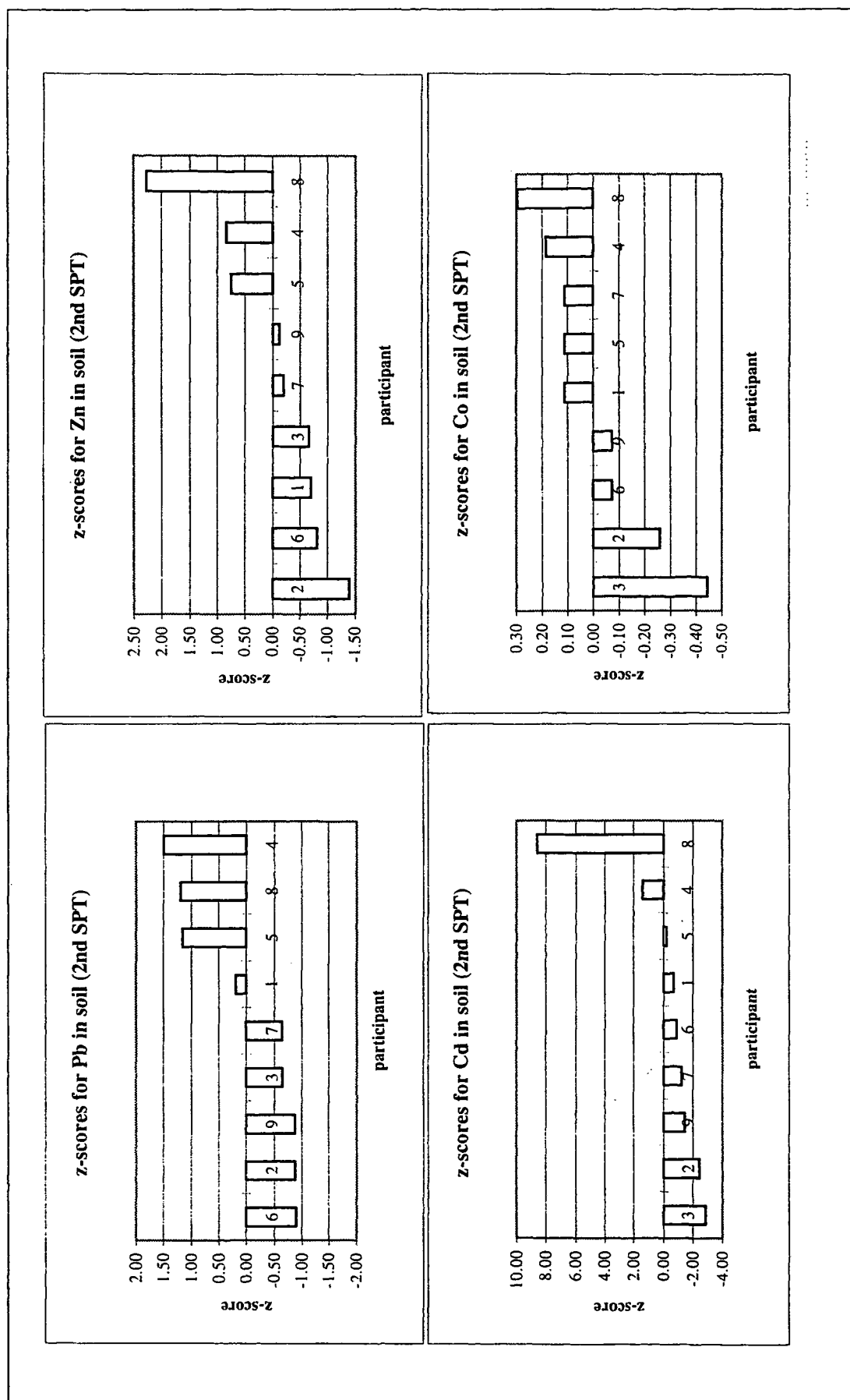


Figure 8.9. z-scores from the second sampling proficiency test in Hounslow Heath.

8.3.3.3 Estimation of Measurement Uncertainty Using the Second SPT Data

The analysis of all the samples for the second SPT was completed in one laboratory. In this way any effect of analytical errors on the results of the trial was eliminated. The assessment of the results gave an estimate of the sampling uncertainty generated when the different organisations sample the site using different sampling protocols. The uncertainty was estimated as random error from the nine participants and was expressed as standard deviation of the nine values for the estimated mean metal concentration in the sampling site.

The resulting scatter in the estimates of the mean concentration of Pb gives an estimate of expanded uncertainty of 41%. This is calculated as $\%U = 200 s / x_{\text{grand}}$, where x_{grand} is the grand mean of the reported results from the second proficiency test and s the standard deviation. For Zn the uncertainty estimate is 45 % for Cd 140 % and for Co 9 %. These estimates of uncertainty are lower than the ones estimated by using the CTS data (values 68%, 53%, 67%, 13% respectively), except for Cd. However, if the value for this element for participant 8 is removed as an outlier, the estimated uncertainty gives a percentage of 41 %, which is lower than the uncertainty value estimated from the CTS.

8.3.4 Conclusions from the Second Sampling Trial in Hounslow Heath

The second sampling trial was successful in showing that the methodology of sampling trials first applied to a rural site in Derbyshire heavily contaminated by smelting are equally applicable to a completed landfill site in Hounslow Heath with different distribution and intensity of metal contamination.

The extremely heterogeneous nature of the Pb contamination at this site is a limiting factor in the measurement uncertainty that was achievable in the collaborative trial (68 % for Pb at 95% confidence). Lower values of measurement uncertainty but still of a high percentage were estimated for Zn (53%) and Cd (67%). For Co which was more evenly distributed in the site the percentage of the uncertainty was 13%.

The 20 m regular grid chosen for the protocol of the collaborative trial is not fit-for-purpose of estimating mean concentrations (to within 20%) for Pb, Zn, and Cd, but is fit-for-purpose for Co, at this site. In general more dense sampling is needed at sites with such high degree of heterogeneity as at an old landfill. Obviously the effectiveness of a sampling protocol can be assessed using the CTS only for the specific site and further experimentation is needed before any attempt for generalisation about the potential suitability of particular protocols at contaminated sites.

The general level of agreement between the participants in the proficiency test can be considered satisfactory. The levels of uncertainty estimated by this test were lower than these from the CTS. This may be due in part to the more appropriate designs chosen by the participants in the SPT. The estimated values of extended uncertainty were 41%, 45%, 41%, 9% for Pb, Zn, Cd and Co respectively.

Z-scores in the SPT reflect the ability of the participants for effective sampling as well as the effect of heterogeneity. More detailed work is needed to allow for the heterogeneity in the assigned values and the target standard deviation used in z-scoring in order to be able to assess better the performance of the participants.

Finally useful information on the estimation of sampling uncertainty could be gained by comparison of the results of the sampling trials in Derbyshire and in Hounslow, discussed in the following chapter.

CHAPTER NINE. DISCUSSION OF ESTIMATING MEASUREMENT UNCERTAINTY FROM SAMPLING

9.1 Introduction

The general discussion on the suggested methodology for estimating measurement uncertainty in the sampling of contaminated land is presented in this chapter. The comparison of the results from the two inter-organisational trials, carried out in this project is made. Also matters related to the vertical variation of the contamination, as well as a geostatistical view of the problem, have been considered. The application of the suggested methodology for a probabilistic classification of contaminated land has been described and some ideas on the estimation of spatial uncertainty have been presented. Finally the potential of the methodology to be applied to the sampling of different environmental media has been discussed.

9.2 Comparison of the Results from the Two Inter-organisational Sampling Trials on Contaminated Land

The two inter-organisational sampling trials were conducted for the purpose of showing the general feasibility of the approach, on two sites with different contamination characteristics, i.e. on an area of land contaminated by smelting in Derbyshire and on a landfill site with lower level of contamination in West London. Two types of trials were tested at both sites: a collaborative trial in sampling (CTS) and a sampling proficiency test (SPT). The statistical analysis of the results of these trials focused on elements with contrasting characteristics in respect to their concentrations in top soil. Lead and Cu were the elements selected for the first sampling trial in Derbyshire, with the first having high concentration in top soil and the latter been closer to the natural background. For the second trial in Hounslow Heath all Pb, Zn and Cd showed elevated levels of concentration in the soil and Co was the selected background element for the comparisons.

9.2.1 Comparison of the Results from the Two CTS

The collaborative trial is traditionally the final stage in determining the performance characteristics of an analytical method (Horwitz, 1988). In this project two collaborative trials in sampling were organised to test the feasibility of the method in the sampling of contaminated land. The validation of the sampling protocols used in these trials and the assessment of the method for the estimation of measurement uncertainty in the sampling of contaminated land were the objectives of the trials. The comparison of their results discussed below was based on the following considerations:

1. The performance of the sampling protocols used for the estimation of the mean metal concentration in top soil, in relation to the estimated value of the uncertainty,
2. The susceptibility of the sampling protocols to the heterogeneity in respect to the metal concentrations in soil, and
3. The potential applicability of the method in sites with different contamination characteristics.

The sampling protocol used by all the participants for the first CTS was a 'W' shaped design collecting the sample increments along the vertices. In the second CTS the common sampling protocol was a regular grid, collecting the sample increments every 20 m. The selection of the sampling protocols was based on several factors listed below:

1. the simplicity and ease to execute,
2. the frequency of use for the purpose of soil sampling in contaminated land investigations in general,
3. preliminary information on the degree of heterogeneity at the sampling targets.

It should be noted here that there is no intention in this project to suggest an optimum sampling design. Sampling protocols commonly used, were tested in respect to their performance characteristics for the particular purpose of sampling top soil.

The estimates of uncertainty for Pb as an example of an elevated element and Cu, Co as examples of background elements for the first and second CTS respectively are presented in Table 9.1:

Table 9.1. Expanded uncertainties for Pb, Cu, Co in soil from the two CTS, assuming coverage factor $k = 2$.

<i>1st CTS (Derbyshire)</i>				
Element	Pb		Cu	
Grand mean $\mu\text{g g}^{-1}$	6575		30	
Uncertainty	U $\mu\text{g g}^{-1}$	U%	U $\mu\text{g g}^{-1}$	U%
analytical	490	7.4	2.2	7.3
single sampler	3366	50.9	2.8	9.3
multiple sampler	3366	50.9	4.4	14.7
<i>2nd CTS (London)</i>				
Element	Pb		Co	
Grand mean $\mu\text{g g}^{-1}$	336		13.9	
Uncertainty	U $\mu\text{g g}^{-1}$	U%	U $\mu\text{g g}^{-1}$	U%
analytical	56	16	1.2	9
single sampler	228	68	1.2	9
multiple sampler	228	68	1.8	13

In both trials the calculated components of uncertainty for Pb, show that the dominant factor of the uncertainty for this element is the within-sampler variance. The between-sampler variations are obscured because of the high within-sampler standard deviation. On the other hand for the background elements the between-sampler variance exceeds the value for the within-sample variance, which is comparable with the results from analytical collaborative trials where the differences between laboratories are greater than the within-laboratory differences.

Greater values of relative uncertainty were observed for Pb in the second CTS. This indicates the poor performance of the sampling protocol used for this particular element. The reason for this poor performance reflects mainly the high degree of within-location heterogeneity of the soil. This is reflected on the sampling variance of Pb (for duplicate samples separated by 2 m) being 5.20% of the total variance for the simple regular grid at the Derbyshire site (see Fig. 4.7). The equivalent sampling variance for the same element at Hounslow Heath is higher with a percentage of 32% of the total variance (see Fig. 8.3). Small variations in the siting of the sampling locations in the CTS would therefore potentially produce a higher uncertainty (relative to the mean) for the Hounslow site than for the Derbyshire site. The way that the participants interpreted the sampling protocols might also have an effect on the estimation of metal concentrations. The accuracy of the marking out of the sampling targets varied from those who paced out the intervals judging directions by eye and those who used measuring tapes and canes to interpret the protocols more rigorously. This would be expected to have some effect on the estimated metal concentration in soil.

Probably the main cause of the bias however was the susceptibility of the protocols to the heterogeneity in the distribution of Pb in both sampling targets. Relatively small variations in the exact location of the sampling points could easily

result in a small 'hot spot' of contamination being included or excluded, thus affecting the composition of the composite sample. The 'W' design used in the first trial was more susceptible to heterogeneity. The large differences in the orientation of the 'W' would be expected to have such an effect, but there was no clear systematic variation of mean concentration with particular orientation.

The regular grid used in the second CTS was expected to increase the agreement between the results because it covers the sampling area more evenly and with a higher sample density. However, the degree of heterogeneity on a landfill site is likely to be greater, so that small scale variation in Pb concentration may still have been the dominant factor. The contrasting agreement of the Cu and Co mean values for the smelter and the landfill site respectively and the lack of spatial variation in these elements further support this idea.

In general the CTS at the two sites were successful in showing the feasibility of such trials in sampling. Ideally the protocol to be validated should be tested at more sites with different levels of contamination. However, the effectiveness of a sampling protocol at a particular site can be evaluated using a CTS. Such a conclusion is obviously site specific and cannot be generalised. When the technique is applied more widely it may become possible to derive generalisation about potential suitability of particular protocols at contaminated sites.

9.2.2 Comparison of the Results from the Two SPTs

Sampling proficiency tests were applied in this project to assess the general applicability of the method and also to estimate the uncertainty generated in the sampling of contaminated land when different samplers sample a contaminated site using different sampling protocols. The particular objective given to each sampling organisation was to design and implement a protocol for the estimation of the mean metal concentration in the top soil at the sampling sites. The general experimental design was the same at both sampling targets, however the two proficiency tests differed at some points, relevant mainly to the set up of the trials. These differences were justified on the basis of the knowledge gained from the first SPT at the Derbyshire site.

They concern the analysis of the collected samples and the assignment of the assigned value of metal concentrations for z-scoring.

In the first proficiency test the participants were responsible for the analysis of their samples in analogy with analytical proficiency tests. They were asked to analyse the samples and to report the values for Pb and Cu concentrations to the organisers. For the separation of the analytical errors a common reference material HRM31 was distributed to be analysed in the same run with their collected samples. In the second SPT all analyses was done at Imperial College. This was decided as some of the participants were not able to undertake the analysis of the samples and also to avoid delays caused by the analytical process in different laboratories. It was judged that this alteration would not have an effect on the experiment, as the main objective was to assess the performance of the participating organisations in sampling. All the samples were analysed in the same analytical batch for the elimination of any analytical bias between the samples.

The results of both SPTs were assessed using z-scores which is the recommended method by the harmonised protocol for proficiency testing in analysis (Thompson and Wood, 1993). This requires the use of an assigned value for the best estimate of the 'true' concentration at a site and a value for the target standard deviation. In the first proficiency test the assigned values for this sampling target were derived from the pilot study of this field in which a high density sampling scheme was employed, collecting samples every 10 m on average. These values were the best estimates of the 'true' values since the 'consensus of experts' suggested by the harmonised protocol was unavailable to be used as the assigned value. Any error in the assigned value, would not substantially affect the degree of scatter measured between the concentration values reported by the participants, but would affect estimates of bias for each participant. In the second SPT the grand means of the metal concentration in soil from all the participants were used as the best estimates of the 'true' values. Moreover it was observed in the first SPT that the robust mean of the Pb results on the field samples ($6171 \mu\text{g g}^{-1}$) was close to the assigned value ($6296 \mu\text{g g}^{-1}$) so there was no evidence for overall bias when the information from all the participants is combined. The target standard deviation in both trials was arbitrarily set at 20% of the assigned value based on a broad estimate of fitness-for-purpose.

The uncertainty for the mean metal concentration estimates from the two SPTs was calculated from the scatter of the results. The values of uncertainty from the two SPTs are presented in Table 9.2.

Table 9.2. Standard deviations and expanded uncertainties for Pb, Cu, Co concentrations in soil, assuming coverage factor of 2 ($U = 2s$). Data from the two SPTs.

<i>1st SPT (Derbyshire)</i>						
Element	Pb			Cu		
Grand mean $\mu\text{g g}^{-1}$	6171			27		
Uncertainty	s $\mu\text{g g}^{-1}$	U $\mu\text{g g}^{-1}$	U%	s $\mu\text{g g}^{-1}$	U $\mu\text{g g}^{-1}$	U%
	2223	4446	72	7	14	52
<i>2nd SPT (London)</i>						
Element	Pb			Co		
Grand mean $\mu\text{g g}^{-1}$	358			13.5		
Uncertainty	s $\mu\text{g g}^{-1}$	U $\mu\text{g g}^{-1}$	U%	s $\mu\text{g g}^{-1}$	U $\mu\text{g g}^{-1}$	U%
	73	146	41	0.6	1.2	9

From the assessment of z-scores individual organisations showed bias outside the assigned values in both proficiency tests. Scores can be considered satisfactory when values lie between -2 and +2, equivalent to being within 20% of the assigned value. For the estimation of mean Pb at the smelting site, four out of nine organisations failed to meet this 'fitness-for-purpose' criterion. For organisation number 6 this is dominantly an analytical problem (mirrored in analysis of HRM31, see Table 6.6. in Chapter Six, section 6.3.2) but for organisation number 1,2 and 8 this failure seems to arise from the field sampling. At the landfill site all the participants 'scored' within the ± 2 z-score range for Pb and the general level of agreement ($s = 20\%$) was improved from the first SPT in Derbyshire ($s = 36\%$).

9.3 Suggested Methodology for the Estimation of Measurement Uncertainty in the Sampling of Contaminated Land

9.3.1 Four Methods for the Estimation of Measurement Uncertainty

A quantitative methodology for the estimation of measurement uncertainty has been suggested, by adapting methods devised originally for estimating analytical uncertainty to estimate sampling uncertainty. In this project field sampling and chemical

analysis have been considered to be two parts of the 'measurement' process. Their combined contribution to the estimated uncertainty has been recognised and quantified by several methods throughout the experiments of this project.

There are four cases in which measurement uncertainty, including that from sampling, could be estimated:

- (i) for single sampler/single protocol
- (ii) for single sampler/multiple protocols
- (iii) for multiple sampler/single protocol
- (iv) for multiple sampler/ multiple protocols

These will be discussed in turn, using their application at the two experimental contaminated sites used in this research.

Method 1:- Single sampler/ single protocol

This method for estimating sampling and analytical precision involves taking duplicate samples at some proportion of the sampling locations, typically 10% (Ramsey et al., 1992). Duplicates are the most cost effective type of replicates, allowing the estimation of precision over the greatest number of locations. The duplicate samples are not taken at exactly the same position as the original, but displaced by a distance that represents the uncertainty of locating the sampling point, by the particular surveying technology used. This distance will produce differences between sample duplicates caused by small scale local variation, but this variation represents that which may arise randomly in locating the sample location by this surveying technique.

Examples of the application of this method can be considered the estimates of uncertainty of metal concentrations in soil in the two initial surveys at the Derbyshire site (discussed in Chapter Four) and at Hounslow Heath (discussed in Chapter Eight). In both surveys the duplicate samples were taken two metres away from the original location in randomly selected directions. Duplicate samples were taken at 20% of the locations in each grid, and both samples were analysed twice in a balanced design.

The three components of the variation were separated using robust analysis of variance (Analytical Methods Committee, 1989). The sampling uncertainty, or within-location variance, will be partially due to small scale geochemical variation within the

location, but represents the uncertainty in all of the samples that could be taken from that 'location' as specified, in this case within a two metre radius.

For the Derbyshire site for the combined duplicates of all the single auger samples, the extended uncertainty for Pb ($1871 \mu\text{g g}^{-1}$) is 49.8% (U%) for the estimated mean concentration of this site ($7516 \mu\text{g g}^{-1}$). However, by using composite samples the uncertainty drops to only 30.9% of the mean (estimated from the duplicate five-fold composite samples). This reduction in variance is statistically significant at the 95% confidence level, using the *F*-test. This finding quantifies the theoretical prediction that taking composite samples should reduce measurement uncertainty. For the site in Hounslow Heath the extended uncertainty for Pb ($249 \mu\text{g g}^{-1}$) is 83.6% (U%) for a single measurement made at the estimated mean concentration of this site ($297 \mu\text{g g}^{-1}$). The question of whether either of these levels of uncertainty is acceptable and therefore 'fit-for-purpose', will be discussed below.

Analytical bias can be added into these estimates of uncertainty. In this research analytical bias (B_a) was estimated independently by the analysis of six certified reference materials with a wide range of concentration of Pb at both experimental sites. The uncertainty interval of any concentration *c* (e.g. the estimated mean for the site) becomes asymmetric. The upper limit can therefore be calculated from:

$$c + U = c (1 + U\%/100 - B_a/100)$$

and the lower limit from:-

$$c - U = c (1 - U\%/100 - B_a/100)$$

The main problem with this method is that it does not allow for bias originating from the sampling process. This may be caused by the variability between different sampling protocols designed to fulfil the same objective (i.e. estimation of the mean Pb concentration of the site) or from different samplers applying the same protocol. The bias between different sampling protocols can be assessed by Method 2, and the bias between different samplers with the same protocol by Method 3.

An application of this method could be considered to be the use of the portable XRF. The in-situ methodology for the estimation of metal concentration in soil can be viewed as one sampling protocol. The advantage of this methodology is the lower analytical time that enables the user to take duplicate measurements at more sampling locations. In the portable XRF experiment performed in this project (described in detail in Chapter Seven) duplicate measurements were taken at all sampling locations on the 30 m regular grid the Derbyshire site. However, the in-situ methodology did not allow duplicate analysis. The total measurement variance (i.e. sampling and analytical combined) has been separated from the total variance using ANOVA.

Method 2:- Single sampler/ Multiple protocols

It is possible in principle to estimate the bias between different sampling protocols, applied by the same sampler on one sampling target. This procedure explained in detail in Chapter Four, was applied to the Derbyshire site using four protocols; (1) a regular grid with single samples, (2) a regular grid with composite samples, (3) a herringbone design and (4) a stratified random design. The bias between the mean values of Pb concentration estimated by each protocol was found not to be statistically significant for the number of samples taken. This was primarily because of the large uncertainties (i.e. standard error) on each of the mean values, which probably originated from the heterogeneity of the site. The estimates of the uncertainty for the concentration values at each location are similar for all four sampling protocols. Combining measurements from the duplicates from all four protocols gives a more reliable estimate of the extended uncertainty value as 60.4%. This is not significantly different from the value calculated for a single protocol by Method 1. The expanded uncertainty in the estimated mean Pb concentration for the whole site, calculated across the four separate protocols, is 32 %. This shows a much smaller uncertainty for the estimated site means than for concentration estimates at individual locations, as would be expected.

Method 3:- Multiple samplers/ single protocol

This method of estimating uncertainty is effectively 'top down' and is based on the application of a Collaborative Trial in Sampling (CTS). One of the original

objectives of the CTS was to validate a particular sampling protocol. It is also possible however, to use the results of a CTS to estimate the uncertainty of measurements resulting from its application at a particular site.

Examples of this method are the two CTS in Derbyshire and in Hounslow Heath. Although essentially a 'top down' approach, the experimental design did allow the separate estimation of the uncertainty due to the chemical analysis within the one laboratory. For Pb this value is 7.4 % for the first CTS and 16% for the second CTS and is clearly very small when compared with the uncertainty from the two applications of the sampling protocol by each participant (50.9 % and 68% for the two CTS respectively). Interestingly there is no increase in this uncertainty estimate when multiple samplers use this same protocol. This implies that the spatial heterogeneity of the sites was a dominant factor in controlling the uncertainty for Pb, rather than differences between samplers.

For Cu and Co in the first and second CTS respectively, which are not at elevated concentrations in these sites, the analytical component of the variance is a high proportion of the single sampler uncertainty. Moreover the uncertainty for multiple samplers is higher than that for single samplers, showing some contribution of different sampling bias of the various samplers (see Table 9.1).

Although this CTS methodology was applied here to estimate the uncertainty on mean concentrations for a whole site, it could equally well be applied to estimate the uncertainty of the concentration at particular locations within a site.

Method 4:- Multiple samplers/ multiple protocol

In a Sampling Proficiency Test (SPT) each sampler uses a different sampling protocol, which is selected by the participant to fulfil the same objective. The uncertainty estimated from such an inter-organisational trial, reflects the widest definition of uncertainty at one site. For the application of the SPTs in Derbyshire and in Hounslow Heath, the objective addressed by all nine sampling protocols was the estimation of the mean metal concentration at the sites.

Although a SPT gives the most complete estimate of uncertainty is also the most expensive method and the most difficult to set up. This is because the participants of such trials are expected to fulfil several requirements listed below:

1. to travel to the sampling target,
2. to apply an appropriate sampling protocol, selected according to their judgement,
3. to be able to perform the chemical analysis of the collected samples.

In the first SPT organised in this project the analytical measurements were each undertaken by the participating organisation. A reference material (HRM31) was specially prepared with material from this same site, and analysed by each participant to separate the contribution of analytical bias to their measurements. In the second SPT the organisers had to undertake the analysis of the samples for logistic reasons.

The resulting scatter in the estimates of the mean metal concentration has been discussed earlier. Uncertainty values from the two SPTs are given in Table 9.2. The general deduction from the application of this methodology is that the values of estimated uncertainty are affected by the site heterogeneity. The scatter of the estimated concentrations were much higher for Pb than for more homogeneously distributed elements such as Cu and Co.

9.3.2 Acceptable Levels of Uncertainty: Fitness-for-Purpose

Once the uncertainty of measurements has been estimated for a particular survey, one question that arises is whether that level of uncertainty is acceptable. This question has barely been addressed in the literature for any application except the accounting of nuclear materials (De Bievre, 1983). In analytical quality control for example, a target for analytical precision in trace determinations (e.g. 10%) is common practise. Such targets are rarely based, however, on a rigorous evaluation of the precision required for the interpretation of the analytical results. When the requirement is for the spatial interpretation of concentration measurements, it has been suggested (Ramsey et al., 1992) that the limit for measurement variance, and therefore uncertainty, is given by:

$$s^2_{\text{meas}}/s^2_{\text{total}} < 20\%$$

When this proportion is exceeded, then the identification of 'hot spots' on contaminated land tends to become progressively less reliable. Such a level is related to the variability

of the environment under study (e.g. the degree of contamination) rather than using a level fixed in relative terms (e.g. 10% RSD). No matter how large the measurement uncertainty becomes however, it is always possible to use the estimated uncertainty values. These can be used to test whether one location is contaminated more than another location (or more than a threshold value), using traditional statistical techniques. In that sense therefore, there is no universal limit to the acceptable level of uncertainty. Retrospectively however, it may prove that the uncertainty of one survey prevents the quantification of a particular objective, such as the delineation of a contamination 'hot spot'. In this case a second survey would be required which would have the value of uncertainty reduced to an appropriate level. This acceptable level of uncertainty needs to be calculated therefore from the results of the first survey.

9.3.3 Uncertainty on the Vertical Variability of Metal Concentration

The geochemical interpretation of the data from both sampling trials focused on the lateral extend of contamination and specifically the estimation of metal concentrations in the top 0-15 cm layer of soil. Since soil sampling was done mainly by augering, all the analysed samples can be considered to give the average metal concentration in this layer. However, it is interesting to study the vertical variation of the contamination at the sampling targets.

For the old smelting site in Derbyshire there are some evidences about the vertical variability. These are based on a separate study of the downwards migration of contamination (Maskall and Thornton, 1994). In the study, drilling to depth of 6 m has revealed metal concentrations at three locations in the field. The vertical migration of Pb in two of the cores appears to be attenuated by the clay/sandstone matrix but in the third core Pb has migrated to a depth of 4 m (Maskall et al., 1995).

Evidences about the metal variability in the top 0-15 cm sampled in this project come from the portable XRF core experiment described in Chapter Seven, section 7.4.3. The Pb profiles produced by this 'sliced core experiment' show little systematic vertical spatial trend in the collected cores. In general the lateral heterogeneity of this site in respect to metal concentration is very high compared with the vertical one in the top 0-15 cm.

In the old landfill site in Hounslow, London there is not enough evidence on the vertical distribution of the contamination. This is primarily because of the shallow nature of soil in this site. Attempts to auger below the top 20 cm were fruitless because of the gravel lying underneath.

In general however, the suggested methodology could be applied for the estimation of uncertainty in the vertical distribution of the contamination. This could be done if bore hole determinations were available. Such an approach has been tested with bore holes by an 'is-situ' geochemical logging tool in a borehole penetrating a cyclothem sequence at Northumberland, UK (Ramsey et al., in press). The uncertainty on the estimates of concentrations of six elements was calculated from a combination of a random and a systematic component estimated independently. Upper and lower limits of uncertainty have been estimated in this study.

Inter-organisational trials could in theory be applied with such an objective. However, difficulties in the set up of these trials would comprise the need for excavation of the soil. The resultant substantial disruption of the site may affect the independence of the participants.

9.4 Probabilistic Classification of Contaminated Land

The current method used to classify contaminated land is to compare element concentrations measured in the soil against a threshold value. In the UK, for example, the values used as threshold trigger concentrations have been set by the Inter-departmental Committee on the Redevelopment of Contaminated Land (ICRCL) (UK Department of the Environment, 1987).

The problem with such a deterministic approach is that it ignores the existence of measurement uncertainty. There is therefore a risk of misclassification because of this uncertainty. There is a definite probability that a single soil sample from a particular location within an area of land will give a measurement of a contaminant concentration below the threshold value, even when repeated samples taken from that location would show that the soil has an average concentration that is over the threshold. Conversely, there is also a possibility that one sample may give a measurement over the threshold, where repeated measurements show that the average concentration at a particular location is under the threshold. What is required is a probabilistic classification of contaminated land that takes into account the uncertainty of a concentration measurement as well as its value.

The proposed method specifies four categories for the level of contamination based on the overlap of the uncertainty with a single threshold level (Fig. 9.1). It is derived from the general approach to uncertainty taken by Thompson (Analytical Methods Committee, 1995). Four categories can be distinguished:

1. Category 'Uncontaminated' where all of the range of the uncertainty is lower than the threshold value (T). The probability of this site being contaminated over the threshold is therefore <0.025 (i.e. 2.5%).
2. 'Possibly contaminated' category, where although the measured concentration value is less than the threshold, the upper side of the range of the uncertainty does cover the

threshold value. In this case the probability of this site being contaminated over the threshold is therefore between 0.025 and 0.50 (i.e. 2.5 - 50%).

3. Category 'Probably contaminated' where the measured concentration exceeds the threshold value, but there is still a probability that the true mean value is less than the threshold (0.025 - 0.50). For this category the probability of this site being contaminated over the threshold is therefore between 0.50 and 0.975. (50 - 97.5%).
4. Category 'Contaminated' with probability of the site being contaminated over the threshold is >0.975, although on the lower boundary of this class there is still a very small chance (2.5%) that the true concentration value at the site is under the threshold value.

There is however, a modification to this calculation which can make the classification even more rapid, although it can lead to misunderstanding in interpretation, unless the logic of the matter is well understood as will be explained. If the uncertainty values for the particular measurements made are calculated for the concentration equal to the threshold value, then the classification of the soil samples can be made directly, without calculating uncertainties for each measurement. The equivalence of this second method of calculation to the first can be seen by comparing Fig 9.1(b) with Fig 9.1(a). The concentration ranges for this method of calculation are shown in column 3 of Table 9.3, and are arithmetically equal to those for the original method shown in column 2.

The potential problem with this 'short cut' method of calculation is that it appears to use a modified 'threshold value' (e.g. $T-U$) and not the regulatory threshold value (T). These are not really modified thresholds, but calculation intermediates to identify measurements that fall into the category of 'uncontaminated' ($C+U < T$). Furthermore this method assumes that only single samples are taken at each location. For sites with duplicate samples the 'modified threshold value' would be $T-(U/\sqrt{2})$.

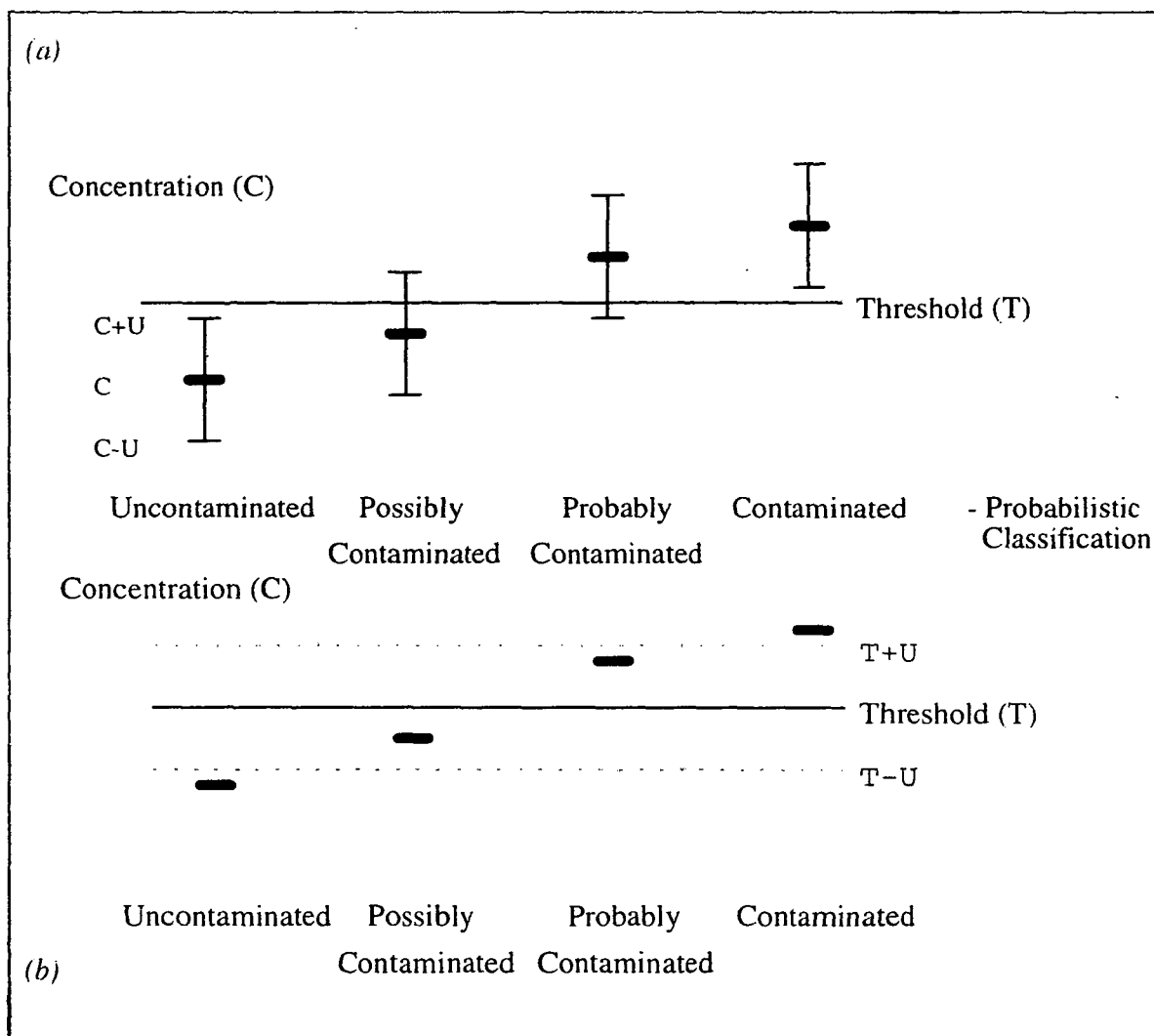


Figure 9.1. Probabilistic classification of contaminated soils, allowing for uncertainty on either the measured concentration values (a), or around the threshold value (b) for a rapid method of classification.

The probability of the location being contaminated over the threshold is given in Table 9.3.

Table 9.3. Probabilistic classification of soils with measured concentration (C) and uncertainty (U), based on the probability that the contaminant concentration is greater than a particular threshold value (T). Uncertainty values are stated for 95% confidence interval, but could be recalculated for what ever confidence interval is considered appropriate.

Classification	Concentration Range <i>with uncertainty on C</i>	Concentration Range <i>with uncertainty on T</i>	Probability of C > T <i>i.e. C being over threshold</i>
Uncontaminated	$C + U < T$	$C < T - U$	<0.025
Possibly contaminated	$C < T < C + U$	$T - U < C < T$	$0.025 - 0.500$
Probably contaminated	$C - U < T < C$	$T < C < T + U$	$0.500 - 0.975$
Contaminated	$T < C - U$	$C > T + U$	>0.975

The probabilities quoted here are given for an expanded uncertainty with a coverage factor of two. It is equally valid to select a different coverage factor that reflects a different probability of misclassification that is considered more appropriate for a particular site. The appropriate coverage factor could give an indication of the degree of heterogeneity in a specific site. The probabilities are also given assuming single estimates of concentration being made at each sampling location. If 'n' multiple measurements have been made at a location, then the uncertainty value for that site should be divided by $\sqrt{n}$, before the classification is made.

9.4.1 Application of Probabilistic Classification to Contaminated Land

In the methods for estimating uncertainty discussed above (2, 3 & 4) the uncertainty (U) is estimated at the mean concentration and converted into relative units (U%) by dividing by the mean. The application of this value to all the measurements of soil concentration assumes that this relative uncertainty remains constant with varying concentration. There is evidence to show that relative precision of analytical methods approaches a constant asymptotic value when measurements are made at concentration levels more than two orders of magnitude above the detection limit of the analytical technique (Thompson and Howarth, 1976). For most applications on contaminated land

this is a reasonable assumption, given the high levels of most threshold values quoted for contaminants. If this is not the case then a slightly more elaborate calculation is needed that uses a mathematical model for the change of uncertainty with concentration.

When considering the calculation of the uncertainty on concentration estimates (C), the upper limit of the uncertainty for each single concentration measurement is given by:-

$$C + U = C(1 + U\%/100) \quad (2)$$

The lower limit is similarly given by:-

$$C - U = C(1 - U\%/100) \quad (3)$$

When the 'short cut' method is being used, with the uncertainty expressed on the threshold value, then the upper and lower limits are given by:-

$$T + U = T(1 + U\%/100)$$

$$T - U = T(1 - U\%/100)$$

If U% has been estimated only for a single sampler (methods 1 or 2) and a separate estimate of measurement bias (B) is available, which has not been used to correct the measurements, then the limits are given by:-

$$T + U = T(1 + U\%/100 - B/100)$$

$$T - U = T(1 - U\%/100 - B/100)$$

9.4.2 Example of Probabilistic Classification

The Pb data from the initial survey in the 270m x 270 m area of Hounslow Heath (described in Chapter Eight, section 8.2.3.3) are used here as an example of the suggested probabilistic classification. These data were selected because of the levels of Pb contamination at this particular site. Specifically, the relatively low metal concentration in soil, close to the regulatory limit could demonstrate the significance of the estimation of the uncertainty in the classification of contaminated land. If this land were to be considered hypothetically for residential development, the Pb concentrations could be classified by the usual deterministic method using the ICRCL trigger threshold for domestic use of $500 \mu\text{g g}^{-1}$ (Fig. 9.2a). A simplistic interpretation of such a map would be that the area appears broadly uncontaminated, with a few (i.e. 8/100) small 'hot spots' that could apparently be avoided for the purposes of development by locating houses in the areas with Pb concentrations $<500 \mu\text{g g}^{-1}$.

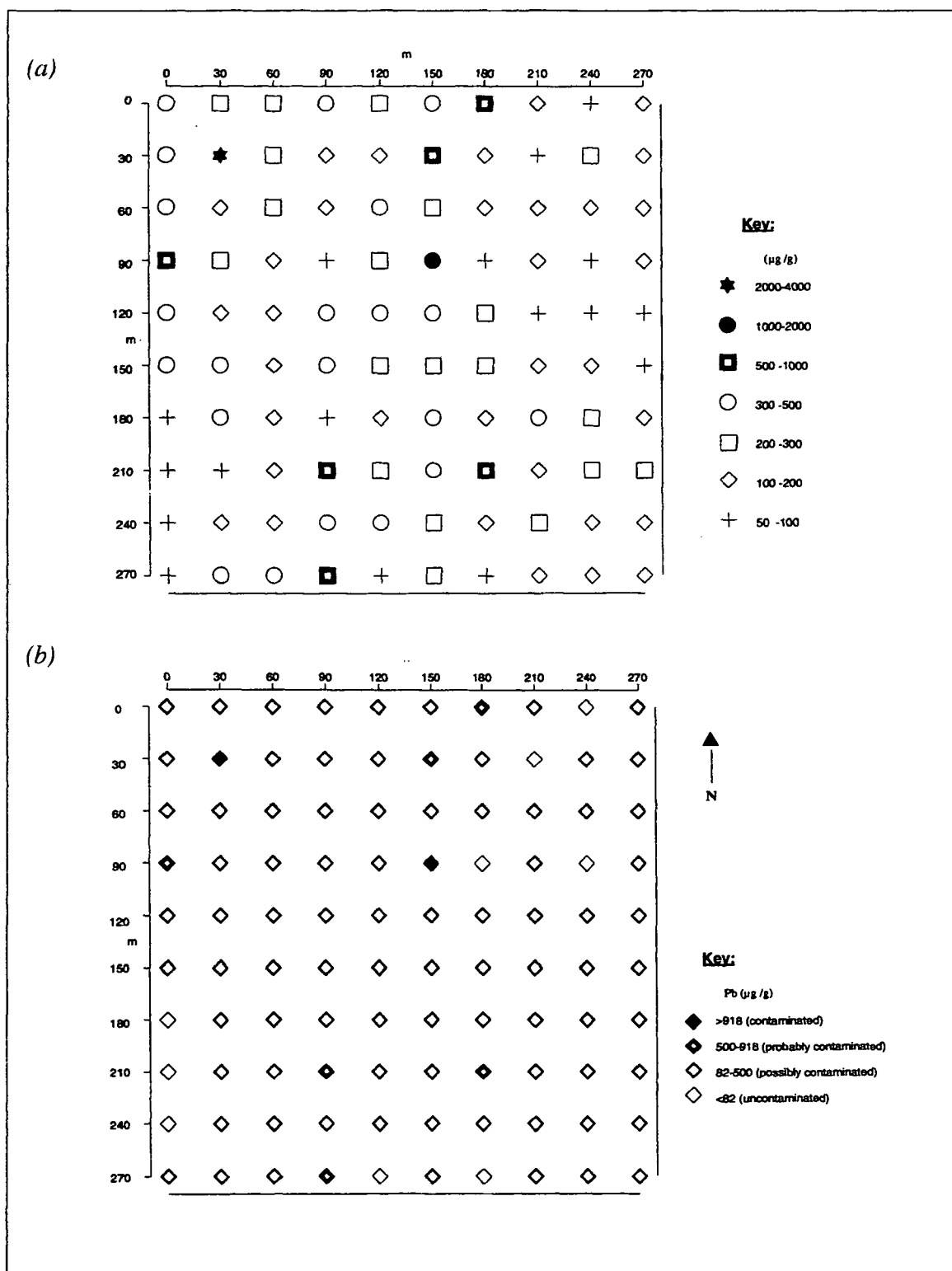


Figure 9.2. Geochemical maps of Pb contamination in the Hounslow Heath site, using a threshold value of $500 \mu\text{g g}^{-1}$. The deterministic approach (a), shows a few isolated locations of contamination over the threshold. The probabilistic map (b), shows that almost the entire area is possibly contaminated over the same threshold, because of the very high measurement uncertainty of the survey.

The probabilistic assessment, using Method 1, required that ten duplicate samples were taken. The very high level of estimated uncertainty (83.6%, as calculated in Chapter Eight, section 8.2.3.1) reflects the extreme small-scale spatial variation of Pb contamination on the site. The probabilistic classification boundaries can be calculated using equations (2) and (3). The concentration equal to the upper limit of the category 'Uncontaminated' ($C = T - U$, equivalent to $C + U = T$) is $82 \mu\text{g g}^{-1}$ and the lower limit of 'Contaminated' ($C = T + U$, equivalent to $C - U = T$) is $918 \mu\text{g g}^{-1}$. These concentration limit values are calculated for 95% confidence, but changes to the coverage factor in the calculation of $U\%$ from equation (1) could be used to specify a different level for the acceptable level of misclassification.

When the map of contamination is redrawn using this probabilistic classification (Fig. 9.3b), then a very different result is evident. Virtually all of this area is at least 'Possibly contaminated' and only nine sporadic sites are classified as 'Uncontaminated'. In effect there are no groups of sites that could be reliably be classified as uncontaminated in the survey area.

9.4.3 Implications of Probabilistic Classification

The use of the probabilistic classification can have large financial implications for the treatment of contaminated land. For this example on Hounslow Heath, the specification of 95% confidence for the classification of the land as 'contaminated' may mean that the cost of remediation may be far higher than if the traditional deterministic classification were used. However, the use of the deterministic method would lead to a high probability of some contaminated areas being miss-classified as uncontaminated and therefore not being treated. The traditional approach could therefore have even greater financial implications than the probabilistic, in the longer term with respect to legal liability.

The *advantages* of the probabilistic approach are:

1. A reduced risk of misclassification of land (e.g. cost savings on either unnecessary remediation, or overlooking the need for remediation).

2. Variable probabilities of misclassification can be incorporated, to reflect the particular circumstances of the site (e.g. intended land use).
3. Cost-benefit analysis can be employed on the design of the sampling survey. For example in assessing the effectiveness of taking composite samples for reducing uncertainty.
4. Uncertainty estimates can also be incorporated in environmental risk assessment to give uncertainties in estimated risks.
5. Probabilistic classification facilitates identification of causes of measurement uncertainty and indicates where improvements need to be made (e.g. in sampling rather than in the analytical method).

Disadvantages of the probabilistic classification:

1. More land will become classified as 'possibly contaminated' than is considered 'contaminated' under the deterministic method. Such areas can be re-sampled by a method with lower uncertainty, to clarify its true status.
2. Increases cost of initial survey, typically +10 to +20% for duplicate samples (i.e. for Method 1). The measurement of uncertainty has a cost, but the information gained will be more useful than unqualified measurements of concentration.
3. It will be an increased complexity in explaining the survey findings. However even the general public is becoming more familiar with measurements expressed in a probabilistic way, in some weather forecasts for example.
4. The estimates of uncertainty made for sampling methods are site specific, and are not generally applicable, as would be the case for uncertainty from analytical methods. Given that the classification of contaminated land is moving towards a more site-specific approach this may not be a serious limitation.

In the longer term there will be a need for quality assurance systems to ensure that organisations engaged in field sampling provide unbiased estimates of measurement uncertainty, as well as for the concentration of contaminants. Sampling proficiency tests could be used to fulfil this role by the evaluation of both concentration and uncertainty estimates.

This study has addressed uncertainty in concentration estimates at particular sampling locations. It has not discussed the uncertainty of concentration estimates

between sampling points, which require geostatistical techniques, such as kriging. However, the estimates of uncertainty at sampling points made by the methods described here could in principle, be fed into geostatistical methods to improve estimates of uncertainty between the sampling points. Some ideas of the potential integration of geostatistical techniques with the estimates of uncertainty are discussed in the following paragraphs.

9.5. Spatial Uncertainty

The estimation of errors in spatial interpretation of the geochemical data is a more complex aspect than the estimation of the errors when estimating the average analyte concentrations or errors on concentration estimates at particular sampling locations. The factor of space and the need for interpolation between the sampling points can introduce additional sources of error. A geostatistical approach could give solutions to the problems of estimating the error in the location of contamination hot spots and also for the plotting of a probabilistic map of areas contaminated over a threshold.

Although such an approach has not been examined in detail in this research, geostatistical techniques could be very useful for studying measurement uncertainty especially for the objective of locating contamination hot spots. This is a very interesting topic for further study. In the following paragraphs a very brief description of potential geostatistical techniques and some ideas of how these could be used in relation to the methodology presented in this thesis is presented.

9.5.1 Variograms and Kriging

In order to estimate errors in the spatial description of metal contamination it is necessary to infer geochemical information between the sampling points. This spatial description of the geochemistry typically involves geostatistical techniques such as variography and kriging. These methods are the classic methods for mapping the spatial distribution of one or more attributes, with emphasis given to characterising the variogram model and using the kriging (error) variance as a measure of estimation accuracy (Deutsch and Journel, 1992).

In general, data close to each other may have more similar values than data far apart. The variogram $\gamma(h)$ is half the average squared difference between the paired data values (Issaks and Srivastava, 1989):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j): h_{ij}=h} (u_i - u_j)^2$$

where, h is the vector separating the pair measurement points.

N is the number of observations,

u_i, u_j the paired data values.

The variogram is a measure of spatial variability. The experimental variogram is calculated from the sample data and the model variogram is chosen by curve-fitting on the experimental data. The features of the model variogram (i.e. range, sill, nugget effect) indicate the degree of spatial variability/ continuity of the data.

Once the variogram has been calculated and the data have been tested for spatial dependence, kriging can be applied to give an average estimation of the property under study. The kriging method is a weighted moving average estimation method where the selection of weights is made such that the estimation variance is minimised (Olea, 1991). Interpolation of the experimental data onto a regular grid using kriging can give an estimate of the spatial distribution. However, the presence of nugget effect influences the calculated estimates and results in a smoothing effect.

For the production of a probabilistic map of areas of land contaminated over a stated threshold geostatistical techniques involving indicator kriging or simulation (discussed briefly in the following section) can be used.

Indicator kriging is a method that can be used to provide a model of uncertainty, by estimating the probability of a parameter having a value below or over a cut-off (z). The binary transform of a random function $Z(u)$ can be defined as:

$$I(u; z) = \begin{cases} 1, & \text{if } Z(u) \leq z \\ 0, & \text{otherwise} \end{cases}$$

where, $Z(u)$ is a random function,

$I(u; z)$ is the binary transform of the random function $Z(u)$,

z is the selected threshold.

Kriging of the indicator data provides the best least squares estimate of the indicator random variable (Deutsch and Journel, 1992). Problems of this method arise when there is equal probability for the value being above or below the threshold. Also the existence of nugget effect reduces the sensitivity of kriging, affecting the decision. To overcome these problems the use of simulation instead of kriging has been suggested (Deutsch and Journel, 1992).

Unlike kriging used as an interpolation algorithm, simulation provides many alternative numerical models each of which is a "good" representation of the reality (Deutsch and Journel, 1992). Geostatistical simulation algorithms generate realisations taking into account the statistical and spatial properties of the data. If the realisations characterise the spatial uncertainty of the parameter the resulting distribution of the system respond values will reflect the uncertainty (Gotway^{and Rutherford}, 1994).

Conditional simulation algorithms honour the data values at their locations, ^{and} overcome the smoothing effect of kriging. For studies of patterns of local variability the conditional simulation approach is therefore preferable (Deutsch and Journel, 1992).

Sequential Gaussian simulation includes all data available in a neighbourhood of a location u , including the original data and all previously simulated values. The conditioning data are first transformed to standard Gaussian values. Kriging based on the variograms of the transformed data is used to obtain estimates of the necessary conditional distributions. Realisations are then drawn randomly from this distribution. Finally the simulated normal values are back-transformed into simulated values for the original variable (Gotway^{and Rutherford}, 1994).

This approach could be applied in theory for the estimation of bias in the location of contamination hot-spots. If the "best" estimate of the area of a contamination hot-spot could be established, different scores could be attribute to estimates derived from different realisations. Fitness-for-purpose would be a compensating factor for the performance in this approach. The technique can also be used for the production of a probabilistic map

with confidence intervals. On this map delineated areas classified as contaminated with confidence of >95% for example can be shown. Moreover, values of uncertainty from the four methods described in this thesis could be used to realistically constrain the simulation.

9.5.2 Spatial Interpretation of the Experimental Data

Variography was applied to the data from the two initial surveys in Derbyshire and in Hounslow Heath to test the spatial continuity of the Pb distribution in the sampling targets. The data from Derbyshire were log transformed and the experimental variogram was plotted using the GEO-EAS¹ software. A spherical model variogram was fitted to the data with a range of 100m, a sill of 0.6 and a nugget effect of 0.2 (Fig. 9.3a). The data from Hounslow Heath were transformed to their Gaussian values in order to plot the experimental variogram. The spherical model fitted to the latter data set has a range of 101 m, sill of 0.5 and a nugget effect of 0.43 (Fig. 9.3b).

¹ GEO-EAS (Version 1.2.1), Geostatistical Environmental Assessment Software.

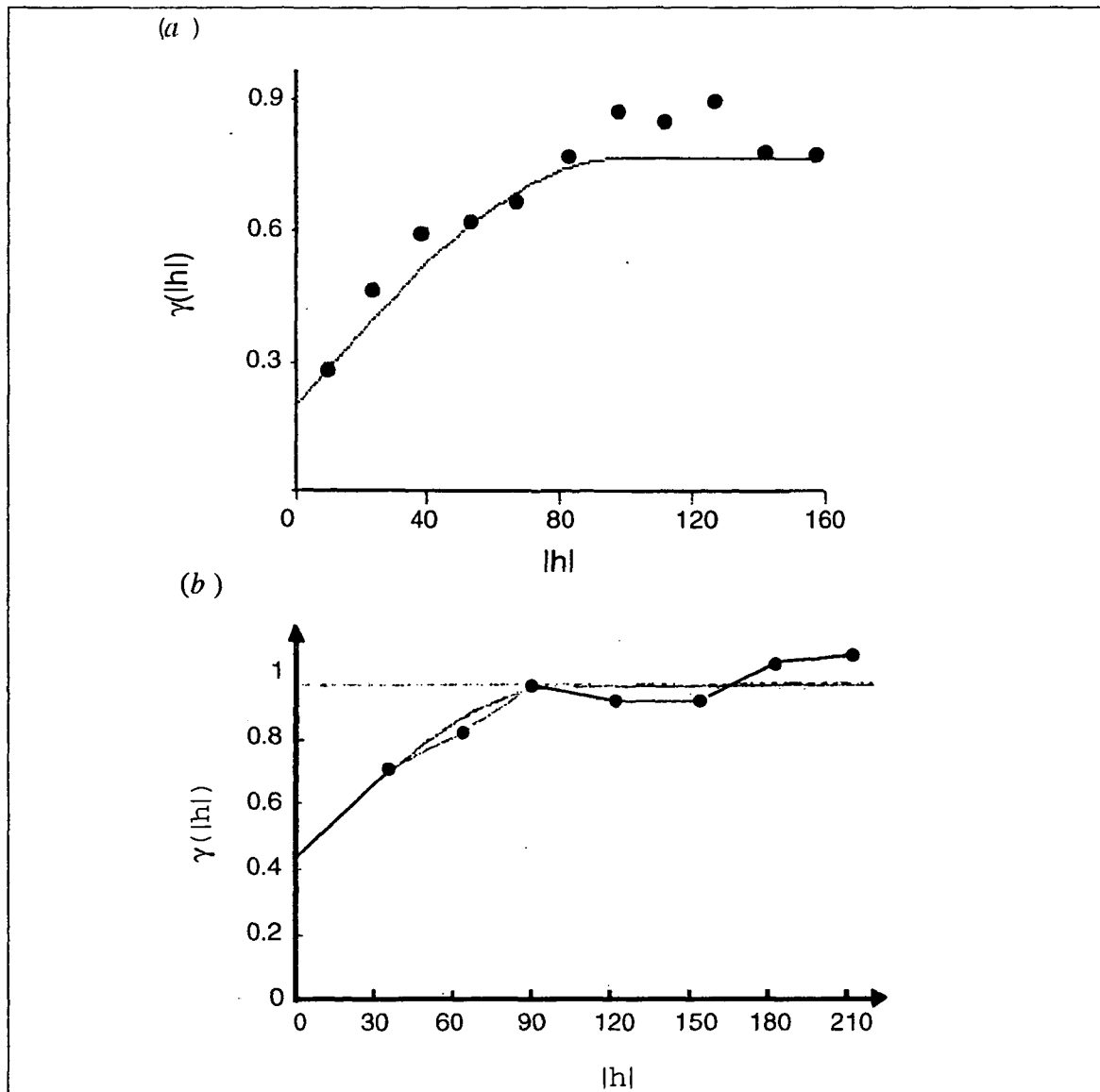


Figure 9.3. Experimental variograms and the fitted spherical models of the ^{transformed} data from the Derbyshire site (a) and the Hounslow Heath site (b).

9.5.3 Comparison of Uncertainty Calculated with Duplicate Samples and the Plotted Variogram

A nugget effect of 0.2 is present in the plotted variogram of the Derbyshire data (Fig. 9.3a) and of 0.43 in the variogram of the Hounslow data (Fig. 9.3b). These variances, present at zero lag distance imply the existence of small scale variability as well as the presence of measurement errors. The higher value of the nugget effect in the old landfill site indicates that small scale heterogeneity is even higher than in the old smelting Derbyshire site. With Method 1 for the estimation of measurement uncertainty, the

measurement variances were calculated from the duplicate samples collected in the surveys. These two estimates of variance can be compared (Kapumpa, 1995).

Specifically, the nugget effect, as a proportion of the total estimated variance (c_0/s_{tot}^2) estimated from the plotted variogram, can be compared with the proportion of the measurement variance over the total variance (s_{meas}^2/s_{tot}^2) calculated from the duplicate samples. The statistics calculated by the different methods are summarised in Table 9.4:

Table 9.4 Average Pb concentration ($\bar{x}$) and total variance (s_{tot}^2) calculated from ANOVA and the plotted variograms. (n ; number of samples)

<i>Derbyshire site</i>			
<i>Method</i>	<i>n</i>	<i>x</i> ($\mu\text{g g}^{-1}$)	<i>s_{tot}²</i>
Classical ANOVA	27	6749	5.95E+07
Robust ANOVA	27	4953	1.40E+07
Variogram	119	6279	3.92E+07
<i>Hounslow Heath site</i>			
Classical ANOVA	10	318	240
Robust ANOVA	10	297	218
Variogram	100	292	380

The above procedure has yielded $c_0/s_{tot}^2 = 28.9\%$ and s_{meas}^2/s_{tot}^2 equal to 21.7% and 10.9% as calculated from the classical and robust ANOVA respectively for the Derbyshire site. For the Hounslow Heath site the ratios were calculated to be: $c_0/s_{tot}^2 = 44\%$ and s_{meas}^2/s_{tot}^2 equal to 32% from both the classical and robust ANOVA. The greater magnitude for the nugget effect ratio compared to the ANOVA ratio might be because it includes both the measurement error and the error due to small scale variation.

In general the estimation of measurement uncertainty by using the duplicate samples technique seems to be more straightforward than the geostatistical approach. Moreover the calculated errors include errors from the treatment of the samples e.g. contamination of the samples in the laboratory. However, in the instance of extremely heterogeneous sites the two approaches would give similar results. This is because a high degree of heterogeneity is equivalent of sampling on a more dense grid and thus having a large data set for the geostatistical model.

The estimation of uncertainty that is based on the collection of duplicate samples (Method 1) is much faster and more flexible than the geostatistical approach. The collection of a few duplicate samples is sufficient for the estimation of uncertainty while in geostatistics the amount of data needed to extract a variation model is much larger. The duplicate sample approach can be applied to any sampling design however small or irregular, whereas the geostatistical approach requires evenly distributed sampling designs to perform well. Also geostatistics might be biased in the way that they rely on one data set while the sampling trial methodology comprises the collection of data from several organisations.

9.6 Potential Application of the Suggested Methodology for Uncertainty Estimation for Other Sampling Media

Theoretically all four of the methodologies described here for the estimation of measurement uncertainty on contaminated land could be applied with satisfactory results to other sampling media. Specifically for sampling trials the methodology could be applied in all instances where duplication and maintenance of the independence of the participants of the inter-organisational trials is possible. However, specific problems related to the nature of the sampling medium have to be considered in each instance. In general, factors that have to be examined in order to recognise the suitability of the sampling target for such trials are the potential stability, the degree of heterogeneity and its size. Some examples are given in the following paragraphs.

9.6.1 Sampling of Bulk Materials

The example of sampling of raw materials such as piles of ore or foodstuffs seems to conform to the requirements mentioned above. In this instance the set up of a sampling trial would be quite straightforward.

The temporal stability of such sampling targets would have to be examined by sampling with a recognised protocol repeatedly over a suitable period of time. Some portion of the sample could be removed and analysed. Attention would have to be given

to the storing of the sampled material, to ensure it changes in the composition did not occur.

The maintenance of the degree of heterogeneity of such sampling targets could be a difficult task. Piles of ore would have to be mixed over time intervals to prevent segregation. However, the organisers of the trials would be responsible for such actions so that the state of the sampling target will be the same for all the participants of the trial.

In the instance of bulk materials, sampling is done along a conveyor by removal of increments by scoops. Duplicate samples could be collected in fixed intervals after the collection of the original sample on the conveyor. Another possible way for the collection of duplicate samples can be considered with the example of sampling a large pile of raw material by coning and quartering. Samples could be collected from the top of the cone by taking the two opposite quarters and rejecting the other two. The duplicate sample could be collected after rebuilding the cone of the rejected material and collecting the other pair of the opposite segments (Thompson and Ramsey, 1995).

The size of such sampling targets would need to be related to particle size as well as the degree of heterogeneity.

9.6.2 Sampling of Temporally Variable Targets, Such as Water

The estimation of measurement uncertainty from sampling temporally variable targets, such as water is a more difficult task. The main characteristic of such sampling targets that affects the set up of inter-organisational trials is the lack of temporal stability even for short time scales. Because of their flow and the existence of this temporal variation it would be very difficult to prevent the effects of rapid changes in composition. Also difficulties may arise depending on the specific type of water sampling, e.g. sampling of rivers and streams, water from lakes or ground water sampling.

Duplication would be a more complicated procedure taking into account not only spatial variation but also variation with time. When sampling along a stream for example, two samples collected at a fixed distance will have to be recollected after the passing of a fixed period of time to quantify differences in the composition.

The depth of sampling will have to be defined carefully, otherwise comparisons would not be possible to be made between the collected samples.

Sampling trials would have to be completed in a relatively short period of time in order to be able to compare the results from the different participants. an experimental design for inter-organisational sampling trials would need to be carefully designed so that significant bias between the samplers could be separated from variation caused by temporal fluctuations. Special care should also be taken during the transportation of the samples because contamination of the samples or loss of the analyte can easily occur.

CHAPTER TEN. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER STUDY

10.1 Introduction

The general conclusions from this research on the estimation of measurement uncertainty in the sampling of contaminated land are presented. Methodologies for the quantification of uncertainty due to sampling have been developed through experimentation and applied to two contaminated sites as examples. Inter-organisational sampling trials have been organised for the first time, by analogy with the methods used for analytical data quality, and four different methods for estimating measurement uncertainty have been suggested. Finally some recommendations for further study are presented.

10.2 The Initial Surveys

Two contaminated sites have been used in this research for experimentation of the estimation of sampling errors. The first one, situated in Derbyshire, was a heavily contaminated site used in the past for Pb smelting. The second was a less contaminated site with a history of landfill use, situated on Hounslow Heath in west London. The pilot studies at these sites were very useful not only for showing the potential of the sampling targets for applying inter-organisational sampling trials, but also to estimate the uncertainty of sampling when only one sampler is involved in sampling.

10.2.1 The Old Smelting Site in Derbyshire

It is possible to estimate uncertainty in field sampling when one sampler is commencing sampling, either using one or multiple sampling protocols on an area of contaminated land. The study focused on two elements with contrasting properties in the soil, Pb and Cu. Four different sampling protocols were

applied for the estimation of the mean concentration of these two elements in top soil. Lead showed a high geochemical variability across the site while Cu showed very little geochemical variation. The natural heterogeneity of the elemental concentration affects the magnitude of precision and makes the sampling variance the dominant factor in the estimation of uncertainty. This is also a limiting factor in the size of systematic differences that can be detected between the different sampling protocols.

The size of the sampling bias that can be detected at any site therefore varies according to the geochemical variance of the element. This has been demonstrated for Pb and Cu at the particular site for which the minimum sampling bias that is demonstrable is $726.5 \mu\text{g g}^{-1}$ for Pb and $0.8 \mu\text{g g}^{-1}$ for Cu. No significant difference was found between the mean concentrations estimated by the four different sampling designs employed for Pb. For Cu a significant difference of 8.7% was detected between the simple regular grid and the stratified random protocols but not between the other protocols. The optimal sampling scheme for this objective at this site for Pb or Cu therefore appears to be the regular grid because it is the fastest to set up and least prone to location errors.

Duplicate sampling is a useful method for the estimation of the overall sampling precision within a site and is applicable to any of the various sampling protocols. The improvement in sampling precision gained by using a five-fold composite sample at each regular grid point is significant (5.2% to 2.7% of the total variance) but judged not to be worthwhile in this case, because the original sampling variance was well below the limit of 20%. Uncertainty estimates were calculated from the duplicate samples collected from all the sampling designs giving a value of extended uncertainty (U) $\pm 49.8\%$ for Pb mean concentration across the site and of $U = \pm 20.8\%$ for Cu.

The results from sampling using different sampling protocols gave useful information on the levels and the extent of contamination. The particular site studied was judged to be suitable for inter-organisational sampling trials such as a sampling proficiency test and a collaborative sampling trial. This was based on the presence of a variety of elements with either low or high concentrations in soil (e.g. Cu, Pb) as well as all the other criteria listed in Chapter Three, section 3.2.1. With these methods an estimate of variability between different samplers (i.e. sampling bias) could be made. This particular site however, may have limited usefulness as a reference sampling target (i.e. the sampling equivalent of a reference material) because of the high degree of heterogeneity in respect to metal concentrations in soil.

10.2.2 The Completed Landfill Site in Hounslow Heath, London

From the initial survey in Hounslow the levels of contamination and the degree of heterogeneity of the soil at this site were determined. The study focused on two elements, Pb and Zn as examples of contaminants in the site. The concentration of Pb in soil was not very high with a mean of $292.1 \mu\text{g g}^{-1}$. However, high values of this element's concentration were observed at isolated sampling locations. The variability of Pb in soil was quite high, giving rise to high levels of extended measurement uncertainty of $\pm 83.6\%$ for the 95% confidence interval, estimated from duplicate samples and analyses.

Concentration of Zn was also elevated with a mean of $283.6 \mu\text{g g}^{-1}$ which is close to the phytotoxic limit of $300 \mu\text{g g}^{-1}$ for this element in soil according to the ICRCL classification for soils. The measurement uncertainty for this element was estimated to be $\pm 50.4\%$, a lower value compared to Pb, but still relatively high.

The elevated values of measurement uncertainty are mainly due to the high degree of heterogeneity of the soil. This was expected since the sampling area was a part of the old landfill site. The used sampling protocol was unable to overcome this heterogeneity and hence spatial interpretation of the data could not be justified.

10.3 Portable XRF Methodology and Sampling Uncertainty

The performance of the portable XRF Spectrace TN9000 used as an in-situ sampling and analytical technique has been evaluated. The technique, used for the estimation of Pb and Zn concentrations in soil at the old smelting site in Derbyshire, has proved to be a useful and powerful tool capable of giving precise and very rapid results.

The method was compared with established methods for the assessment of contaminated land, such as ICP-AES analysis of solutions derived from the soil samples and laboratory analysis of powdered samples using XRF. A rotational bias of -57% was calculated between the in-situ P-XRF and the ICP-AES measurements for Pb. The magnitude of bias between the in-situ P-XRF measurements and the P-XRF results in laboratory mode was also estimated to be -52% for the same element. These differences in the performance of the analytical technique were shown to be due largely to factors

related to the nature of the samples, i.e. the soil moisture and the surface roughness of the samples. When a correction was made for these factors at two locations within the field, the bias for Pb measurements was effectively reduced to a statistically insignificant level.

In spite ^{of} the bias of the method, in-situ P-XRF measurements were capable of rapidly locating contamination hot-spots in the field. Furthermore the detected contamination hot spots on the initial 30 m grid could be investigated during the same working day, by setting up a more dense sampling grid around the areas with the highest metal concentrations. The measurement bias does not affect the spatial coherence of the hot spots but only affects their intensity and extent. Consequently the effect of the bias, if uncorrected, would be vital in the instance of metal concentrations being close to a regulatory limit.

A very important characteristic of the in-situ P-XRF methodology is the ability to investigate the small scale variability of metal content in soil. This is particularly useful in sampling targets showing high degree of heterogeneity in metal distributions. The degree of heterogeneity can easily be quantified by estimating the variability in different spatial scales within the field.

The limitations of the methodology at this particular site include the small effective sample size ($\sim 1 \text{ cm}^3$) that might result in poor representativity, the small depth of analysis, the bias on the estimated concentration values and the relatively high cost of the equipment. The recognition of the cause of the bias and the application of corrections of the soil moisture and surface roughness effect on the results were needed to eliminate the differences detected between the on-site and off-site analyses.

The portable XRF methodology can also be viewed as an alternative sampling protocol, utilising iterative sampling and on-site estimation of the contamination. From this point of view measurement uncertainty of $\pm 55\%$ for Pb was estimated by taking duplicate measurements at every sampling location 2 m away from the initial sampling point.

10.4 Inter-organisational Sampling Trials

10.4.1 The Collaborative trial in Sampling and the Sampling Proficiency Test

Two types of inter-organisational trials were applied for the first time in the sampling of contaminated land: the collaborative trial in sampling (CTS) and the sampling proficiency test (SPT). The trials were carried out at the two sites in Derbyshire and in Hounslow Heath. For the CTS the participants separately visited the sites and applied the sampling protocol supplied twice at each site, to collect top soil samples. The samples were then analysed in duplicate under randomised repeatability conditions. The resultant data were analysed statistically by nested analysis of variance (ANOVA) to provide estimates of the analytical repeatability standard deviation, the sampling repeatability standard deviation and the between-sampler standard deviation. In this way the within- and between- sampler precision of the estimates of the mean metal concentrations over the sites could be quantified. The estimation of the mean metal concentration in top soil was selected as the objective for the participants in these trials to give a simple statistic for comparison. More realistic yet more complicated objectives such as spatial location of contamination hot spots should be possible in future trials.

With the SPTs the performance of the participants was assessed when they were using sampling protocols of their choice for the estimation of average metal concentrations in the sampling targets. The analytical results of these trials were converted to *z*-scores, using appropriate values of assigned value and target standard deviation selected on the basis of fitness-for purpose.

Both trials were also used to assess the measurement uncertainty in sampling when different samplers are sampling a contaminated site either using a common protocol (CTS) or different protocols (SPT).

The feasibility and usefulness of CTS has been demonstrated by applying the methodology at the two experimental sites. The performance of two sampling protocols has been evaluated and compared with criteria based on fitness-for-purpose considerations. The “W” design used in the first CTS trial proved to be fit-for-purpose for Cu (element with concentrations close to the natural background at this site) but suspect for Pb (element with high concentrations at the site). The 20 m regular grid chosen for the

protocol in the second CTS was not fit-for-purpose of estimating mean concentrations (to within 20%) for Pb, Zn, and Cd, but is fit-for-purpose for Co. at this site. From both CTS it was clear that the effectiveness of a sampling protocol can be assessed using the CTS only for a specific site and further experimentation is needed before any attempt for generalisation about the potential suitability of particular protocols at contaminated sites.

The main factor affecting the suitability of the sampling protocol is the degree of heterogeneity of the sampling target. It should be noted that even for analytes with fairly homogeneous distribution in the sampling target, the sampling protocols have been applied at only one level, e.g. close to background concentrations of Cu in soil for the first CTS. Ideally the collaborative trial in sampling should be repeated on several sampling targets with different analyte concentrations. In this way the sampling protocol could be validated at different concentration levels.

Estimates of the uncertainty associated with the sampling protocol could also be made. In the instance of Pb the large sampling precision dominates the total uncertainty at both sites. The main reason is the great degree of heterogeneity of the sample targets in respect to the Pb concentration in soil. Within-sampler precision is as big as between-sampler precision and consequently the between-sampler variance does not contribute any extra uncertainty for Pb.

On the contrary there is a clear effect of the within-sampler and between-sampler variances on the estimation of the measurement uncertainty for the more homogeneously distributed elements. Both of these variances contribute to the total measurement uncertainty of the mean concentration for these elements. The magnitude and the proportion of their contribution were quantified using the data from the CTS.

Proficiency testing has been shown to be a concept that is usefully applicable to the practice of sampling. Inter-organisational errors in sampling can be estimated and the performance of the participants can also be assessed. Subsequently any need for improvement of the used sampling procedures can be identified.

Assigned values in normal circumstances will have to be based on the consensus of the participants in each round - it seems unlikely that the result of an expert organisation would be generally acceptable as an assigned value. Organisations exceeding acceptable limits for some of the analytes were detected in the first SPT, although in this

instance the problems for one organisation were demonstrably analytical rather than due to sampling. The general agreement between the participants was improved in the second SPT compared with the first one.

The detection and separation of the sources of error is also possible. The distribution and the analysis of a common reference material proved to be very useful for the identification and quantification of the analytical error in the first SPT. The analysis of all samples in one laboratory was chosen for the elimination of analytical bias between the participants in the second SPT. This does however go against the general philosophy of proficiency testing where the participants are responsible for all ^{components} of their measurements.

The effect of heterogeneity of the sampling target has been well recognised. Its effects limit the analogy with proficiency testing in analysis and also makes the results of such trials on contaminated land to be site specific. However, since its effect on the results is recognised it is a matter of continuing practical experience on sampling proficiency tests before it could be taken into account when scoring the performance of the participating organisations.

Sampling proficiency testing can be used as a method for the quantification of measurement uncertainty in sampling. In spite the high cost of this method it is possible to gain valuable information. Ideally, rounds of sampling proficiency tests have to be repeated on a regular basis so as to validate the methodology and alert the interested bodies on the effect of sampling error in contaminated land investigations.

10.4.2 Four Methods for the Estimation of Measurement Uncertainty in Field Sampling

A quantitative methodology for the estimation of measurement uncertainty has been suggested, by adapting methods devised originally for estimating analytical uncertainty to estimate sampling uncertainty. In this project field sampling and chemical analysis have been considered to be two parts of the 'measurement' process. Their combined contribution to the estimated uncertainty has been recognised and quantified by several methods throughout the experiments of this project.

Specifically, four methods for the estimation of this uncertainty have been considered depending on the manner the samples are collected:

- (i) for single sampler/single protocol
- (ii) for single sampler/multiple protocols
- (iii) for multiple sampler/single protocol
- (iv) for multiple sampler/ multiple protocols

With the first method it is possible to estimate sampling and analytical precision by taking duplicate samples at some proportion of the sampling locations, typically 10% (Ramsey et al., 1992). The duplicate samples are not taken at exactly the same position as the original, but displaced by a distance that represents the uncertainty of locating the sampling point, by the particular surveying technology used. Examples of the application of this method can be considered the estimates of uncertainty of metal concentrations in soil in the two initial surveys at the Derbyshire site and at Hounslow Heath. The three components of the variation (analytical, sampling and geochemical) are then separated using robust analysis of variance (Analytical Methods Committee, 1989). The sampling uncertainty, or within-location variance, will be partially due to small scale geochemical variation within the location, but represents the uncertainty in all of the samples that could be taken from that 'location' as specified, in this case within a two metre radius. Analytical bias can be also added into these estimates of uncertainty. However, this method does not allow for bias originating from the sampling process. The bias between different sampling protocols can be assessed by Method 2, and the bias between different samplers with the same protocol by Method 3.

Method 2 was applied to the Derbyshire site using four sampling protocols. The bias between the mean values of Pb concentration estimated by each protocol was found not to be statistically significant for the number of samples taken. This was primarily because of the large uncertainties (i.e. standard error) on each of the mean values, which probably originated from the heterogeneity of the site. The estimates of the uncertainty for the concentration values at each location were similar for all four sampling protocols.

Method 3 of estimating uncertainty is based on the application of a Collaborative Trial in Sampling (CTS). One of the original objectives of the CTS was to validate a particular sampling protocol. It is also possible however, to use the results of a CTS to estimate the uncertainty of measurements resulting from its application at a particular site. Examples of this method are the two CTS in Derbyshire and in Hounslow Heath.

Although essentially a 'top down' approach, the experimental design did allow the separate estimation of the uncertainty due to the chemical analysis within the one laboratory. For Pb this value is 7.4% for the first CTS and 16% for the second CTS. The spatial heterogeneity of the sites was the dominant factor in controlling the uncertainty for Pb, rather than differences between samplers. Some contribution of different sampling bias of the various samplers was observed for Cu and Co in the first and second CTS respectively. These elements were more homogeneously distributed in the soil and the uncertainty for multiple samplers was higher than that for a single sampler. Although this CTS methodology was applied here to estimate the uncertainty on mean concentrations for a whole site, it could equally well be applied to estimate the uncertainty of the concentration at particular locations within a site.

Method 4 was applied at the two experimental sites by organising Sampling Proficiency Tests (SPT). For the application of the SPTs in Derbyshire and in Hounslow Heath, the objective addressed by all nine sampling protocols was the estimation of the mean metal concentration at the sites. Although a SPT gives the most complete estimate of uncertainty is also the most expensive method and the most difficult to set up. The scatter in the estimates of the mean metal concentration by this method is affected by the site heterogeneity. The scatter of the estimated concentrations were much higher for Pb than for more homogeneously distributed elements such as Cu and Co. The problem of separating the analytical error from that originating from sampling had been overcome by distributing to the participants a common sample to be analysed together with their samples from the site in the first SPT. In the second SPT all the analyses of the samples was done in the one analytical run, in one laboratory.

10.4.3 Probabilistic Classification of Contaminated Land

The findings of the suggested methodology for the estimation of measurement uncertainty can be incorporated to give a solution to the problem of the classification of contaminated land. The current method used to classify contaminated land is to compare element concentrations measured in the soil against a threshold value. The problem with such a deterministic approach is that it ignores the existence of measurement

uncertainty. A probabilistic classification of contaminated land has been suggested, using the findings of this research. This, takes into account the uncertainty of a concentration measurement as well as its value.

Four categories can be distinguished:

1. Category 'Uncontaminated' where all of the range of the uncertainty is lower than the threshold value (T). The probability of this site being contaminated over the threshold is therefore <0.025 (i.e. 2.5%).
2. 'Possibly contaminated' category, where although the measured concentration value is less than the threshold, the upper side of the range of the uncertainty does cover the threshold value. In this case the probability of this site being contaminated over the threshold is therefore between 0.025 and 0.50 (i.e. 2.5 - 50%).
3. Category 'Probably contaminated' where the measured concentration exceeds the threshold value, but there is still a probability that the true mean value is less than the threshold (0.025 - 0.50). For this category the probability of this site being contaminated over the threshold is therefore between 0.50 and 0.975. (50 - 97.5%).
4. Category 'Contaminated' with probability of the site being contaminated over the threshold is >0.975 , although on the lower boundary of this class there is still a very small chance (2.5%) that the true concentration value at the site is under the threshold value.

The probabilities quoted here are given for an extended uncertainty with a coverage factor of two. It is equally valid to select a different coverage factor that reflects a different probability of misclassification that is considered more appropriate for a particular site. The appropriate coverage factor could give an indication of the degree of heterogeneity in a specific site. The probabilities are also given assuming single estimates of concentration being made at each sampling location. If 'n' multiple measurements have been made at a location, then the uncertainty value for that site should be divided by $\sqrt{n}$, before the classification is made.

10.5 Recommendations for Further Study

This research has demonstrated the importance of the estimation of sampling uncertainty for contaminated land investigations. However, several aspects of

uncertainty have yet to be explored. Some recommendations for further study are given below.

The feasibility of inter-organisational trials in sampling has been demonstrated in this research. Further experimentation is needed for the validation of this methodology. One objective could be to evaluate sampling targets that could be considered for the development of the initial idea of establishing reference sampling targets (Thompson and Ramsey, 1995). These would be analogous to the reference materials used in analysis. The main consideration for these targets would be the elemental concentrations of the contaminants.

More general objectives could be to conduct ongoing sampling trials at sites with different levels and types of contamination (e.g. organic contaminants) to deduce general conclusions from the trials. Even for analyte concentrations within a narrow range, the repetition of the trials would be a good practice.

The problem of the heterogeneity of the sampling targets with respect to the metal concentrations also needs to be further studied. Heterogeneity is a basic characteristic of contaminated land sites. The use of portable equipment, e.g. portable XRF is very promising for the study of small scale heterogeneity of the sampling target. Estimates of the degree of heterogeneity have to be integrated with the data during the interpretation of the results of the inter-organisational trials. The calculation of z-scores for example could be modified in such way to allow for the factor of heterogeneity. Attention has also to be given however, to the fitness-for-purpose criterion, that must not be ignored.

It has been mentioned early in this research (Chapter Two) that there are two approaches for the evaluation of measurement uncertainty, by using either a “bottom-up” or a “top-down” method. The “bottom-up” approach requires the recognition of all sources of errors contributing to the total uncertainty (EURACHEM, 1995). On the other hand with the “top-down” approach for the estimation of measurement uncertainty, errors are viewed from a higher level enabling the user to assess measurement uncertainty based on the results of inter-organisational trials (Analytical Methods Committee, 1995). Comparisons of the estimated uncertainty between these two methodologies should be made, and this would give very useful information related

to the subject. This would require parallel processing of the data according to the two approaches and would give a more complete estimate of uncertainty.

An other important issue raised in this research is the problem of the spatial uncertainty. It is well understood that one of the main concerns in contaminated land investigations is the location of contamination hot spots. The estimation of the uncertainty in locating these hot spots is definitely an aspect needing to be further studied. Integration of these new methods with traditional geostatistical techniques should prove very useful in this instance. A very interesting task would be the integration of the results of the inter-organisational trials with the geostatistical data. Inter-organisational trials with the objective of estimating spatial uncertainty would be more difficult to organise but would provide very useful feedback to practitioner of contaminated land investigations.

Finally, the consideration of different sampling media for the estimation of the uncertainty generated from sampling, seems a very interesting potential subject for further experimentation. Environmental sampling of water and air are potential sampling targets for applying inter-organisational trials. Such trials will have to be modified however, to take temporal variation into account which is a crucial factor in the sampling of these media.

When such methods have been devised and applied to a wide range of environmental media, scientists will have a much clearer idea of the uncertainty that will continue to arise from their initial acts of measurement; that of sampling.

REFERENCES

- Alloway, B.J. (1995) *Heavy Metals in Soils* (second edition), Blackie, Glasgow, p. 354.
- Alloway, B.J. and Ayres, D.C. (1997) *Chemical Principles of Environmental Pollution* (second edition), Blackie Academic and Professional, London, p. 5.
- Analytical Methods Committee (1989) Robust statistics-How not to reject outliers: Part 1. Basic concepts, *Analyst*, **114**, 1693.
- Analytical Methods Committee (1995) Uncertainty of measurement: Implications of its use in analytical science, *Analyst*, **120**, 2303.
- Barcelona, M.J. (1988) "Overview of the Sampling Process", in: Keith, *Principles of Environmental Sampling*, L.H. Keith, Ed., Washington, D.C., p.3-23.
- Black, S. C. (1988) "Defining Control Sites and Blank Sample Needs", in *Principles of Environmental Sampling*, L.H. Keith, Ed., American Chemical Society Washington, D.C., p. 109-117.
- Bosman, R. (1994) "International Standards for Site Investigation" *Site Investigations and Clean- up Technologies for Contaminated Sites*, (5/ 6 October 1994, London).
- British Geological Survey (1981) Geology map of South London, Sheet 270, solid and drift edition, London.
- Cochran, W.G. (1977) *Sampling Techniques*, John Wiley & Sons, Inc. (third edition), New York.

- Davis, J.C. (1986) *Statistics and Data Analysis in Geology*, John Wiley and Sons, New York.
- De Bievre, P., Baumann, S., Schott, R., Dalton, J.C., Berg, R., De Regge, P., Gorgnyl, T., Hoflich, M., Deron, S., Kuhn, e., Mainka, E., Bradshaw, R.P., and Pietri, C. (1987) 1988 target values for random uncertainties in sampling and element assay of nuclear materials, *ESARDA Bulletin*, No. 13, 8.
- Deutsch, C.V., and Journel, A.G. (1992) *GSLIB: Geostatistical software Library and User's Guide*, Oxford University Press, New York.
- Dixon, W.J. (1950) Processing data for outliers, *Biometrics*, **9**, 74.
- Ferguson, C.C. (1992) The Statistical Basis for Spatial Sampling for Contaminated Land, *Ground Engineering*, **25**, 34.
- Flatmann, G.T., Englund, E.J. and Yfantis, A.A. (1988) "Geostatistical Approaches to the Design of Sampling Regimes", in: Keith, L.H., Ed., *Principles of Environmental Sampling*, American Chemical Society, Washington, D.C., p. 73-84.
- Garrett R.G. (1969) The Determination of sampling and analytical errors in exploration geochemistry, *Economic Geology*, **64**, 568.
- Goldschmidt, V. M. (1954) *Geochemistry*, Clarenton Press, Oxford.
- Gotway, C.A., and Rutherford B.M. (1994) "Stochastic Simulation for Imaging Spatial Uncertainty: Comparison and Evaluation of Available Algorithms", in *Geostatistical Simulations*, M. Armstrong and P.A. Dowd (eds.), Kluwer Academic Publishers, The Netherlands.
- Gy, P.M. (1979) *Sampling for Particulate Materials: Theory and Practice*, Elsevier, Amsterdam.

- Gy, P.M. (1992) *Sampling of Heterogeneous and Dynamic Material Systems, Theories of Heterogeneity, sampling and Homogenizing*, Elsevier, Amsterdam, p.568.
- Hoffmann, P. (1994) General Aspects of Environmental Sampling, in: B. Markert, Ed., *Environmental Sampling for Trace Analysis*, VCH, Weinheim, 11-71.
- Holcombe, L.J. (1988) "Relations of Sampling Design to Analytical Precision Estimates", in: Keith, L.H., Ed., *Principles of Environmental Sampling*, American Chemical Society, Washington, D.C., p. 395-408.
- Horwitz, W. (1988) Protocol for design, conduct and interpretation of collaborative studies, *Pure Applied Chemistry*, **60**, 855.
- International Vocabulary for Basic and General Terms in Metrology (VIM)*, (1993) ISO, Geneva, 2nd ed.
- Isaaks, E.H. and Srivastava, R.M. (1989) *Applied Geostatistics*, Oxford University Press, New York.
- ISO 3534- 1: 1993 (E/F), Statistics, *Vocabulary and Symbols- Part 1 Probability and General Statistical Terms*.
- ISO Guide 30: 1992, *Terms and Definitions Used in Connection with Reference Materials*.
- Journel, A.G. and Huijbregts, C.J. (1978) *Mining Geostatistics*, Academic Press, London.
- Kapumpa, C.C. (1995) *The Relationship Between Geology and Grade Distribution in the Baluba Centre Limb, Zambian Copperbelt*, PhD Thesis, University of London.

- Keith, L.H., Ed. (1988) *Principles of Environmental Sampling*, American Chemical Society, Washington, D.C.
- Keith L.H. (1991) *Environmental Sampling and Analysis: a Practical Guide*, Lewis Publishers, Chelsea MI, USA.
- Kratochvil, B., Wallace, D. And Taylor, J. K. (1984) Sampling for chemical analysis, *Analytical Chemistry*, **56**, 113R-129R.
- Kuharic, C. A., and Cole, W. H. (1995) FPXRF, EDXRF and ICP comparison of Pb contaminated soils from Leadville, Colorado, *Advances in X-Ray Analysis*, **38**, 725.
- Kuharic, C. A., Cole, W. H., Singh, A. K., and Gonzales, D. (1993) X-ray fluorescence survey of lead contaminated residential soils in Leadville, Colorado: a case study, *EPA Report*, EPA/600/R-93/073.
- Maskall, J., and Thornton, I. (1993) Metal contamination of soils at historical lead smelting sites, *Land Contamination and Reclamation*, **1**, 92.
- Maskall, J., Whitehead, K. and Thornton, I. (1995) Heavy metal migration in soils and rocks at historical smelting sites, *Environmental Geochemistry and Health*, **17**, 127.
- Ministry of Agriculture, Fisheries and Food (1979) Leaflet 655 "Sampling soils for Analysis", London.
- Moore, J. (1994) "Site Assessment Techniques", *Site Investigations and Clean- up Technologies for Contaminated Sites*, (5/ 6 October 1994, London).
- Olea, R. (1991) *Geostatistical Glossary and Multilingual Dictionary*, Oxford University Press, New York.

- Ramsey, M. H., Thompson, M., and Hale, M. J. (1992) Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance, *Geochemical Exploration*, **44**, 23.
- Ramsey, M.H., Watkins, P.J., and Sams, M.S. (in press) Estimation of measurement uncertainty for in-situ bore hole determinations using a geochemical logging tool.
- Ripley, B.D., and Thompson, M. (1987) Regression techniques for the detection of analytical bias, *Analyst*, **112**, 377.
- Rose, A. W., Hawkes, H.E., and Webb, J.S. (1979) *Geochemistry in Mineral Exploration*, Academic Press, London.
- Stock, H.D. (1995) On-Site Versus Off-Site Analysis; A Matter of Balance Between Heterogeneity and Accuracy, in: W.J. van den Brink, R. Bosman and F. Arendt (eds.), *Contaminated Soil '95*, Kluwer Academic Publishers, Netherlands.
- Swift, R. P. (1995) Evaluation of a field-portable X-ray fluorescence spectrometry method for use in remedial activities, *Spectroscopy*, **10**, 31.
- Thompson, M. And Howarth, R. J. (1976) Duplicate analysis in geochemical practice. Part 1: Theoretical approach and estimation of analytical reproducibility, *Analyst*, **101**, 690.
- Thompson, M. and Maguire, M. (1993) Estimating and using sampling precision in surveys of trace constituents of soils, *Analyst*, **118**, 1107.
- Thompson, M. and Ramsey, M.H. (1995) Quality concepts and practices applied to sampling - an exploratory study, *Analyst*, **120**, 261.

- Thompson, M., and Walsh, J.N. (1989) *A Handbook of Inductively Coupled Plasma Spectrometry*, Blakie, Glasgow.
- Thompson, M., and Wood, R. (1993) the international harmonized protocol for the proficiency testing of (chemical) analytical laboratories (technical report), *Pure and Applied Chemistry*, **65**, 2123.
- Thornton, I. (1983) *Applied Environmental Geochemistry*, Academic Press, London.
- UK Department of the Environment (1987) Interdepartmental Committee on the Redevelopment of Contaminated Land. Guidance on the assessment and redevelopment of contaminated land, DOE, London.
- Vincent, H. A., and Boyer, D. M. (1995) An examination of an XRF method used for the determination of lead in urban soil and dust samples, *ASTM Special Technical Publication*, STP 1226, 215.
- Visman, J. (1969) A general theory of sampling, *Material Research and Standards*, 9- 13.
- Watson, W., Walsh, J.P., and Glynn, B. (1989) On-site x-ray fluorescence spectrometry mapping of metal contaminants in soils at Superfund sites, *American Laboratory*, Boulder, CO.
- Webster, R. (1977) *Quantitative and Numerical Methods in Soil Classification and Survey*, Clarendon Press, Oxford.
- Yee, C.S. (1995) *Reconnaissance Survey of Potential Heavy Metal Contamination of Soils on Hounslow Heath, West London*, Msc Thesis, University of London.

APPENDIX 1

Analytical Method

Acid digestion of rocks, soils and sediments for multi-element analysis

1. APPARATUS

1) test tube; 2) wire test tube rack (plastic coated); 3) stainless steel test tube rack; 4) aluminium heating block (deep, 252 holes); 5) shallow aluminium heating block (315 holes); 6) Oxford dispenser; 7) centrifuge tube; 8) vortex tube mixer; 9) balance (top pan); 10) centrifuge

2. REAGENTS

1) deionised water; 2) nitric acid (A.R. 70% w/w); 3) perchloric acid (A.R. 60% w/w); 4) hydrochloric acid 5M (A.R.): dilute 430ml conc. HCl (36% w/w) to 1,000ml with DIW

3. PROCEDURE

1) weigh 0.100g (± 0.001) samples as per weighing list into clean dry test tube in wire test tube racks, using top pan balance
2) add 4.0ml nitric acid into each tube from Oxford dispenser
3) add 1.0ml perchloric acid into each tube from the dispenser
4) place tubes in the aluminium heating block and switch programmer to 'MANUAL' mode, then set up as follows:

rise rate (sec/deg)	dwelt time (hrs)	dwelt temperature ($^{\circ}\text{C}$)
001	3.0	50
001	3.0	150
001	18.0	190
001	0.1	195

5) turn on the fume cupboard and switch programmer to 'AUTO'
6) when attack cycle complete, check each tube if residue is dry and transfer tubes to the stainless steel racks
7) when tubes are cool, add 2.0ml hydrochloric acid (5M) to each tube from an Oxford dispenser
8) transfer tubes to wire racks and allow to cool
9) add 8.0ml deionised water from an Oxford dispenser and mix each tubes using a vortex mixer
10) decant into polystyrene tubes and cap, if necessary, centrifuge at 2,000 rpm for 2 minutes
11) chemical analysis for general elements using ICP-AES

APPENDIX 2
Total geochemical data

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.1 Regular grid, composite samples.

Sample	Li	Na	K	Be	Mg	Ca	Sr	Ba
R11C	32	205	3740	1	2130	31087	86	1563
R12C	70.3	330	7170	1.78	3940	4570	75.6	1120
R13C	49	278	5380	1.47	3270	4030	108	2210
R14C	50.5	268	6200	1.535	3820	8300	112	1980
R15C	28.9	175	2580	1.405	1570	4570	99.5	1800
R16C	25	185	2617	1	1510	5527	128	2085
R17C	32.7	190	3350	1.09	1950	2830	51.9	1290
R18C	25.6	130	2790	0.86	1830	2710	84.6	2420
R21C	58	295	6510	1.565	3350	3960	47.1	430
R22C	36.8	216.5	4345	1.1625	2415	5355	109.65	1388.5
R23C	54.3	295	5940	1.59	3580	5720	88.5	2050
R24C	61.2	313	6730	2.06	4500	7030	106	2120
R25C	31.6	218	3360	1.155	1880	9240	103	1850
R26C	26.5	138	2910	0.98	1680	4720	140	2330
R27C	59	324	6680	2	3623	3880	58	862
R28C	31.1	233	4070	0.985	2160	3180	75.4	2320
R32C	51.9	313	6080	1.47	3030	4300	62.3	966
R33C	63.4	348	7240	1.63	3840	5900	78.5	1540
R34C	58.3	320	6190	1.76	4040	5770	78.2	1710
R35C	61.5	298	8620	2.185	5540	4500	83.3	1500
R36C	51	264	5973	2	3213	3860	87	988
R37C	36	228	4250	1.13	2280	3100	84.9	1970
R44C	42	227	4540	1	2517	5200	92	1458
R45C	26.3	170	2870	1.12	1710	5580	136	1880
R46C	36.6	250	4450	1.215	2420	3800	93.3	2160
R47C	36.5	288	4770	1.395	2400	4500	89.6	1880
R54C	43.8	298	4940	1.17	2660	9220	61.7	1200
R55C	28.4	153	2560	0.94	1550	4550	79.1	2160
R56C	69	339	7153	2	4003	7913	62	587
R57C	31.3	268	4140	1.06	2190	66300	118	2060
R65C	33.5	195	3510	1.185	1970	4180	160	2500
R66C	21.5	118	2110	1.4	1300	7680	203	2820
R67C	29	204	3263	1	1760	3050	52	1267
R75C	24	159	2783	1	1600	7353	279	2570
R76C	35.1	163	4070	1.23	2210	7920	172	2100
R84C	50.1	265	5950	1.35	3140	976	50.6	638
R85C	58.9	320	6400	1.74	3360	4890	69	1290
R86C	55.9	298	5770	1.755	3520	6480	52.2	891
R94C	61.2	278	6750	1.685	3800	1790	50.7	822
R95C	61.2	320	7090	1.61	3510	3950	66.6	883

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.1 Regular grid, composite samples. (continue)

Sample	Al	La	Ti	V	Cr	Mo	Mn	Fe
R11C	40233	24	921	49	25	7	969	23400
R12C	73200	31.3	1180	65.5	35.5	5.5	1000	31800
R13C	55700	26.5	1010	52.8	27	5.5	857	25200
R14C	57700	29	1080	57.5	31	5.5	543	23800
R15C	37200	22.3	941	43.3	26	5.5	1010	28000
R16C	32933	23	776	45	22	6	784	19233
R17C	39600	22	952	55.5	30	5.5	813	22800
R18C	32900	21.8	1010	48.3	29.5	4	437	17200
R21C	63100	27.8	1170	58	37	5	1580	32600
R22C	45300	22.5	980	46.8	23	5	657	21450
R23C	64000	31.3	1110	62.5	28	7.5	766	28500
R24C	69900	36.8	1320	68.5	40	8	789	32100
R25C	38700	21.3	904	44	22	3.5	957	23700
R26C	35400	18	797	41.8	26	6.5	569	22500
R27C	67167	32	1277	66	33	7	1080	29267
R28C	40700	24.3	1100	51.8	29	4.5	426	21000
R32C	57700	29.3	1100	54.8	37.5	5.5	930	28200
R33C	72900	33.5	1240	67	35	6	1070	33500
R34C	64900	31.3	1220	61.3	33.5	5.5	671	30000
R35C	67900	36.5	1150	66.3	28.5	5	578	37200
R36C	58100	30	1127	57	32	7	1128	31067
R37C	45400	29	1090	60.5	36	5.5	851	27100
R44C	49000	26	1103	51	34	6	1124	28167
R45C	35400	22.8	977	41.8	31.5	5.5	582	21600
R46C	45600	25	1090	48.3	28	4	802	24000
R47C	49100	28.5	1110	60.5	37	7.5	569	23200
R54C	52500	26.8	1220	53.8	33.5	6.5	714	24400
R55C	37400	19.8	926	44.3	21.5	3	1040	24200
R56C	73033	34	1207	65	34	6	1230	32367
R57C	38100	23	925	54.3	31	6	559	19500
R65C	44200	22	941	46.5	25.5	7	615	20700
R66C	28900	20.8	686	35.3	22.5	5.5	718	17300
R67C	38233	23	950	58	27	7	745	22233
R75C	31033	18	602	37	27	7	475	17233
R76C	44400	23.8	960	48.3	22	6	556	20500
R84C	55800	26.3	1080	54	22.5	4.5	750	28300
R85C	63300	30.8	1270	59.5	37	5.5	1440	28900
R86C	65000	30.8	1190	65	25.5	4.5	987	29200
R94C	63600	29.5	1150	60.5	26.5	5	882	31400
R95C	67500	31.3	1240	63.5	34	5	1210	27000

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.1 Regular grid, composite samples. (continue)

Sample	Co	Ni	Cu	Ag	Zn	Cd	Pb	P
R11C	10	19	28	1	501	6	3263	1134
R12C	13.8	27.5	24.8	0.4	279	2	4360	947
R13C	13	22	26.1	0.6	640	3	8900	988
R14C	12	25.5	34.2	1	904	6.25	12400	1100
R15C	17	23	42.8	1	1160	13.8	4490	1030
R16C	11	16	27	1	632	5	8912	1059
R17C	11	16	26.5	0.5	278	2.25	3280	1080
R18C	6.75	14	19.6	0.3	442	3.75	3990	779
R21C	15.8	22.5	20.7		193	0.5	1850	910
R22C	10.75	19.25	24.85	0.6	484.5	3.375	4917	805.5
R23C	14.3	25.5	24.2	1.3	399	3	7280	941
R24C	18.8	35.5	40	1.2	911	9.25	3740	874
R25C	12.3	19	31.1	0.5	735	5.75	7230	1190
R26C	12	20.5	28.6	1.3	1080	5.75	12900	1310
R27C	17	28	26	1	325	3	1330	741
R28C	7	16	22.9	0.3	438	3.5	3850	1450
R32C	11.5	27	20.9		278	2	4620	935
R33C	13	26.5	26	0.9	465	3.5	7590	1110
R34C	12.5	28.5	27.5	0.4	471	3.5	5830	866
R35C	16.8	42.5	35.7		513	4.5	2420	854
R36C	15	24	27	1	451	4	4750	1087
R37C	12.8	20	25.6	0.8	510	5	2080	1230
R44C	15	21	29	1	596	5	5630	1153
R45C	10.5	21.5	27.6	1.9	729	6.75	13100	1300
R46C	13.5	19.5	24.2	0.6	585	4	6350	1160
R47C	10	20	27.6	1.3	558	6	4870	1530
R54C	11	22	23.1	1	287	2.75	4280	914
R55C	10.8	19	30.6		621	6	3310	890
R56C	16	29	25	1	250	2	1903	835
R57C	7.75	21.5	32.2	0.5	666	8.25	2820	1870
R65C	9.25	19	25.5	1.1	632	4.75	15700	1250
R66C	10.8	19	31.5	1	962	7.75	15800	1220
R67C	11	16	25	0	352	4	2643	959
R75C	7	15	27	1	1253	6	30033	1957
R76C	12.3	21	30.9	0.7	807	6.5	9200	1060
R84C	13.3	24	22.8	0.3	171		1030	715
R85C	17	29.5	27	0.5	444	3.75	2450	1010
R86C	16	28	27.8	0.3	417	3.75	1880	718
R94C	17	27.5	24.2		254	1.75	1020	666
R95C	18.5	26	24.8	0.6	385	4.25	1100	746

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.2 Regular grid, single auger samples.

Sample	Li	Na	K	Be	Mg	Ca	Sr	Ba
R11S	34	218	4007	1	2250	30990	106	1670
R12S	70.9	348	7480	1.86	4050	4500	69.6	996
R13S	47.6	253	5550	1.43	3270	3950	99.8	2220
R14S	49.6	290	6270	1.615	3890	7890	108	2150
R15S	29.2	173	2630	1.375	1580	3710	99.2	1930
R16S	41	241	4683	1	2700	11928	95	1627
R17S	31.8	175	3300	1.075	1870	2650	78.3	1320
R18S	23.8	175	2630	0.835	1770	2930	67.7	2250
R21S	56.8	318	6630	1.57	3380	7100	52	531
R22S	41	241	4707	1	2705	12799	107	1696
R23S	59.4	285	6410	1.555	3780	6320	70.6	1500
R24S	59.2	303	6410	1.91	4420	6080	117	1760
R25S	30.2	218	3190	1.17	1800	13400	121	1870
R26S	19.5	130	2120	0.685	1520	110000	159	2010
R27S	59	315	6420	2	3560	3840	62	1058
R28S	29.8	193	4060	1.045	2200	3590	74.4	2120
R32S	48.8	335	5670	1.425	2980	4720	56.1	888
R33S	56	303	6460	1.595	3480	6540	79	1530
R34S	58.6	300	6090	1.815	4130	3920	115	1980
R35S	61	355	9310	2.365	5890	6080	77.6	1410
R36S	45.2	224	5210	1.395	2890	4345	104.15	1287.5
R37S	36.7	250	4340	1.185	2260	3310	98.4	2370
R44S	51	268	5707	1	3007	4603	70	926
R45S	26.8	208	2820	1.075	1740	6010	132	1520
R46S	34.3	200	4160	1.195	2350	3460	98.7	2100
R47S	35.6	240	4580	1.475	2340	4540	117	2030
R55S	25.8	128	2520	0.96	1520	3200	58.7	1630
R56S	42.3	253	4930	1.105	2690	8260	57.3	949
R57S	28.6	233	3770	1.115	2180	81800	142	2110
R65S	30.9	168	3320	1.145	1880	2910	119	1950
R66S	20.8	168	2020	1.25	1290	8240	202	2910
R67S	29	201	3213	1	1730	2813	53	1260
R75S	26.3	169	3087	1	1700	6390	234	2363
R76S	35.7	223	4180	1.225	2200	9380	184	2230
R84S	37.9	210	4510	1.1	2630	1330	35.3	547
R85S	57.2	268	6050	1.79	3350	4290	76.8	1440
R86S	56.9	315	5870	1.8	3440	4990	56.6	935
R94S	62.6	363	7340	1.775	3890	2240	52.2	800
R95S	58.7	293	6830	1.68	3540	4410	64.9	850

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.2 Regular grid, single auger samples. (continue)

Sample	Al	La	Ti	V	Cr	Mo	Mn	Fe
R11S	42333	23	926	50	25	8	610	21433
R12S	73500	32	1170	66.3	31.5	9	962	31700
R13S	55500	27.5	1030	53.3	32	6.5	850	25600
R14S	58300	30.3	1070	60.3	29	8.5	515	23600
R15S	38100	19.5	942	44	22.5	6.5	1430	28300
R16S	48660	26	1008	54	27	7	869	25912
R17S	39200	23.3	923	52.3	25	4	651	24000
R18S	30600	21.3	850	61.8	27	4.5	595	17800
R21S	64000	31.5	1180	59	33.5	6.5	1670	33200
R22S	48812	26	1004	53	27	7	827	25445
R23S	66900	30	1090	61.3	28	5.5	841	28600
R24S	64200	30	1200	64.3	32	7.5	763	31600
R25S	38000	21.3	941	44.8	27	5	1010	23500
R26S	26000	21.3	592	30	15.5	5	717	15300
R27S	65767	32	1253	65	33	7	1180	29833
R28S	39500	23.5	950	52.5	24	8	476	21500
R32S	54700	26.5	998	55	28.5	5.5	864	28200
R33S	66300	30.8	1130	63.5	34	8.5	1020	31300
R34S	63700	30.3	1200	61.5	26.5	9	570	30500
R35S	73800	42.8	1330	71.8	35.5	8	707	38400
R36S	52400	27.4	1081	53.65	30.5	6.5	1016.5	29250
R37S	47500	27.5	1220	63.3	40.5	8	1080	30000
R44S	57733	28	1173	57	35	7	1066	30733
R45S	34300	18.5	912	39	25.5	6	640	20900
R46S	41600	24.5	958	46.5	20	6	697	23100
R47S	45100	27.5	975	55	24.5	5.5	490	22700
R55S	34800	21.8	877	42.8	18.5	5	1620	27000
R56S	48700	25.5	999	48.8	22	7	731	23800
R57S	37200	24.8	887	56.3	34	8	557	19400
R65S	41100	20	893	45	18.5	7.5	543	21100
R66S	27500	22.8	623	33	16.5	6	728	16500
R67S	38367	23	930	56	27	6	840	21967
R75S	33333	18	682	38	25	6	414	17433
R76S	47600	21.5	1040	48.8	26.5	6.5	752	20200
R84S	43000	23.5	920	44.8	19.5	3.5	562	22700
R85S	61400	31.5	1150	60.3	31.5	7.5	1390	29200
R86S	69200	34	1360	68.3	35.5	6	1070	29000
R94S	66700	31.5	1250	65.3	33	7.5	1080	31300
R95S	65600	30.8	1220	65.8	29	8	1090	27500

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.2 Regular grid, single auger samples. (continue)

Sample	Co	Ni	Cu	Ag	Zn	Cd	Pb	P
R11S	9	20	27	1	513	5	7047	1272
R12S	14	27.5	25	0.6	302	2.75	3910	942
R13S	13	22	25.6	0.7	606	2.75	8930	889
R14S	13.3	26.5	35.5	1.4	799	4.75	11900	1150
R15S	15.8	21.5	35.2	1.2	949	9	4990	1060
R16S	12	22	27	1	529	5	5595	1062
R17S	11.3	14.5	23.2	0.3	259	2	3450	979
R18S	6.75	14	17.3	0.6	398	3.5	3920	960
R21S	16.3	23	20.3	0.4	192	1.25	1790	988
R22S	12	22	27	1	584	5	6683	1095
R23S	12.3	24.5	25.1	0.7	399	2.5	6380	920
R24S	18.5	33.5	44.5	0.8	1000	9.75	3370	850
R25S	11.8	18.5	28.5	0.7	678	6.25	6750	1030
R26S	7.75	11.5	18.6	0.7	691	4.75	11900	930
R27S	17	29	27	1	355	4	1733	777
R28S	7.75	16	23.2	0.7	417	4	3740	1820
R32S	11.5	25.5	23.2	0.4	268	1.5	4570	1050
R33S	13	24.5	24.3	1.2	356	3	7800	944
R34S	13.3	29	29.6	0.6	394	3.25	5270	852
R35S	17.8	44.5	38.6	0.8	395	3.25	1930	812
R36S	14.4	23	28.4	0.8	577	4.875	6430	1125
R37S	15	19.5	25.6	0.9	735	8	2480	1110
R44S	16	23	25	1	354	3	3850	1110
R45S	9.5	18	28	1.4	1450	17.3	13000	1220
R46S	13	19	24.1	0.5	609	3.75	6780	1010
R47S	9.25	21	28.5	0.5	503	5	4390	1490
R55S	11.3	17	28.6	0.4	537	6.25	2550	745
R56S	9.75	20.5	23.6		276	2.25	3720	956
R57S	8.5	20.5	31.1	1.1	689	8	3520	1700
R65S	8.5	17.5	25.5	1.1	574	4.25	13900	1160
R66S	10.3	16.5	30.5	0.8	900	7	15200	1200
R67S	11	15	23	1	363	4	2623	918
R75S	7	15	27	1	1063	6	24533	1853
R76S	11.5	19	30.1	0.6	799	6.25	8950	962
R84S	10	19.5	19.6		170	0.5	884	649
R85S	19.3	30.5	27.3	0.8	470	5.5	2290	909
R86S	15.8	26.5	25.9	0.5	304	2.75	1900	690
R94S	17	29	26.4	0.5	292	2.25	1010	731
R95S	18.3	27	26.1	0.8	380	3.5	1080	803

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.3 Herringbone design, single auger samples.

Sample	Li	Na	K	Be	Mg	Ca	Sr	Ba
H11	63.1	334	7280	1.615	3525	2900	48.35	345
H12	67.7	355	7550	1.835	4320	4620	58.9	540
H13	51.7	248	5580	1.56	3530	14200	223	1370
H14	53.5	328	6340	1.9	4130	4600	119	2050
H15	44.3	230	5580	1.585	3130	3430	66.5	1620
H16	33.1	185	3910	1.11	2350	5200	127	2260
H17	30.4	198	3000	1.04	1730	5600	80.3	1850
H18	24.7	208	2650	0.9	1740	2860	66.5	2180
H21	60.4	308	6640	1.65	3730	11500	56.8	558
H22	57.55	294	6780	1.57	3610	8690	47.6	397.5
H23	61.9	323	6680	1.795	3680	6370	79.4	1470
H24	51.6	278	5540	1.835	3840	5290	90	2330
H25	42.3	233	5700	1.645	3880	8740	96.5	2240
H26	28.3	155	2890	1.145	1800	3590	110	2040
H27	28.25	194	3360	1.1275	1820	4710	91.05	2245
H28	28	168	3580	1.04	1970	3490	90.7	2270
H33	60.15	322.5	6475	1.5775	3455	4065	65.4	1355
H34	53.5	335	5920	1.485	3230	5950	80.2	1730
H35	59.3	305	6780	2.07	4480	4950	118	2610
H36	30.7	210	3380	1.11	1960	6540	188	2300
H37	40.6	263	4760	1.31	2620	3550	151	1960
H45	38.4	243	4220	1.5	2800	7840	103	2120
H46	46.2	250	6430	1.775	3750	6970	97.3	2270
H47	36.5	248	4600	1.125	2230	3510	78.4	1960
H55	30	198	2720	1.01	1640	4210	98.2	2070
H56	25.6	195	2660	1.08	1750	18800	194	2080
H57	31.4	218	3790	1.03	1980	11800	80.6	1500
H65	40.1	263	4520	1.275	2400	4840	181	2300
H66	26.9	120	2210	0.995	1320	4530	113	1930
H67	40.3	285	5270	1.705	2810	7490	167	2150
H68	47.4	278	5610	1.45	2630	6010	87.7	1540
H75	47.9	298	6010	1.445	3020	4250	58	1050
H76	36.3	298	4080	1.215	2150	5000	138	2320
H77	44.2	285	4970	1.615	2810	4450	102	2070
H85	58	330	6330	1.69	3240	4880	54.4	739
H86	52.3	268	6130	1.835	2990	5900	90.4	1660
H87	59.2	298	6550	1.745	3580	5070	75.5	1910
H94	58.4	320	6720	1.82	3460	2560	50.7	711
H95	59.5	298	6660	1.555	3190	4170	58.8	744

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.3 Herringbone design, single auger samples. (continue)

Sample	Al	La	Ti	V	Cr	Mo	Mn	Fe
H11	67000	32.4	1245	61.8	33.75	5.5	1445	32950
H12	73800	34.3	1270	65.8	39.5	6	831	32900
H13	58800	28.3	1080	56.5	29.5	6	845	27900
H14	64000	34.3	1240	66.3	37	7.5	503	27100
H15	56200	30.8	1320	57.5	28.5	5.5	1180	31100
H16	38400	22	943	45	26	4.5	617	24200
H17	38200	20.8	1050	55	32	6	893	21700
H18	31500	22.5	974	42.5	27.5	5	331	18000
H21	67200	31.8	1240	60	33	6	855	32500
H22	61850	30.55	1205	59.8	33.5	4.75	1340	35450
H23	71000	37.8	1390	67.3	40.5	8	987	30500
H24	61900	32	1270	62	36.5	6.5	824	30500
H25	55500	29.8	1080	61.8	27	7	531	27500
H26	35200	18.8	870	38.8	15	5	501	23400
H27	35150	21.4	916	54.55	25.25	6.25	621	21850
H28	36300	20.3	959	48	28.5	5	528	20500
H33	65050	30.75	1085	61.4	33.5	5.5	1092.5	32600
H34	61200	29.8	1260	58.5	37	6.5	1350	27500
H35	68300	36.3	1380	67.8	39	7	422	29200
H36	39700	16	823	44.3	32.5	6.5	553	20900
H37	49800	25	1180	62.5	34.5	6	838	28900
H45	48200	26	1140	51.8	32.5	6	725	26000
H46	57500	34.8	1310	62	38	8.5	624	28900
H47	46100	24.5	1090	49	30	5.5	516	20500
H55	38800	24.8	945	43.8	21	5	853	22700
H56	37500	20	868	41.5	28.5	5.5	754	20300
H57	41200	23.5	1090	48	36	5.5	562	19100
H65	49600	19.5	1040	52	33.5	8	1080	22500
H66	35200	17	808	39.3	16	6.5	939	20700
H67	51300	28	1090	55.3	38	6.5	646	21900
H68	56900	36.3	1330	65	37.5	6	284	18800
H75	56500	33	1270	55.5	38.5	7	501	23700
H76	47300	20.5	1030	52.3	28	7	516	21800
H77	56700	35	1160	61.3	32	7	588	23800
H85	64900	33.5	1290	63	36	7.5	911	29700
H86	58200	28.5	1190	59.5	37.5	5.5	700	24100
H87	71800	32.8	1380	71.3	35.5	7	613	28600
H94	63400	32.3	1220	63.3	36	6.5	683	31600
H95	65800	31.3	1430	61	35.5	6	1030	25000

**Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.**

2.1.3 Herringbone design, single auger samples. (continue)

Sample	Co	Ni	Cu	Ag	Zn	Cd	Pb	P
H11	16.55	25	21.25	0.55	191	1	1395	1060
H12	14.8	30	24.5	0.5	230	1.25	1860	809
H13	13.5	27.5	32.3	1	2810	32	7120	912
H14	14.3	31.5	32.8	1.5	952	8.5	9670	928
H15	18	29.5	38.4	0.6	578	4.5	2970	802
H16	11.5	21	36	0.8	724	5	7740	1120
H17	11	14.5	24.2	0.5	682	7.25	3900	1090
H18	7	14	18.1	0.3	345	2.5	4000	858
H21	15.3	24.5	21.5	0.3	227	1.5	1240	884
H22	15.4	28.25	21.5	0.35	188.5	1	1570	962
H23	14.8	27.5	25.3	1.3	384	4.25	4470	917
H24	14.8	29	29.4	1	496	3.75	7050	853
H25	14	30	32.6	1.4	616	4.75	6440	974
H26	11.5	21	33.1		688	5	5580	960
H27	9.775	19	28.95	0.45	625.5	5.75	4420	1120
H28	8.5	17	23.8	0.5	482	5	4030	1330
H33	14.05	24.75	24.15	0.95	396.5	3	8430	952
H34	14.3	23	22.4	0.8	388	3.25	6690	817
H35	11	31.5	31.3	0.8	554	4.75	7380	1020
H36	8.5	17	28	1.2	1270	6.5	20400	1490
H37	14.3	24.5	35.2	0.6	761	9.5	2660	1100
H45	12.8	25.5	32.3	1.1	675	5.25	9040	1020
H46	15.8	31.5	31.8	0.9	467	3.5	4370	1010
H47	9.75	16	24.1	0.2	410	3.5	4560	1080
H55	10.3	18	28.2	0.5	636	6	5420	1120
H56	9.25	20	31.8	1.1	1410	13	16300	1360
H57	8.5	16	27	0.6	488	5.25	3020	1790
H65	10.8	20	24.8	0.9	630	2	14100	1140
H66	11.8	17.5	25.8	0.2	798	9.75	4800	1010
H67	12	24.5	38.5	0.8	938	9	10900	1230
H68	5.75	20	22.4	0.4	368	3.25	1810	1680
H75	12	23	32.2	0.6	384	3.5	1380	856
H76	16	19	24.5	1.3	616	3.5	11400	1040
H77	14.8	23	32.3	1	599	7.5	2700	911
H85	17.3	27.5	23.3	1.3	294	3.5	1230	774
H86	13.3	29.5	43.6	1	664	5.25	2530	2220
H87	11.8	27.5	25	0.9	413	4	5050	822
H94	16.3	28.5	27.3	0.6	310	2.5	1050	823
H95	16	23.5	22.3	0.4	315	2.75	1060	767

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.4 Stratified random design, single auger samples.

Sample	Li	Na	K	Be	Mg	Ca	Sr	Ba
S1	58.65	300.5	6410	1.495	3570	12145	50.5	368
S2	65.5	318	7290	1.77	3850	4600	65.9	1120
S3	61.7	313	6430	1.935	4150	5270	75	1830
S4	57.4	283	7140	1.675	4060	5760	84.2	2470
S5	47	230	5740	1.405	2950	9160	68.1	1720
S6	39	233	5250	1.285	2770	31700	183	2010
S7	29.9	223	3370	1.01	1780	2310	73.7	2340
S8	28.7	205.5	3095	0.95	1850	4285	76.9	2175
S9	61	300	6750	1.565	3480	5010	49.8	392
S10	63.2	330	7570	1.65	3770	8020	52.6	364
S11	60.1	290	6380	1.445	3410	4540	99.1	2220
S12	67.5	350	7490	1.895	4820	7590	75.9	1720
S13	39.1	233	4235	1.4975	2355	5065	79.45	1910
S14	23.9	140	2550	0.945	1440	4400	136	2330
S15	30.3	183	3210	0.985	1850	3510	68	1830
S16	31.5	238	3740	1.065	2250	3960	62.1	1900
S18	56.6	283	6510	1.585	3470	4180	51.2	605
S19	78.1	355	8070	1.91	4140	5940	63.4	597
S20	58.3	328	6390	1.64	3700	4430	77	1860
S21	42.7	208	4710	1.545	3100	5850	105	2080
S22	24.6	145	2720	1.01	1600	6600	297	2450
S23	28.1	175	3230	0.92	1680	6080	152	2320
S27	61.3	305	6950	1.995	4860	5680	83.6	1780
S28	34.5	223	4160	1.565	2280	6610	135	2080
S29	37	200	4840	1.15	2540	4090	77.7	1870
S32	43.45	243	4650	1.165	2475	5580	63.85	1540
S33	30	195	2910	1.04	1690	4820	103	2320
S34	33.1	200	3910	1.015	1980	4190	92.5	2290
S37	40	238	4310	1.19	2450	5950	160	2670
S38	22	143	2100	0.925	1280	9030	217	3070
S39	27.2	178	2800	1.21	1490	7910	108	2130
S42	50.8	263	5840	1.61	3050	3040	79.1	2220
S43	28.9	166.5	3295	1.435	1840	5355	138.85	2380
S44	53.5	318	6690	2.04	3520	4810	96.2	2050
S48	64.3	325	8230	1.89	4450	2020	51.6	479
S49	43.5	223	4410	1.39	2290	3940	66.7	1580
S50	51.6	270	5740	1.37	2940	4920	59.7	1260
S52	73.1	359	8160	1.93	3850	6210	67.3	861
S53	60.1	303	6680	1.73	3180	7660	82.2	1480
S56	66.3	325	7440	1.65	3510	2580	60.3	798

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

2.1.4 Stratified random design, single auger samples. (continue)

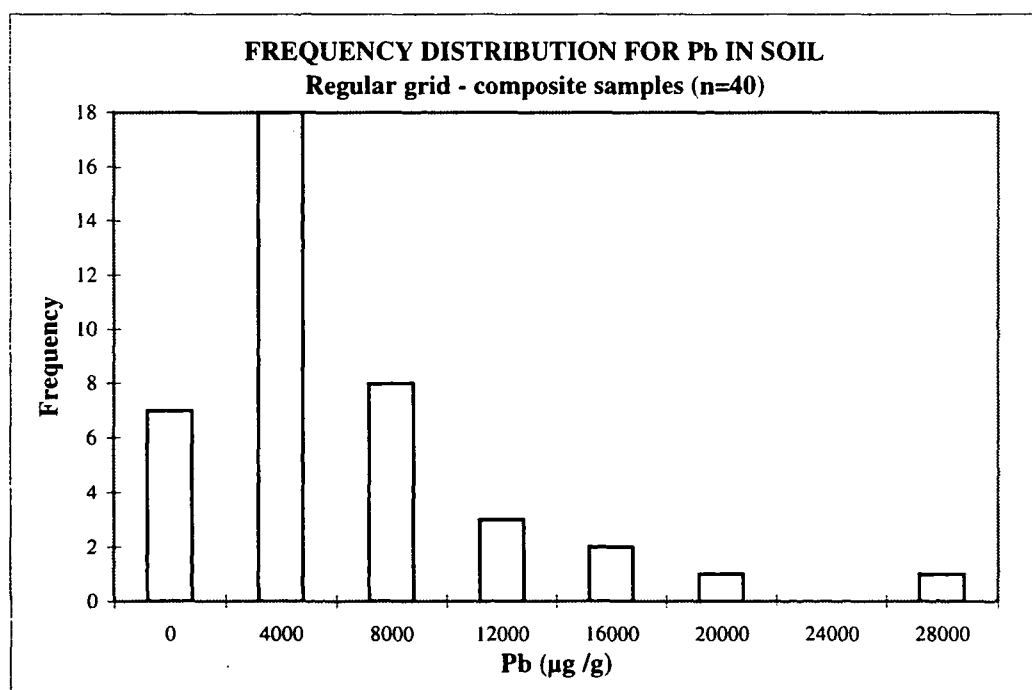
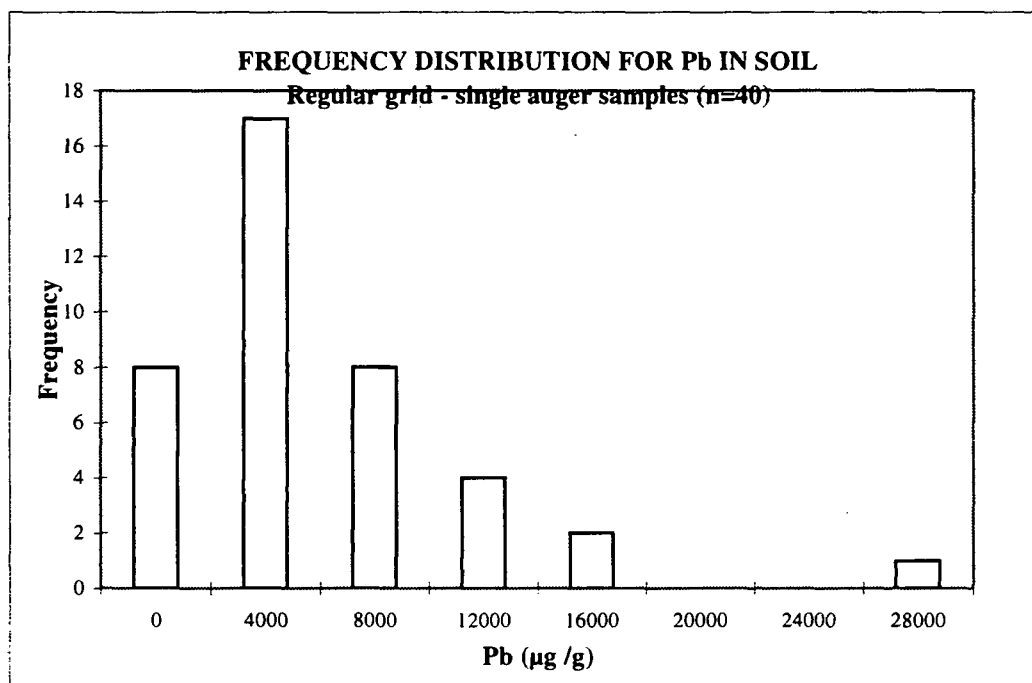
Sample	Al	La	Ti	V	Cr	Mo	Mn	Fe
S1	62450	31.8	1180	58.05	30.75	5.25	1465	31350
S2	69200	30	1110	64.8	32	6	799	32300
S3	68100	34	1260	64.3	34	6	907	33800
S4	63900	30.5	1280	62.5	28.5	5.5	469	23900
S5	53600	30.3	1100	54.3	25.5	5	1230	24900
S6	45300	19.3	875	53.8	26.5	7	667	29600
S7	38500	21.3	867	53	26	5	1140	32400
S8	35850	23.65	1029	51.65	29.75	3.5	443.5	18700
S9	65700	32.5	1110	60	28	5	1220	34000
S10	66700	32	1160	62	31.5	5	1270	33100
S11	63400	28	1030	60	27.5	6	811	28000
S12	70800	34.3	1240	65.5	36.5	5	705	28800
S13	47150	23.15	1080	51.3	28.75	5	861	28950
S14	31100	16.3	796	35	23.5	4.5	627	20900
S15	38000	23	950	52.8	25.5	4	410	21900
S16	36500	24.5	978	54	32.5	3.5	643	21700
S18	60800	31.8	1140	59	27	5	1060	35100
S19	78400	33	1300	69	32	6	1510	35900
S20	64200	30.8	1150	60.3	33	5	690	27100
S21	47500	25.8	994	51.5	25	5.5	355	21000
S22	28700	17.3	587	35.5	25	6	457	17000
S23	34700	14	727	39.3	40	4.5	508	17300
S27	67000	33.5	1240	68.3	31.5	5.5	445	26700
S28	42700	20	906	66.3	29.5	8	714	24700
S29	47500	27.3	1100	52.8	25	4	556	22400
S32	50650	25.8	1085	54.4	29.25	5	573.5	23700
S33	39100	23	1040	45.5	27.5	5	1210	22200
S34	41800	23.5	874	45	23	4.5	426	18600
S37	47000	19.8	932	48.3	28.5	5.5	748	21000
S38	26800	17	581	33.8	21.5	5.5	673	16300
S39	35600	21	913	40.8	21.5	5	616	20200
S42	58300	27.5	1210	59	34	5.5	713	27000
S43	38100	20.9	929.5	42.05	28.25	4.25	927.5	27600
S44	66300	34	1290	68.3	35	6	586	26500
S48	72700	36	1410	69.3	36.5	6.5	596	35700
S49	47900	24.3	981	50.8	26.5	5	1010	22700
S50	61200	28.3	1140	65	23.5	6.5	490	26800
S52	71150	33.8	1315	67	37.25	5.75	1625	31850
S53	61100	31	1140	61.3	32.5	5.5	1020	25800
S56	69000	32.5	1280	65.3	33.5	5.5	1020	29100

Appendix 2-1. Raw ICP-AES analytical data from the initial survey in Derbyshire.
Concentrations in µg /g.

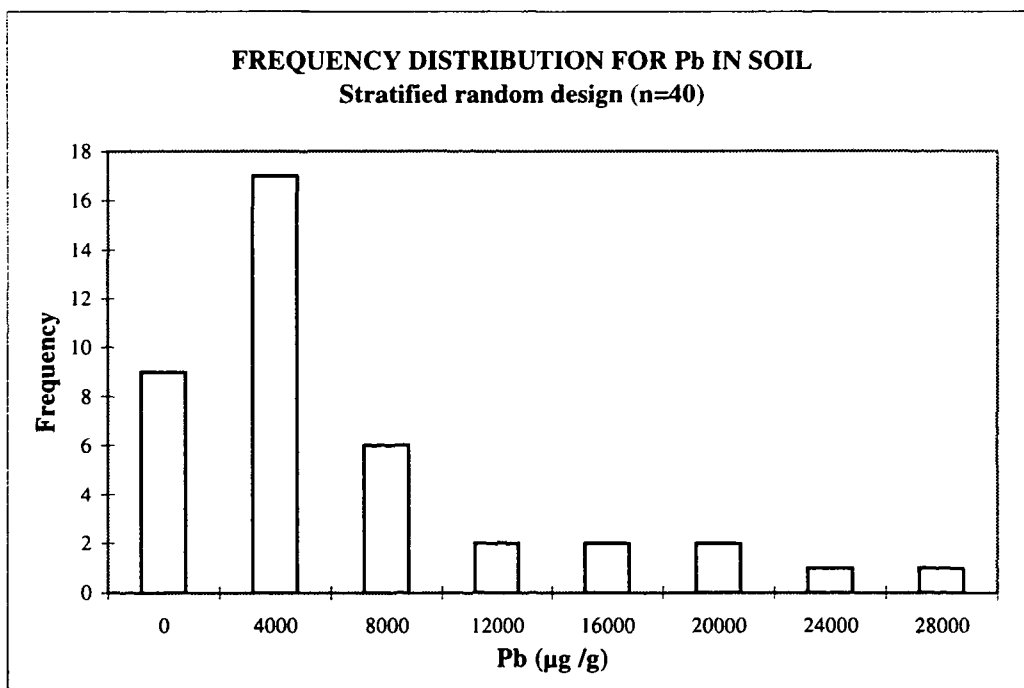
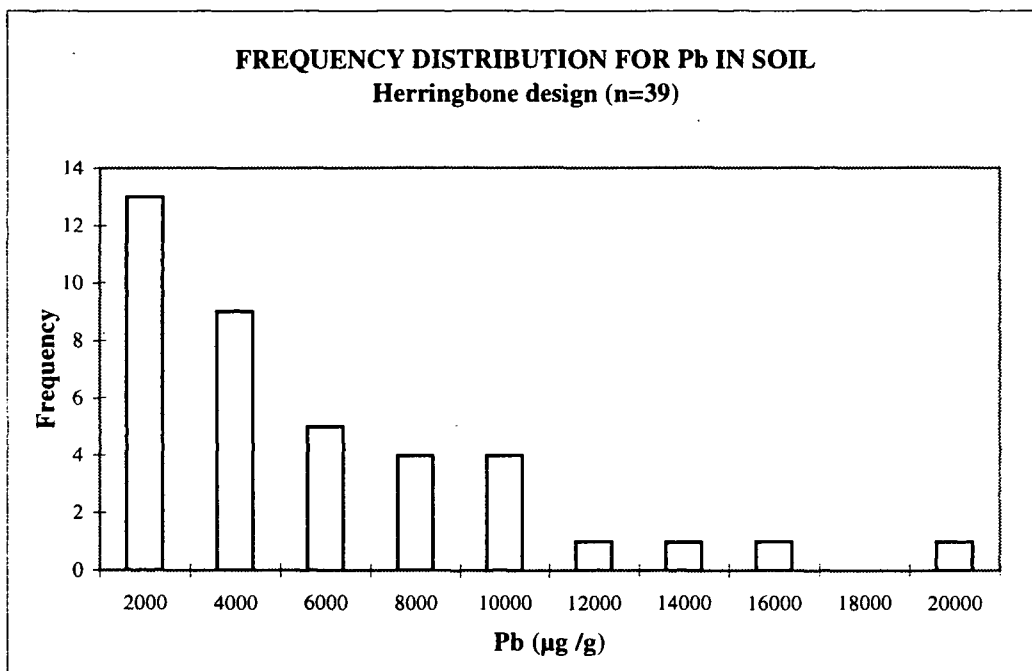
2.1.4 Stratified random design, single auger samples. (continue)

Sample	Co	Ni	Cu	Ag	Zn	Cd	Pb	P
S1	15.65	24	22.05	0.4	194.5	1.25	1410	974
S2	13.8	28	26.2	0.4	370	1.25	6330	1040
S3	15.8	31	34	0.8	545	3.25	5270	827
S4	12.3	26.5	33.3	0.5	617	3.5	5500	837
S5	19.8	23.5	33.2	0.3	565	2.75	3960	957
S6	12.3	27	56.8	0.9	7140	13.5	20200	2450
S7	12.8	18	31	0.3	388	2.25	4830	1880
S8	6.625	14.75	18.6	0.55	433.5	4.125	3335	783.5
S9	16.3	23.5	21.6		176		1400	1030
S10	17	26.5	23.1	0.5	183	1	1230	958
S11	11.8	22.5	23.9	0.6	434	2	8270	1050
S12	14	32	30.1	0.8	541	3	4600	802
S13	14.3	25.5	41.6	0.7	816.5	6	6625	1120
S14	8.75	17	30.5	0.9	954	7	14600	1460
S15	7	14.5	21.8	0.3	304	2.5	3640	1150
S16	10.8	19	22.5	0.8	386	3.75	3130	900
S18	17.3	37.5	21.8		218	1.5	1400	807
S19	18.8	30	23.7	0.2	247	1	1660	823
S20	12.8	25.5	28.5	0.9	427	2.25	6820	954
S21	8.5	22	30.8	1	685	7.5	12300	1270
S22	6.25	15	29.1	1.1	1350	9	29500	2000
S23	6.75	14.5	25.3	1	1000	5	20100	1740
S27	14	33.5	35	0.3	636	6.25	3050	749
S28	12.8	26	39	1.4	960	7.5	16400	1610
S29	10.8	20.5	28	0.4	426	3.75	2400	941
S32	8.75	21.25	26.15	0.65	351	2.5	4970	873
S33	9.75	19	27.5	0.8	687	6.25	5950	1040
S34	7.5	15	23.3	0.4	591	5	7350	1150
S37	9.75	20.5	25.6	0.8	800	4.5	13900	1060
S38	6.75	14	27.9	0.8	1040	5	24000	1470
S39	9	18.5	32.1	0.5	757	10.8	7340	1270
S42	11.5	26.5	29.7	0.7	410	3.5	4890	904
S43	19.9	31	43.8	0.7	948	8.65	5380	1006
S44	16.3	25.5	38.5	0.6	505	5.75	1840	760
S48	15.3	33.5	34.5	0.7	244	2	864	775
S49	16.3	23	30.1	0.8	450	3	2460	1600
S50	8.75	22.5	22.1	0.2	341	2.75	2710	887
S52	23.3	30.25	26.6	0.5	356.5	3	1365	1045
S53	15.3	29	33	0.6	626	3.75	2610	1900
S56	20.5	26	25.1		272	1.5	1140	708

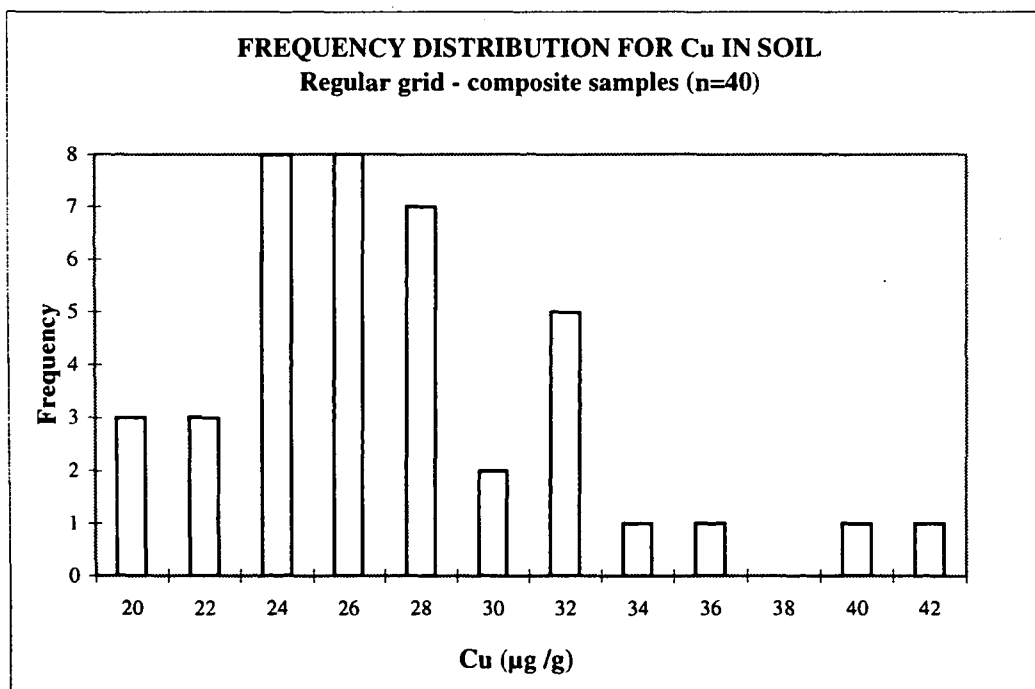
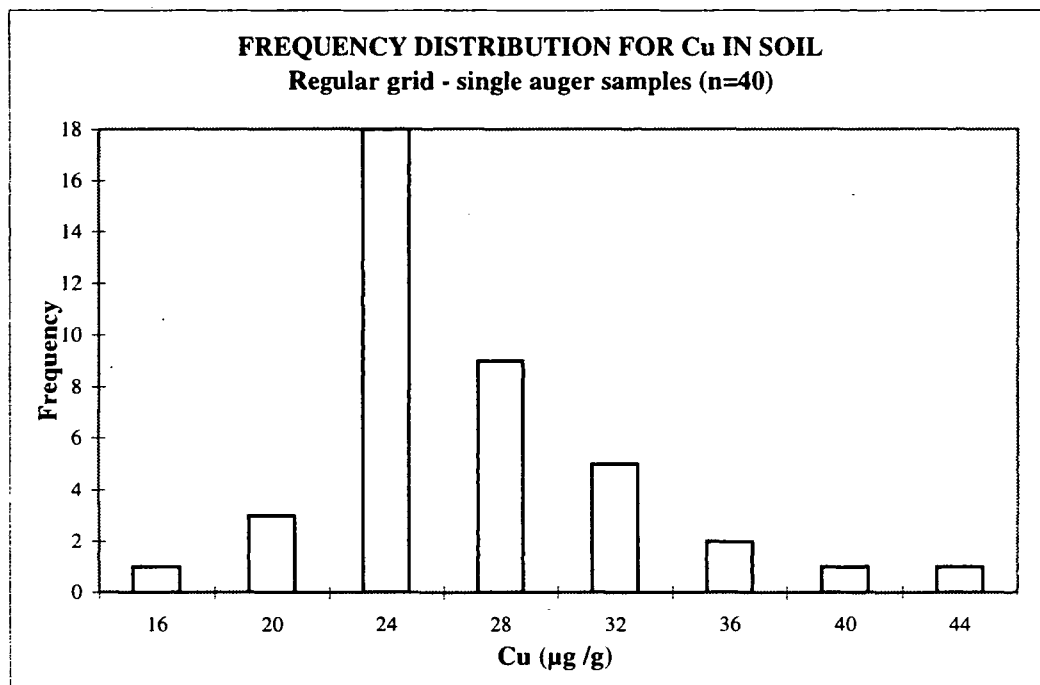
Appendix 2-2. Frequency distributions of Pb for the four sampling designs from the initial survey in Derbyshire.



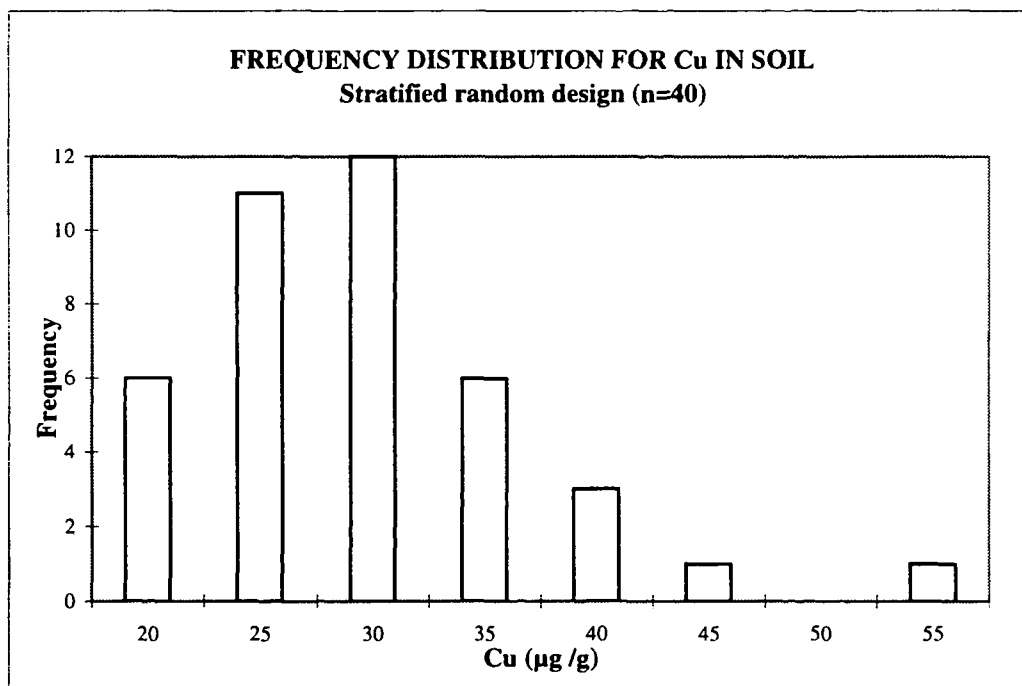
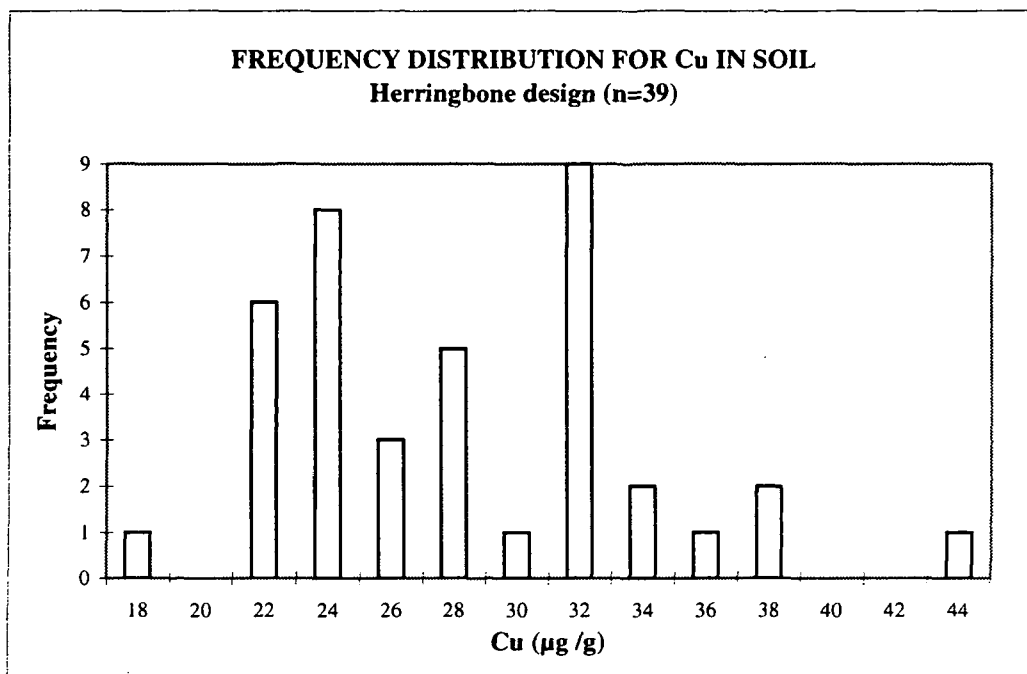
Appendix 2-2. Frequency distributions of Pb for the four sampling designs from the initial survey in Derbyshire. (continue)



Appendix 2-3. Frequency distributions of Cu for the four sampling designs from the initial survey in Derbyshire.



Appendix 2-3. Frequency distributions of Cu for the four sampling designs from the initial survey in Derbyshire. (continue)



Appendix 2-4. Raw ICP-AES analytical data from the initial survey in Hounslow Heath, London. Concentrations in µg /g.

Sample	x (m)	y (m)	Li	Na	K	Be	Mg	Ca	Sr
HA1	0	270	29.3	635	6230	1.705	5160	20600	97.6
HA2	0	240	43.4	858	8490	2.935	7480	19300	130
HA3	0	210	50.4	873	10300	2.445	9330	22500	121
HA4	0	180	33.2	811.5	5780	2.71	6370	40900	218.5
HA5	0	150	26.2	543	5760	1.515	5180	40100	123
HA6	0	120	27.6	380	6350	1.47	4790	19600	69.2
HA7	0	90	20.8	230	4790	1.085	2960	2170	25.5
HA8	0	60	20.5	233	4480	1.18	3060	1850	26.7
HA9	0	30	17	198	3580	0.69	2140	3750	32.3
HA10	0	0	16.1	205	3080	0.94	1930	2830	26.9
HB1	30	270	28.6	413	6670	1.795	5240	13600	66.1
HB2	30	240	19.4	1140	4950	1.51	3880	17900	93.5
HB3	30	210	30.6	363	6530	1.58	4620	5170	45.9
HB4	30	180	17.8	310	3880	1.86	2410	4710	43
HB5	30	150	38.4	413	8220	1.88	7170	16000	68.5
HB6	30	120	35.1	433	7420	1.755	6500	16500	74.3
HB7	30	90	21.75	389	4600	1.215	3595	30350	95.1
HB8	30	60	19.9	203	4190	1.07	2610	3030	29.1
HB9	30	30	30.9	295	6300	1.365	4550	6560	41.4
HB10	30	0	24.6	378	5440	1.19	4070	97200	158
HC1	60	270	37.05	551.5	8245	1.8125	6575	15000	86.25
HC2	60	240	32.8	573	7060	1.865	5410	10700	63.3
HC3	60	210	47	538	10300	2.12	7790	9560	67.1
HC4	60	180	24.1	363	4740	1.465	3280	6470	49.6
HC5	60	150	24.4	313	5560	1.47	3870	5460	39.8
HC6	60	120	22.6	300	5590	1.205	3590	7490	40.3
HC7	60	90	23.8	325	5030	1.435	3520	4460	36.8
HC8	60	60	21.7	238	4550	1.145	2930	3000	29.4
HC9	60	30	20.2	260	4600	0.985	2790	3100	28.4
HC10	60	0	32.1	463	6530	1.715	5340	19500	88.4
HD1	90	270	27.3	633	6150	1.965	4260	15300	93.4
HD2	90	240	44.4	515	9480	1.84	8160	18500	105
HD3	90	210	34.5	458	7440	1.59	5420	8790	57.1
HD4	90	180	65.3	558	12900	2.165	12000	21500	146
HD5	90	150	33.8	525	7120	1.705	5710	15600	74.3
HD6	90	120	23.8	333	5490	1.265	3920	10800	50.1
HD7	90	90	23.9	213	5520	1.125	3540	2600	28.2
HD8	90	60	22.8	288	4720	1	2990	5220	40.2
HD9	90	30	33.3	362.5	7360	1.5475	5555	13300	60
HD10	90	0	25.8	340	5480	1.335	4420	20600	70.6
HE1	120	270	51.4	715	9930	2.105	8780	18600	99.8
HE2	120	240	32	583	6520	1.67	5350	19600	91.3
HE3	120	210	31	540	6340	1.555	5260	21400	116
HE4	120	180	43.9	648	9670	2.04	7850	15500	102
HE5	120	150	42.3	700	9030	1.785	6840	14400	83.2
HE6	120	120	29.9	435	6500	1.425	5020	22200	72.6
HE7	120	90	28.1	398	5340	1.28	3810	7310	49
HE8	120	60	37.95	571.5	7700	2.205	6340	15550	86.3
HE9	120	30	18.4	293	3820	1.19	3010	12900	60.5
HE10	120	0	17.8	165	3680	0.775	2290	1960	23.4

Appendix 2-4. Raw ICP-AES analytical data from the initial survey in Hounslow Heath, London. Concentrations in µg /g. (continue)

Sample	x (m)	y (m)	Li	Na	K	Be	Mg	Ca	Sr
HF1	150	270	24.3	378	4900	1.345	3640	8150	50.4
HF2	150	240	46.8	610	9770	1.96	8390	18100	86.9
HF3	150	210	46.7	568	9090	1.98	7930	20300	107
HF4	150	180	21.2	400	4210	1.53	2880	13300	65.3
HF5	150	150	40.7	448	8680	1.77	6940	10800	67
HF6	150	120	24.8	400	5990	1.53	4360	11600	59.7
HF7	150	90	32.85	458	7335	1.58	6035	16200	80.95
HF8	150	60	30.1	415	5740	1.875	5210	14900	76.3
HF9	150	30	36.4	423	7790	1.61	7160	31900	74
HF10	150	0	28.7	440	5110	1.49	3930	7810	56
HG1	180	270	24.1	430	5340	1.435	3560	9540	57.5
HG2	180	240	41.7	500	8930	1.965	7550	10900	70.7
HG3	180	210	56.8	528	13000	2.295	11100	9810	73.2
HG4	180	180	57	505	12300	2.13	10800	6560	56.3
HG5	180	150	39.7	715	8370	1.875	6630	23300	97.7
HG6	180	120	42.8	773	9010	2.125	6560	26700	87.8
HG7	180	90	57.1	582.5	11900	2.47	10780	13200	81.7
HG8	180	60	42.5	460	8250	1.885	7270	13800	83.4
HG9	180	30	28.7	300	5680	1.52	4170	5690	47
HG10	180	0	25.9	260	5260	1.185	4020	3000	30.6
HH1	210	270	55.2	578	10800	1.965	10800	31300	131
HH2	210	240	53.7	463	11700	2.09	11000	6830	60.7
HH3	210	210	47.2	470	9400	1.775	8260	15600	81.9
HH4	210	180	23.6	275	5260	1.265	3470	6900	41.4
HH5	210	150	26.1	306.5	4525	1.2675	3540	65500	75.45
HH6	210	120	32.3	463	6620	1.555	4840	9100	52.9
HH7	210	90	41.4	453	8100	1.81	7030	12000	72.6
HH8	210	60	35.1	343	7470	1.73	6420	12200	57.3
HH9	210	30	25.1	345	5940	1.295	4670	9530	53.8
HH10	210	0	25.3	270	5290	1.42	3640	3250	34.1
HI1	240	270	27.4	225	4110	1.42	3600	71200	83.5
HI2	240	240	30.5	438	6550	1.73	5050	27000	89.5
HI3	240	210	32.1	395	6270	1.445	4850	14100	62.2
HI4	240	180	53.7	450	11900	2.4	10000	31300	84.7
HI5	240	150	22.9	200	3910	1.12	3060	46300	61.9
HI6	240	120	24.3	268	5140	1.375	3790	11600	46.6
HI7	240	90	40.7	480	7850	2.36	6480	19000	99.3
HI8	240	60	38.5	470	7980	1.965	6650	10900	74.6
HI9	240	30	28.65	321.5	4970	1.5125	3610	10515	66.6
HI10	240	0	28.6	320	6230	1.365	4130	5290	41.9
HJ1	270	270	21.6	295	4320	1.085	2740	8050	39.7
HJ2	270	240	23.9	313	4350	1.16	2990	7250	38.8
HJ3	270	210	28.7	270	4480	1.39	3560	9200	48.4
HJ4	270	180	25	288	3880	1.42	2810	8930	37.5
HJ5	270	150	46	414	9560	1.8	7920	6465	50.65
HJ6	270	120	32.7	300	7200	1.415	4980	3010	33.7
HJ7	270	90	27.2	285	5020	1.395	3610	17500	59.4
HJ8	270	60	35.2	405	6440	1.67	5360	41600	97.3
HJ9	270	30	32.3	300	5980	1.6	4570	6800	47
HJ10	270	0	28.6	268	6120	1.445	4160	5170	41.5

Appendix 2-4. Raw ICP-AES analytical data from the initial survey in Hounslow Heath, London. Concentrations in µg /g. (continue)

Sample	x (m)	y (m)	Ba	Al	La	Ti	V	Cr	Mo
HA1	0	270	349	30800	22.5	1110	75.3	40.5	2.5
HA2	0	240	474	41800	27.5	1390	101	53.5	4.5
HA3	0	210	399	48200	29.5	1570	106	54	3
HA4	0	180	509.5	31300	27.75	1118	84.15	39.75	3
HA5	0	150	293	29200	21.3	1180	71.8	35.5	2
HA6	0	120	235	30800	22.3	1300	74	42	1.5
HA7	0	90	122	25500	22.5	1240	55.5	43.5	1.5
HA8	0	60	103	24600	20.5	1050	59	31.5	2.5
HA9	0	30	100	19400	18.5	1210	42.8	38.5	1
HA10	0	0	98.2	18400	18.8	1170	48	39.5	2.5
HB1	30	270	246	32400	25	1020	77.5	32.5	3
HB2	30	240	1100	21400	17.3	582	59.3	94	10
HB3	30	210	201	35100	25.8	1180	75.5	42	2.5
HB4	30	180	235	19200	19.8	855	61	34	3.5
HB5	30	150	201	40800	27.3	1460	91.5	39.5	2.5
HB6	30	120	250	34700	24	1120	84	35	3
HB7	30	90	277.5	23700	19.4	1075	58.8	32.25	1.5
HB8	30	60	111	22500	20.3	993	54.5	25.5	2.5
HB9	30	30	210	32000	22	1400	72.3	48	1.5
HB10	30	0	251	24400	18	990	59.5	30.5	-32768
HC1	60	270	349.5	38200	25.05	1215	86.75	63	3.5
HC2	60	240	225	33500	26.8	1360	77.8	42.5	3
HC3	60	210	264	47000	28.5	1410	103	44	3.5
HC4	60	180	200	26600	23.3	1140	66	35	3.5
HC5	60	150	179	29100	24.3	1140	68.3	39	3
HC6	60	120	166	26200	19.3	1140	59	36	1.5
HC7	60	90	178	27700	23	1050	65.8	34	3
HC8	60	60	133	24300	22.5	1080	56.3	29	2.5
HC9	60	30	127	23700	21	1230	53.3	41	1.5
HC10	60	0	638	32200	21.5	1270	79.5	43.5	2
HD1	90	270	274	28900	20	936	76.5	41	2.5
HD2	90	240	292	44500	29.5	1520	97	66	2
HD3	90	210	199	35000	21.8	1200	85	40.5	1.5
HD4	90	180	1160	59200	31.3	1800	113	44.5	7.5
HD5	90	150	285	34400	23.3	1310	81.8	42.5	2
HD6	90	120	241	27200	19	1180	66	41	1
HD7	90	90	151	30100	19	1270	60.3	44	-32768
HD8	90	60	325	22800	17.5	1150	52.8	50.5	9.5
HD9	90	30	261	34100	22.9	1365	80.25	40	1.5
HD10	90	0	271	26400	22	1090	68.3	36	1.5
HE1	120	270	274	48100	27	1320	103	37.5	2.5
HE2	120	240	256	33700	23.8	1150	77.8	34.5	2
HE3	120	210	477	30900	22.5	1050	73.8	122	2.5
HE4	120	180	310	45000	25.8	1400	96.3	46.5	2.5
HE5	120	150	331	40500	23.8	1360	88	44	6.5
HE6	120	120	230	30000	22	1260	72.8	39	1
HE7	120	90	275	26300	18.5	1220	61.5	38.5	1
HE8	120	60	235.5	37950	25.75	1340	85.3	39.5	2
HE9	120	30	383	20100	18.3	956	63	49.5	2.5
HE10	120	0	113	21500	17.5	1200	45	48	6

Appendix 2-4. Raw ICP-AES analytical data from the initial survey in Hounslow Heath, London. Concentrations in µg /g. (continue)

Sample	x (m)	y (m)	Ba	Al	La	Ti	V	Cr	Mo
HF1	150	270	220	25300	18.5	1040	60.5	30.5	2.5
HF2	150	240	504	46400	26.3	1660	101	41.5	1.5
HF3	150	210	500	44500	25.3	1390	95	95.5	2
HF4	150	180	355	21900	18	806	56.5	42.5	2.5
HF5	150	150	256	40300	22.5	1450	89.3	50	1
HF6	150	120	227	27000	20.3	1030	70	36.5	3
HF7	150	90	340.5	33800	21.15	1220	80.5	41.75	2
HF8	150	60	300	29800	21.3	1080	76.8	28	2
HF9	150	30	183	36700	23.5	1410	87.5	41.5	2.5
HF10	150	0	222	28400	20.5	1270	84.5	44.5	2.5
HG1	180	270	270	25700	20.3	687	59.8	37.5	2
HG2	180	240	232	41900	28	1240	96.5	37.5	2.5
HG3	180	210	238	59500	30.5	1710	116	45.5	1.5
HG4	180	180	210	55900	29.3	1730	110	49	2
HG5	180	150	366	40100	25.5	1300	87.8	52	9
HG6	180	120	295	43300	27.8	1380	92	50.5	2.5
HG7	180	90	245	57150	31.05	1630	117	39.25	2.5
HG8	180	60	356	39600	23.5	1430	95.5	53	3
HG9	180	30	180	30100	21	1180	75.3	36.5	3
HG10	180	0	158	29900	23.5	1380	63	49.5	1.5
HH1	210	270	199	51000	27	1700	104	41	-32768
HH2	210	240	195	54600	29.3	1600	105	36	2
HH3	210	210	198	45000	25.3	1520	89.8	41.5	1.5
HH4	210	180	233	26000	22.8	1020	58	52	1.5
HH5	210	150	186	27900	25.05	1025	57.4	37	-16383.3
HH6	210	120	220	35100	22	1270	76.5	42.5	2.5
HH7	210	90	506	38900	23.3	1290	86.8	44	2
HH8	210	60	196	37700	25.8	1310	80.8	40.5	1.5
HH9	210	30	245	26200	18	591	63.8	29	2
HH10	210	0	189	29800	23	1310	68.5	62.5	3
HI1	240	270	274	31400	28.5	961	59.3	45	-32768
HI2	240	240	298	33000	27.5	1060	73.8	51.5	3
HI3	240	210	206	34000	23.5	1090	70	34	1.5
HI4	240	180	242	57800	36.3	1580	123	40.5	3.5
HI5	240	150	217	26200	23.3	1050	51.8	48	-32768
HI6	240	120	170	26200	20.8	939	65.5	39	1.5
HI7	240	90	327	40100	28.5	1680	101	62.5	4
HI8	240	60	259	38200	24.3	1200	86.8	37	3
HI9	240	30	418.5	29550	27.5	1075	59.05	76.5	2
HI10	240	0	193	31000	22.3	1190	68.8	44	2.5
HJ1	270	270	190	25000	20	1150	55	48.5	-32768
HJ2	270	240	216	26000	23.8	1150	53	51.5	-32768
HJ3	270	210	300	29500	28.8	1100	57.8	58.5	1
HJ4	270	180	181	26700	25	1220	62.3	44	2.5
HJ5	270	150	212	47750	27.15	1550	94.4	38.25	1.25
HJ6	270	120	136	34000	22	1240	75.8	37.5	1
HJ7	270	90	289	28100	25.3	1180	61.3	66	1.5
HJ8	270	60	241	34800	24.3	1370	89.8	53.5	1.5
HJ9	270	30	260	34000	27.8	1230	70	52	1.5
HJ10	270	0	240	31700	23.8	1150	66.5	38	1.5

Appendix 2-4. Raw ICP-AES analytical data from the initial survey in Hounslow Heath, London. Concentrations in µg /g. (continue)

Sample	x (m)	y (m)	Mn	Fe	Co	Ni	Cu	Ag	Zn
HA1	0	270	407	31100	12.5	39.5	186	3.4	460
HA2	0	240	582	41900	18.3	64	284	1.7	536
HA3	0	210	439	42800	16.5	54	140	1.7	500
HA4	0	180	855.5	34000	13.15	43.5	175.5	1.7	613
HA5	0	150	368	31400	10.5	34	67.8	-32768	269
HA6	0	120	357	32300	11	30	56.2	-32768	198
HA7	0	90	321	26100	8.25	18.5	17.6	0.4	67.6
HA8	0	60	191	25100	9	20	18.8	0.5	76.2
HA9	0	30	286	18100	6	14.5	18.6	-32768	63.5
HA10	0	0	216	18700	6.75	15.5	17.2	0.8	58.2
HB1	30	270	482	32800	14.5	43	46.7	0.6	193
HB2	30	240	618	76200	24.5	130	660	2.5	3930
HB3	30	210	390	34200	13.3	33.5	88.3	1.2	249
HB4	30	180	323	31700	13	31.5	52.9	1.3	164
HB5	30	150	302	33800	14	34.5	48.7	1	150
HB6	30	120	359	35100	13.3	38.5	73.1	0.6	251
HB7	30	90	370.5	26650	10.25	25.25	61.7	-32768	214.5
HB8	30	60	185	21000	8.25	19	32	0.5	90.6
HB9	30	30	359	31300	11	28	47.5	-32768	183
HB10	30	0	404	24500	9.25	29.5	71.3	-32768	238
HC1	60	270	377	36100	13.9	51.75	153	10.75	379
HC2	60	240	411	33500	14	35.5	97.3	2.4	218
HC3	60	210	583	40400	19	41	74.7	2.4	211
HC4	60	180	260	24700	9.5	26	56.4	1	234
HC5	60	150	270	28100	11	26.5	54.3	1.2	171
HC6	60	120	346	26700	9.5	24	52.3	-32768	132
HC7	60	90	296	26700	11.5	27.5	43.5	0.6	193
HC8	60	60	232	22300	9.75	21	25.4	0.7	116
HC9	60	30	262	23200	8.25	19.5	24.8	0.3	103
HC10	60	0	408	33000	12.5	36	98	-32768	511
HD1	90	270	506	39900	14.5	45	145	2.4	351
HD2	90	240	351	34300	13.3	51	114	11.6	286
HD3	90	210	417	34000	13.3	66.5	85.1	1.5	186
HD4	90	180	361	42500	15.5	46.5	41	-32768	118
HD5	90	150	364	31500	12.3	52	118	1.9	266
HD6	90	120	327	28400	9.25	30	107	19	240
HD7	90	90	224	30300	7.5	20.5	22.8	-32768	80
HD8	90	60	357	25000	10.8	35	114	-32768	996
HD9	90	30	441	34000	12.55	33	52.85	-16383.6	207.5
HD10	90	0	334	29100	10.5	30	68.8	1.2	273
HE1	120	270	402	37300	15.8	48	130	3	295
HE2	120	240	453	37100	13.5	40.5	121	0.9	304
HE3	120	210	392	30300	12	56	229	33.2	505
HE4	120	180	342	38800	15.3	44.5	138	2.8	279
HE5	120	150	352	32500	13.3	41	168	5.3	350
HE6	120	120	486	33100	12.5	32.5	65.2	1.3	188
HE7	120	90	413	28200	9.75	27	47.9	0.2	166
HE8	120	60	448	36400	14.65	40.75	67.8	-16383.8	155.5
HE9	120	30	440	66400	13.5	41	94.6	0.8	391
HE10	120	0	259	19500	6.75	14	19.5	-32768	90.7

Appendix 2-4. Raw ICP-AES analytical data from the initial survey in Hounslow Heath, London. Concentrations in µg /g. (continue)

Sample	x (m)	y (m)	Mn	Fe	Co	Ni	Cu	Ag	Zn
HF1	150	270	517	29700	11.5	31.5	61.7	0.5	252
HF2	150	240	372	42000	14.3	40.5	60.2	0.5	220
HF3	150	210	500	38800	14.5	96.5	200	19.9	461
HF4	150	180	366	30400	10.3	33.5	120	0.7	922
HF5	150	150	417	37000	13.3	38.5	72.6	0.7	242
HF6	150	120	348	28800	12.8	33	108	2.5	257
HF7	150	90	404	33300	12.55	41.25	148	1.5	363.5
HF8	150	60	427	30900	13	37	70.1	-32768	254
HF9	150	30	282	31600	11.5	36.5	54	0.6	566
HF10	150	0	370	28900	10.8	30.5	55	0.6	232
HG1	180	270	292	28000	10.5	37.5	168	1.5	380
HG2	180	240	537	36600	17	44.5	73.1	2.2	200
HG3	180	210	447	43200	17.5	48	43.4	0.9	133
HG4	180	180	334	40800	16.3	42.5	35	1.3	113
HG5	180	150	395	34600	14	44.5	181	6.1	377
HG6	180	120	445	37300	14.3	43	136	3.4	329
HG7	180	90	332	44650	17.25	55.5	85.35	-16383.4	165.5
HG8	180	60	374	36200	14.3	50	195	0.9	347
HG9	180	30	240	32500	12	42.5	30.8	0.6	139
HG10	180	0	467	26300	9.5	22	19	-32768	75
HH1	210	270	350	38200	13.8	40.5	42.5	0.3	133
HH2	210	240	270	40600	15	40	30.3	-32768	117
HH3	210	210	310	35400	12.3	34.5	41.2	0.4	134
HH4	210	180	413	28200	10.3	34	69.1	7.4	185
HH5	210	150	383	26250	8.775	28.5	50.4	2.4	138.6
HH6	210	120	327	32000	12	34	86.8	1.2	205
HH7	210	90	280	35300	13.3	41	97.2	0.8	332
HH8	210	60	393	35300	12.5	38	50	0.8	216
HH9	210	30	240	27600	10.8	32.5	155	0.6	224
HH10	210	0	272	27800	11.8	27.5	41.9	1.9	136
HI1	240	270	610	34500	10.8	36.5	63.3	3.8	174
HI2	240	240	366	32200	13.3	41.5	116	7	277
HI3	240	210	424	30800	12.5	32	71.5	1.5	191
HI4	240	180	372	43600	17.3	50.5	33.3	1.1	102
HI5	240	150	384	26100	7.75	27	57.1	5.1	155
HI6	240	120	362	30900	9.75	29	59.9	0.8	163
HI7	240	90	458	36900	15.3	44	215	1.2	357
HI8	240	60	354	35900	14	42.5	109	1.4	238
HI9	240	30	494	33700	11.3	49	153	21.35	343
HI10	240	0	412	30300	10.5	31	44	2.2	152
HJ1	270	270	348	26800	8.75	23	66.2	1.5	182
HJ2	270	240	323	25900	10	27.5	74.6	5.2	161
HJ3	270	210	400	30200	10	40	78.3	14.1	211
HJ4	270	180	439	28500	13.3	28	40.7	1.6	114
HJ5	270	150	299.5	37200	13.375	36.75	28.4	-16383.9	92.8
HJ6	270	120	316	36300	13.8	35	40.2	-32768	107
HJ7	270	90	445	31900	14.3	41	93.1	11.5	249
HJ8	270	60	545	32300	11.5	41.5	114	0.8	282
HJ9	270	30	451	35100	12.5	39.5	65.1	5.6	196
HJ10	270	0	452	38700	12	36.5	80.3	1.3	213

Appendix 2-4. Raw ICP-AES analytical data from the initial survey in Hounslow Heath, London. Concentrations in µg /g. (continue)

Sample	x (m)	y (m)	Cd	Pb	P
HA1	0	270	1.75	474	3000
HA2	0	240	2	378	1780
HA3	0	210	1.5	327	1670
HA4	0	180	1.75	787	1985
HA5	0	150	1.25	395	988
HA6	0	120	-32768	453	822
HA7	0	90	-32768	72	421
HA8	0	60	1.25	70.5	467
HA9	0	30	0.5	72	589
HA10	0	0	1	88.5	560
HB1	30	270	1	287	1010
HB2	30	240	32.3	3590	1970
HB3	30	210	1.75	197	1220
HB4	30	180	1.25	207	1060
HB5	30	150	2	165	769
HB6	30	120	2.75	371	889
HB7	30	90	-32768	470	723.5
HB8	30	60	1.75	101	738
HB9	30	30	-32768	188	750
HB10	30	0	0.75	366	970
HC1	60	270	6.125	250	2765
HC2	60	240	2	260	1570
HC3	60	210	2	240	1700
HC4	60	180	2	197	1370
HC5	60	150	1.75	188	938
HC6	60	120	-32768	155	823
HC7	60	90	4.25	194	517
HC8	60	60	3	108	615
HC9	60	30	0.75	104	635
HC10	60	0	-32768	495	1060
HD1	90	270	0.75	338	2890
HD2	90	240	6	152	2530
HD3	90	210	0.75	159	1780
HD4	90	180	-32768	87	874
HD5	90	150	0.75	344	2030
HD6	90	120	1	462	1000
HD7	90	90	-32768	82.5	322
HD8	90	60	3.75	521	675
HD9	90	30	-32768	462.5	1100
HD10	90	0	0.5	779	1110
HE1	120	270	1	212	2240
HE2	120	240	0.5	197	2030
HE3	120	210	12.5	327	4340
HE4	120	180	0.75	254	2520
HE5	120	150	1	314	3570
HE6	120	120	-32768	258	1320
HE7	120	90	-32768	162	828
HE8	120	60	-32768	217.5	801.5
HE9	120	30	-32768	482	1320
HE10	120	0	-32768	60	592

Appendix 2-4. Raw ICP-AES analytical data from the initial survey in Hounslow Heath, London. Concentrations in µg /g. (continue)

Sample	x (m)	y (m)	Cd	Pb	P
HF1	150	270	-32768	458	1290
HF2	150	240	-32768	711	1080
HF3	150	210	8.75	264	3660
HF4	150	180	-32768	1840	1690
HF5	150	150	-32768	302	1310
HF6	150	120	2.5	245	1920
HF7	150	90	2.25	441	1790
HF8	150	60	-32768	327	864
HF9	150	30	1	228	1010
HF10	150	0	-32768	206	824
HG1	180	270	2.25	713	2210
HG2	180	240	2.75	165	1630
HG3	180	210	-32768	105	1020
HG4	180	180	1	78	1000
HG5	180	150	2.75	284	3590
HG6	180	120	2.5	237	1970
HG7	180	90	-16383.4	199	1017
HG8	180	60	1.25	540	1950
HG9	180	30	1.25	135	608
HG10	180	0	-32768	55.5	468
HH1	210	270	-32768	125	943
HH2	210	240	-32768	69	948
HH3	210	210	-32768	137	1050
HH4	210	180	3.5	102	1490
HH5	210	150	1.625	88.75	1280
HH6	210	120	1.75	173	1560
HH7	210	90	1.25	326	1010
HH8	210	60	1	132	831
HH9	210	30	1.5	285	1140
HH10	210	0	1.75	135	630
HI1	240	270	4.5	76.5	1290
HI2	240	240	4.5	206	2550
HI3	240	210	1	131	1670
HI4	240	180	2	70.5	929
HI5	240	150	2.5	87	1420
HI6	240	120	1	152	1430
HI7	240	90	2.75	290	2070
HI8	240	60	2.25	258	1230
HI9	240	30	9.625	181	2680
HI10	240	0	2	137	938
HJ1	270	270	-32768	168	1530
HJ2	270	240	1.75	126	1370
HJ3	270	210	6.5	102	1950
HJ4	270	180	1	107	953
HJ5	270	150	-32768	83.25	739.5
HJ6	270	120	-32768	82.5	837
HJ7	270	90	5	164	1660
HJ8	270	60	1.75	246	1020
HJ9	270	30	4.5	146	1440
HJ10	270	0	1	149	1300

Appendix 2-5. Raw analytical data from the three analytical techniques for Pb and Zn from the portable XRF study in Derbyshire.

Element	Pb (µg/g)			Zn (µg/g)		
Sample	P-XRF field	P-XRF lab	ICP-AES	P-XRF field	P-XRF lab	ICP-AES
B11	1319	6300	7340	192	311	365
B12	4177	8110	8815	266	402	409
B13	1840	4893	1522	247	479	467
B14	1108	1348	1290	296	479	685
B15	2547	6951	9340	390	640	829
B16	1240	3017	3080	403	950	1240
B21	1568	2705	4180	213	309	271
B22	1183	1816	1926	170	254	251
B23	3626	3812	3670	439	520	723
B24	8555	6536	6718	628	644	665
B25	2869	4898	5630	311	504	485
B26	2690	3576	3630	323	405	283
B32	881	5694	6880	192	620	739
B33	5002	8455	9370	333	415	364
B34	454	1428	1522	144	281	289
B35	10919	20328	21877	669	1065	1144
B44	3788	6525	5230	412	700	683
B45	7556	17363	18784	788	1631	1728
B46	1963	9465	2800	445	833	660
B54	6050	9830	10584	356	590	599
B55	3745	7074	7316	522	691	733
B56	1384	2187	2235	323	559	532
B64	1964	4279	3860	362	516	497
B65	2188	4491	5210	259	414	495

APPENDIX 3.
Duplicate Data and ANOVA Results

Appendix 3-1. Duplicate data and ANOVA results from the initial survey in the Derbyshire site.

SAMPLE I.D.	S1A1Pb ($\mu\text{g g}^{-1}$)	S1A2Pb ($\mu\text{g g}^{-1}$)	S2A1Pb ($\mu\text{g g}^{-1}$)	S2A2Pb ($\mu\text{g g}^{-1}$)
R11C	1555.8	1565.3	1746.1	1810.5
R16C	8128.0	8092.9	7413.6	6823.0
R22C	1817.6	1822.7	1894.6	1855.6
R27C	2448.2	2335.1	2222.3	2236.4
R36C	18327.7	18273.2	36695.0	36601.3
R44C	6403.1	5905.2	7354.0	6952.5
R56C	17130.2	16738.0	21268.2	20040.2
R67C	3418.0	3628.3	3374.2	3409.0
R75C	2559.5	2665.9	4079.5	4187.2
R11S	1374.5	1406.7	1739.4	1587.9
R27S	2534.0	3964.8	2444.0	2459.6
R36S	18560.6	18871.3	34868.6	34698.0
R44S	5402.8	5479.4	6793.8	6858.3
R56S	13333.9	13710.5	19611.0	20147.0
R67S	2858.9	2828.3	3589.8	3513.7
R75S	2493.1	2574.5	4313.1	4494.5
H11	1436.3	1457.7	1331.0	1354.1
H16	7701.2	7782.8	10430.8	12160.6
H22	1570.5	1593.7	1529.5	1599.1
H27	4103.5	4013.3	4954.6	4599.3
H33	6785.6	6725.5	10170.3	9958.1
S1	1329.3	1373.9	1477.0	1471.3
S8	2837.2	3009.5	3655.5	3834.8
S13	5874.2	5814.7	7392.2	7429.4
S32	4354.7	4462.3	5585.3	5480.7
S43	3933.0	3857.2	6823.1	6887.6
S52	1386.5	1426.7	1347.7	1292.7

ANOVA - All sampling designs

Element Pb

Number of duplicate samples n=27

Classical results: Mean= 6748.926

Sums of squares- 5.5117619E+09 6.9734466E+08 4248204

sigma values (geochem, sampling, analysis) - 6822.083, 3588.099, 280.483

sigma (total) - 7713.232

Robust results: Mean= 4952.998

sigma values (geochem, sampling, analysis) - 3531.825, 1228.315, 105.573

sigma (total) - 3740.815

$U = 200s_{\text{meas}} / \text{mean} = 49.8 \%$ (Robust results)

ANOVA - Regular grid (single auger samples)

Element Pb

Number of duplicate samples: n=7

Classical results: Mean= 8661.144

Sums of squares- 2.1587251E+09 3.0519580E+08 1341235.9

sigma values (geochem, sampling, analysis) - 8890.841, 4663.882, 309.520

sigma (total) - 10044.633

Robust results: Mean= 7515.946

sigma values (geochem, sampling, analysis) - 7968.633, 1864.781, 157.621

sigma (total) - 8185.436

$U = 200s_{\text{meas}} / \text{mean} = 49.8 \%$ (Robust results)

ANOVA - Regular grid (five-fold composite samples)

Element Pb

Number of duplicate samples: n=9

Classical results: Mean= 8132.719

Sums of squares- 2.5996294E+09 3.5486635E+08 1259916.8

sigma values (geochem, sampling, analysis) - 8448.730, 4436.192, 264.566

sigma (total) - 9546.248

Robust results: Mean= 6093.206

sigma values (geochem, sampling, analysis) - 5520.622, 930.843, 134.599

sigma (total) - 5600.165

$U = 200s_{\text{meas}} / \text{mean} = 30.9 \%$ (Robust results)

ANOVA - Herringbone design

Element Pb

Number of duplicate samples: n=5

Classical results: Mean= 3847.355

Sums of squares- 9.4745022E+07 13179221 52928.76

sigma values (geochem, sampling, analysis) - 2046.489, 1046.930, 66.413

sigma (total) - 2299.694

Robust results: Mean= 3847.355

sigma values (geochem, sampling, analysis) - 2372.575, 953.784, 67.406

sigma (total) - 2557.999

$U = 200s_{\text{meas}} / \text{mean} = 49.7 \%$ (Robust results)

ANOVA - Stratified random design

Element Pb

Number of duplicate samples: n=6

Classical results: Mean= 5062.869

Sums of squares- 2.2843033E+08 24103318 1594129

sigma values (geochem, sampling, analysis) - 3615.485, 1526.639, 399.265

sigma (total) - 3944.841

Robust results: Mean= 5062.869

sigma values (geochem, sampling, analysis) - 4097.672, 1758.070, 103.083

sigma (total) - 4460.084

$U = 200s_{\text{meas}} / \text{mean} = 69.6 \%$ (Robust results)

SAMPLE I.D.	S1A1Cu ($\mu\text{g g}^{-1}$)	S1A2Cu ($\mu\text{g g}^{-1}$)	S2A1Cu ($\mu\text{g g}^{-1}$)	S2A2Cu ($\mu\text{g g}^{-1}$)
R11C	23.37	23.60	21.45	21.98
R16C	31.37	31.29	34.20	31.58
R22C	23.87	25.52	25.46	25.11
R27C	22.24	21.74	26.12	25.77
R36C	26.72	27.06	27.40	27.33
R44C	26.77	25.91	24.69	23.77
R56C	31.78	32.20	31.79	32.54
R67C	27.83	29.32	20.63	20.46
R75C	23.04	23.68	28.12	26.74
R11S	24.77	24.72	23.38	22.21
R27S	20.63	21.11	26.73	26.99
R36S	25.11	25.30	27.92	27.75
R44S	28.40	28.44	23.23	23.39
R56S	32.72	32.86	32.63	35.00
R67S	30.30	29.36	21.52	20.16
R75S	23.10	23.17	24.85	25.47
H11	20.53	20.79	21.65	22.08
H16	36.07	35.84	34.57	31.90
H22	21.82	21.32	20.82	21.94
H27	34.74	33.20	24.90	22.87
H33	23.49	23.30	24.32	25.33
S1	22.61	23.38	21.46	20.89
S8	17.77	18.84	18.47	19.31
S13	43.06	43.23	39.83	40.27
S32	30.64	31.06	21.92	20.88
S43	51.74	50.48	36.31	36.73
S52	25.83	26.38	27.25	26.93

ANOVA - All sampling designs

Element Cu

Number of duplicate samples n=27

Classical results: Mean= 27.040931

Sums of squares- 3572.223 695.0069 25.428101

sigma values (geochem, sampling, analysis) - 5.283, 3.555, 0.686

sigma (total) - 6.405

Robust results: Mean= 26.271772

sigma values (geochem, sampling, analysis) - 4.051, 2.670, 0.607

sigma (total) - 4.889

$U = 200s_{\text{meas}} / \text{mean} = 20.8 \%$ (Robust results)

ANOVA - Regular grid (single auger samples)

Element Cu

Number of duplicate samples: n=7

Classical results: Mean= 26.115003

Sums of squares- 267.7107 158.68297 5.260297

sigma values (geochem, sampling, analysis) - 2.343, 3.339, 0.613

sigma (total) - 4.124

Robust results: Mean= 25.300902

sigma values (geochem, sampling, analysis) - 0.000, 3.474, 0.451

sigma (total) - 3.503

$U = 200s_{\text{meas}} / \text{mean} = 26 \%$ (Robust results)

ANOVA - Regular grid (five-fold composite samples)

Element Cu

Number of duplicate samples: n=9

Classical results: Mean= 26.456946

Sums of squares- 373.2888 107.31079 8.715252

sigma values (geochem, sampling, analysis) - 2.947, 2.392, 0.696

sigma (total) - 3.859

Robust results: Mean= 26.168552

sigma values (geochem, sampling, analysis) - 2.908, 2.078, 0.609

sigma (total) - 3.625

$U = 200s_{\text{meas}} / \text{mean} = 16.5 \%$ (Robust results)

ANOVA - Herringbone design

Element Cu

Number of duplicate samples: n=5

Classical results: Mean= 26.074001

Sums of squares- 515.6342 112.63868 8.243702

sigma values (geochem, sampling, analysis) - 5.157, 3.294, 0.908

sigma (total) - 6.186

Robust results: Mean= 26.074001

sigma values (geochem, sampling, analysis) - 6.303, 1.711, 0.893

sigma (total) - 6.592

$U = 200s_{\text{meas}} / \text{mean} = 14.8 \%$ (Robust results)

ANOVA - Stratified random design

Element Cu

Number of duplicate samples: n=6

Classical results: Mean= 29.802918

Sums of squares- 2177.521 316.3745 3.208850

sigma values (geochem, sampling, analysis) - 9.782, 5.122, 0.517

sigma (total) - 11.054

Robust results: Mean= 29.802918

sigma values (geochem, sampling, analysis) - 11.365, 4.605, 0.584

sigma (total) - 12.276

$U = 200s_{\text{meas}} / \text{mean} = 31.2 \%$ (Robust results)

Appendix 3-2. Duplicate data and ANOVA results from the Portable XRF experiment in the Derbyshire site.

3.2.1. P-XRF field measurements

SAMPLE I.D.	S1Pb ($\mu\text{g g}^{-1}$)	S2Pb ($\mu\text{g g}^{-1}$)
B11	1005	1633
B12	4631	3723
B13	1415	2264
B14	865	1350
B15	2899	2216
B16	721	1758
B21	2122	1014
B22	1321	1043
B23	3348	3904
B24	11543	5570
B25	2904	2833
B26	2617	2762
B32	976	786
B33	6127	3874
B34	331	576
B35	12878	8948
B44	3246	4332
B45	9006	6098
B46	1936	1989
B54	5811	6289
B55	4611	2880
B56	1326	1442
B64	1215	2713
B65	2070	2305

ANOVA - Portable XRF (field measurements)

Element Pb

n= 24

Classical results: Mean= 3275.542

Sums of squares- 6.4625465E+08 7.6826864E+07
sigma values (geochem, measurement)- 2494.84, 1265.132
sigma (total)- 2797.282

Robust results: Mean= 2856.595

sigma values (geochem, measurement)- 1893.474, 785.614
sigma (total)- 2049.984

U= 200smeas. / mean = 55% (Robust results)

SAMPLE I.D.	S1Zn ($\mu\text{g g}^{-1}$)	S2Zn ($\mu\text{g g}^{-1}$)
B11	202	181
B12	246	285
B13	237	257
B14	359	233
B15	395	384
B16	410	395
B21	247	179
B22	220	119
B23	369	510
B24	821	435
B25	303	320
B26	288	358
B32	214	169
B33	319	347
B34	185	103
B35	693	646
B44	417	406
B45	963	613
B46	359	531
B54	338	373
B55	478	564
B56	319	328
B64	236	488
B65	257	260

ANOVA - Portable XRF (field measurements)

Element Zn

n= 24

Classical results: Mean= 361.6459

Sums of squares- 2395089 443637

sigma values (geochem, measurement)- 146.330, 96.138

sigma (total)- 175.085

Robust results: Mean= 343.1936

sigma values (geochem, measurement)- 131.824, 63.086

sigma (total)- 146.142

U= 200smeas. / mean = 37% (Robust results)

3.2.2 P-XRF laboratory measurements

SAMPLE I.D.	S1A1Pb ($\mu\text{g g}^{-1}$)	S1A2Pb ($\mu\text{g g}^{-1}$)	S2A1Pb ($\mu\text{g g}^{-1}$)	S2A2Pb ($\mu\text{g g}^{-1}$)
B24	4721	4901	8380	8140
B54	10060	10000	9620	9640
B35	24670	23870	16530	16240
B56	1942	2019	2427	2359
B13	5661	5651	4041	4219
B45	20090	20280	14610	14470
B34	415	501	2427	2367
B55	9290	9370	4854	4780
B12	8220	9100	7540	7580
B22	2033	2115	1518	1596

ANOVA P-XRF laboratory measurements

Element Pb

Number of duplicate samples n=10

Classical results: Mean= 7956.175

Sums of squares- 1.4834468E+09 1.3419787E+08 863908.5

sigma values (geochem, sampling, analysis) - 6152.390, 2586.174, 207.835

sigma (total) - 6677.080

Robust results: Mean= 7369.509

sigma values (geochem, sampling, analysis) - 5742.615, 2471.017, 114.017

sigma (total) - 6252.723

$U = 200s_{\text{meas}} / \text{mean} = 67.1 \%$ (Robust results)

SAMPLE I.D.	S1A1Zn ($\mu\text{g g}^{-1}$)	S1A2Zn ($\mu\text{g g}^{-1}$)	S2A1Zn ($\mu\text{g g}^{-1}$)	S2A2Zn ($\mu\text{g g}^{-1}$)
B24	662	645	587	681
B54	594	599	600	566
B35	1142	1034	998	1085
B56	555	578	587	515
B13	510	509	451	442
B45	2101	1872	1183	1364
B34	241	278	309	293
B55	678	649	706	730
B12	369	382	418	438
B22	332	250	196	238

ANOVA P-XRF laboratory measurements

Element Zn

Number of duplicate samples n=10

Classical results: Mean= 659.1750

Sums of squares- 6119900 528227.8 66317.51

sigma values (geochem, sampling, analysis) - 395.969, 157.332, 57.584

sigma (total) - 429.954

Robust results: Mean= 597.2188

sigma values (geochem, sampling, analysis) - 308.011, 27.930, 45.030

sigma (total) - 312.536

$U = 200s_{\text{meas}} / \text{mean} = 17.7 \%$ (Robust results)

3.2.3 Laboratory ICP-AES measurements.

SAMPLE I.D.	S1A1Pb ($\mu\text{g g}^{-1}$)	S1A2Pb ($\mu\text{g g}^{-1}$)	S2A1Pb ($\mu\text{g g}^{-1}$)	S2A2Pb ($\mu\text{g g}^{-1}$)
B12	9641	9602	8026	7990
B13	449	451	2533	2654
B22	2205	2225	1662	1612
B24	4791	4928	8361	8791
B34	449	451	2533	2654
B35	26102	26949	16799	17658
B45	21449	23833	14814	15040
B54	11016	10708	10245	10368
B55	9775	9514	5020	4957
B56	1998	2073	2365	2505

ANOVA - ICP-AES measurements

Element Pb

Number of duplicate samples n=10

Classical results: Mean= 8129.849

Sums of squares- 1.8887971E+09 1.9402168E+08 3816912

sigma values (geochem, sampling, analysis) - 6900.439, 3099.300, 436.859

sigma (total) - 7577.108

Robust results: Mean= 7570.399

sigma values (geochem, sampling, analysis) - 6734.701, 2702.919, 154.140

sigma (total) - 7258.493

$U = 200s_{\text{meas}} / \text{mean} = 71.5 \%$ (Robust results)

SAMPLE I.D.	S1A1Zn	S1A2Zn	S2A1Zn	S2A2Zn
B12	364.321	362.172	438.36	471.817
B13	521.969	516.13	408.8	421.323
B22	239.814	273.881	247.422	241.931
B24	686.035	731.541	619.312	624.394
B34	209.741	211.891	356.168	379.47
B35	1174.63	1185.78	1076.52	1137.82
B45	2082.26	2193.33	1308.37	1329.21
B54	640.03	591.49	559.278	604.858
B55	728.779	785.59	712.973	706.576
B56	480.13	511.569	564.576	572.719

ANOVA - ICP-AES measurements

Element Zn

Number of duplicate samples n=24

Classical results: Mean= 681.8246

Sums of squares- 7203779 736390.1 15279.763

sigma values (geochem, sampling, analysis) - 426.257, 190.885, 27.640

sigma (total) - 467.864

Robust results: Mean= 609.6934

sigma values (geochem, sampling, analysis) - 322.659, 74.895, 25.240

sigma (total) - 332.197

$U = 200s_{\text{meas}} / \text{mean} = 25.9 \% \text{ (Robust results)}$

Appendix 3-3. Duplicate data and ANOVA results from the initial survey in the Hounslow Heath site.

SAMPLE I.D.	S1A1Pb ($\mu\text{g g}^{-1}$)	S1A2Pb ($\mu\text{g g}^{-1}$)	S2A1Pb ($\mu\text{g g}^{-1}$)	S2A2Pb ($\mu\text{g g}^{-1}$)
HA4	787	769	811	780
HB7	338	327	651	563
HC1	289	297	211	204
HD9	662	702	238	246
HE8	229	215	208	218
HF7	346	374	525	520
HG7	324	321	77	73
HH5	56	61	116	120
HI9	189	189	176	168
HJ5	61	61	91	119

ANOVA - Regular grid at Hounslow Heath site

Element Pb

Number of duplicate samples n=10

Classical results: Mean= 317.8000

Sums of squares- 1738031.8 370075.5 6473.000

sigma values (geochem, sampling, analysis) - 197.552, 135.432, 17.990

sigma (total) - 240.192

Robust results: Mean= 297.3088

sigma values (geochem, sampling, analysis) - 179.674, 123.814, 11.144

sigma (total) - 218.488

$U = 200s_{\text{meas}} / \text{mean} = 83.6 \%$ (Robust results)

SAMPLE I.D.	S1A1Zn ($\mu\text{g g}^{-1}$)	S1A2Zn ($\mu\text{g g}^{-1}$)	S2A1Zn ($\mu\text{g g}^{-1}$)	S2A2Zn ($\mu\text{g g}^{-1}$)
HA4	624	625	599	602
HB7	179	177	255	247
HC1	439	448	319	311
HD9	228	246	179	177
HE8	157	150	154	159
HF7	299	309	425	421
HG7	219	214	114	115
HH5	92	96	181	184
HI9	374	367	322	317
HJ5	110	110	76	76

ANOVA - Regular grid at Hounslow Heath site

Element Zn

Number of duplicate samples n=10

Classical results: Mean= 267.4000

Sums of squares- 893441.2 62061.50 433.0000

sigma values (geochem, sampling, analysis) - 152.533, 55.608, 4.653

sigma (total) - 162.420

Robust results: Mean= 250.92950

sigma values (geochem, sampling, analysis) - 130.942, 63.055, 4.468

sigma (total) - 145.402

$U = 200s_{\text{meas}} / \text{mean} = 50.4 \% \text{ (Robust results)}$

APPENDIX 4

Statistical Tests

Appendix 4-1. Student's *t*-test

The *t*-test is based on the *t* probability distribution, which is similar to the normal distribution, but dependent upon the size of sample taken. When the number of observations in the sample is infinite, the *t* distribution and the normal distribution are identical (Davis, 1986). The *t*-test is useful for testing hypotheses about the equivalency of two samples (e.g. examining the significance of analytical bias by comparison with a CRM or for the comparison of mean metal concentrations estimated from different sampling designs)

The *t* value is calculated as:

$$t_{calc} = (X - \mu_0) / s \sqrt{1/n}$$

where,

X = mean of the sample

μ_0 = hypothetical mean of population

n = number of observations

s = standard deviation of observations

This value is then compared with the *t* value from a Table for specified degrees of freedom and level of confidence:

Number of Degrees of Freedom, ν	Significance Level, α (%)					
	10	5	2.5	1	0.5	0.1
1	3.078	6.314	12.706	31.821	63.657	318.310
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
∞	1.282	1.645	1.960	2.326	2.576	3.090

Source: From Table 21, *The Penguin-Honeywell Book of Tables*, copyright F. W. Kellaway (ed.) and Honeywell Controls Ltd. (E.D.P. Division), 1968.

Appendix 4-2. Mann-Whitney Test

The Mann-Whitney test can be used as a nonparametric substitute for the t test of the equality of the means of two samples. Thus it is used for distributions that are not normal (Davis, 1986). Critical values for T are given in the following Table:

Critical Values for T for the Mann-Whitney Test; Values are for the Lower Critical Limit; Corresponding Limit for the Upper Critical Area is Given by $T_{1-\alpha} = nm - T_{\alpha}$

n	α	$m = 2$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	.01	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	2	2
	.05	0	0	0	1	1	1	2	2	2	2	3	3	4	4	4	4	5	5	5
	.10	0	1	1	2	2	2	3	3	4	4	5	5	5	6	6	7	7	8	8
3	.01	0	0	0	0	0	1	1	2	2	2	3	3	3	4	4	5	5	5	6
	.05	0	1	1	2	3	3	4	5	5	6	6	7	8	8	9	10	10	11	12
	.10	1	2	2	3	4	5	6	6	7	8	9	10	11	11	12	13	14	15	16
4	.01	0	0	0	1	2	2	3	4	4	5	6	6	7	9	8	9	10	10	11
	.05	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19
	.10	1	2	4	5	6	7	8	10	11	12	13	14	16	17	18	19	21	22	23
5	.01	0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	.05	1	2	3	5	6	7	9	10	12	13	14	16	17	19	20	21	23	24	26
	.10	2	3	5	6	8	9	11	13	14	16	18	19	21	23	24	26	28	29	31
6	.01	0	0	2	3	4	5	7	8	9	10	12	13	14	16	17	19	20	21	23
	.05	1	3	4	6	8	9	11	13	15	17	18	20	22	24	26	27	29	31	33
	.10	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	35	37	39
7	.01	0	1	2	4	5	7	8	10	12	13	15	17	18	20	22	24	25	27	29
	.05	1	3	5	7	9	12	14	16	18	20	22	25	27	29	31	34	36	38	40
	.10	2	5	7	9	12	14	17	19	22	24	27	29	32	34	37	39	42	44	47
8	.01	0	1	3	5	7	8	10	12	14	16	18	21	23	25	27	29	31	33	35
	.05	2	4	6	9	11	14	16	19	21	24	27	29	32	34	37	40	42	45	48
	.10	3	6	8	11	14	17	20	23	25	28	31	34	37	40	43	46	49	52	55
9	.01	0	2	4	6	8	10	12	15	17	19	22	24	27	29	32	34	37	39	41
	.05	2	5	7	10	13	16	19	22	25	28	31	34	37	40	43	46	49	52	55
	.10	3	6	10	13	16	19	23	26	29	32	36	39	42	46	49	53	56	59	63
10	.01	0	2	4	7	9	12	14	17	20	23	25	28	31	34	37	39	42	45	48
	.05	2	5	8	12	15	18	21	25	28	32	35	38	42	45	49	52	56	59	63
	.10	4	7	11	14	18	22	25	29	33	37	40	44	48	52	55	59	63	67	71

Abridged from Conover (1980), Table 8.

Appendix 4-3. The F test.

The F -test is used to determine equality of variances and it is based on the F probability distribution, which describes the probabilities of obtaining specified ratios of sample variances drawn from the same population (Davis, 1986). The F ratio between the variances is calculated as:

$$F = s^2_1 / s^2_2$$

The calculated value is then compared to tabulate values of F ratios given in the following Table:

Critical Values of F for ν_1 and ν_2 Degrees of Freedom and 5% ($\alpha = 0.05$) Level of Significance															
Degrees of Freedom for Denominator, ν_2	Degrees of Freedom for Numerator, ν_1														
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.74	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46

Source: From Table 22, *The Penguin-Honeywell Book of Tables*, copyright F. W. Kellaway (ed.) and Honeywell Controls Ltd. (E.D.P. Division), 1968.

Appendix 4-4. Dixon's Q -test

Dixon's Q is used for rejecting outlying results. To assess a suspect measurement the difference between it and the measurement nearest to it in size is compared with the difference between the highest and lowest measurements. The ratio of these differences (without regard to sign) gives the Q value (Dixon, 1950):

$$Q = | \text{suspect value} - \text{nearest value} | / (\text{largest value} - \text{smallest value})$$

The critical values of Q are given in the following table:

n	10%	5%	1%
3	0.886	0.941	0.988
4	0.679	0.765	0.889
5	0.557	0.642	0.780
6	0.482	0.560	0.698
7	0.434	0.507	0.637
8	0.479	0.554	0.683
9	0.441	0.512	0.635
10	0.409	0.477	0.597

LIST OF PUBLICATIONS

Argyraki, A., Ramsey, M. H., and Potts, P.J., "Evaluation of Portable XRF for in-situ Measurements of Lead on Contaminated Land", *Analyst* (in press).

Ramsey, M.H., and Argyraki, A., "Estimation of Measurement Uncertainty from Field Sampling: Implications for the Classification of Contaminated Land", *The Science of the Total Environment* (in press).

Argyraki, A, and Ramsey, M.H., "Evaluation of Inter-Organisational Sampling Trials on Contaminated Land: Comparison of Two Contrasting Sites", Geological Society, Special Publication. (in press).

Argyraki, A., Ramsey, M. H., and Thompson, M., (1995), "Proficiency Testing in Sampling: Pilot Study on Contaminated Land", *Analyst*, **120**, 2799-2803.

Ramsey, M.H., Argyraki, A., and Thompson, M., (1995), "On the Collaborative Trial in Sampling", *Analyst*, **120**, 2309-2312.

Ramsey, M.H., Argyraki, A., and Thompson, M., (1995), "Estimation of Sampling Bias Between Different Sampling Protocols on Contaminated land", *Analyst*, **120**, 1353-1356.