Measurement Modelling and Mapping of Arsenic Bioaccessibility in Northampton, UK

MARK R. CAVE, JOANNA WRAGG and HEATHER HARRISON

British Geological Survey, Keyworth, Nottingham, UK, NG12 5GG

ABSTRACT

The human ingestion bioaccessibility of As was measured on 50 representative samples of soils selected from a 281 soil sample geochemical survey of Northampton. The major and trace element content, pH and near infra-red (NIR) spectra of the 281 soils were determined. A multiple linear regression (MLR) model using total As, major element composition and pH identified total As, pH and P to be the significant predictor variables for bioaccessible As (r square = 0.72, median standard error of prediction=1.5 mg kg^{-1} bioaccessible As). When spectral components (SC) derived from chemometric analysis of the NIR spectra were also included in the MLR, total As, pH, Mg and two NIR spectral components were found to be significant predictor variables (r square = 0.84, median standard error of prediction=1.2 mg kg^{-1} bioaccessible As). Correlation analysis of the SC with major element data suggested that the two NIR SC in the second model were related to different forms of Fe oxides in the soil. When plotted over a geological map of Northampton interpolated predictions of bioaccessible As showed clear geological control. The median total As concentration of the soils in Northampton was 30.2 mg kg^{-1} and the median bioaccessible As was 3.0 mg kg^{-1}.

*Address correspondence to Mark Cave, British Geological Survey, Keyworth, Nottingham, UK;E-mail:mrca@bgs.ac.uk

Keywords: Bioaccessibility, arsenic, soil, near infra-red

INTRODUCTION

For the UK, there are extensive areas of the country with naturally elevated concentrations of total arsenic (As) [1]. A study by Appleton et al., [2] estimating UK soil As concentrations and a recently published BGS geochemical soil atlas of England and Wales [3] using soil samples collected for the National Soil Inventory (NSI), highlighted the extent to which the soil guideline value (SGV) for As of 32 mg kg^{-1} for residential land use [4] was exceeded.

Rawlins et al., [5] reported that in recently glaciated areas such as the British Isles parent material (PM) is the primary control on soil geochemistry. For elevated soil As concentrations in rural soils, studies have shown this to generally be the case [1]. In contrast, where high As concentrations are present in urban soils, this can be a reflection of both the PM and Britain's industrial past, with contaminant sources resulting from industries such as mining and mineral processing as well as metal production.

Since As is a highly toxic element, it is important to understand the exposure of humans to this element from their interaction with soil, avoiding unnecessary public concern about its presence in the environment. Despite available guidance, there are concerns that risks to human health can be under or over-estimated at sites [6]. This may be due to the application of total concentrations alone and the many variations in soil properties, which can affect the fate and transport of individual contaminants [7; 8]. For As in soil, the human ingestion pathway is the primary route of exposure [9]. Therefore, the fraction of a soil contaminant that is soluble in the gastro-intestinal tract and is available for uptake into the circularory sytem (known as the bioaccessible fraction) can be used as a measure of exposure to As.

A good example of an urban centre in the UK, which has elevated As concentrations due to parent material control and likely point and diffuse pollution from industrial sources, is Northampton. Northampton is a large market town in central England (Figure 1), with a

population of ca. 200, 000, located on Jurassic ironstones (Northampton Sand Formation and Marlstone Rock) with As concentrations exceeding 30 mg kg^{-1}. Figure 2 shows the Jurassic system PM across the Northampton area, highlighting the Lias clays, the Oolite group and the Northampton Sand Formation [10]. The industrial development of the county of Northamptonshire, including the growth of Northampton, was supported by open cast quarrying of the abundant iron ore in the middle of the 19th century [11] and the ease of access to two major railways (the London & North Western and the Midland Railways). There were also considerable leather currying and tanning works, breweries, iron foundries, and brick and tile works distributed across the area. It is, however, the shoemaking industry, which was very large, at one time employing 75% of the population of the county that is more often associated with Northampton. More recently Northampton has become a hub for the finance and distribution industries because of the good road and rail links afforded to businesses.

The purpose of this study is to gain information on the processes that control As bioaccessibility and estimate its spatial distribution across the Northampton urban area using a multi-parameter approach, which includes total element and bioaccessible As analysis, and near infra-red properties of the soil. With this approach, it is hoped that, in addition to gaining a process understanding, the possibilities of using simple field based methods for As bioaccessibility screening will be explored.

MATERIALS AND METHODS

Sample Collection and Preparation

In total 281 <2 mm topsoil (5–20 cm depth) samples were collected from the Northampton urban area under the British Geological Survey's Geochemical Baselines Project, G-BASE, as part of the urban geochemical surveying programme [12]. Samples were collected on a 500 m grid at a density of approximately 4 samples per km^2, from open ground and as close as possible to the centre of each 500 m grid cell. At each sample site, composite samples based on 5 sub-samples taken at the centre and four corners of a 20 m square. Samples were stored in uniquely labelled kraft bags and dried in a calibrated fan assisted oven, set at 35 ± 2ºC, on metal trays for at least 12 hours until visually dry. Dried samples were gently disaggregated to ensure the breakage of aggregates but retention of clasts [11] and sieved to <2 mm according to British Standard BS 933-1[13].

Soil Geochemical Characterisation

Total element concentrations

Approximately 40 major and trace elements were determined in the 281 <2mm size fraction soils by X-ray fluorescence analysis. Full details of the analytical method and quality control procedures are described in [14]. The resulting geochemical data set was subjected to hierarchical clustering in order to reduce the total number of soil samples for further to 50 bioaccessibility and near infra-red determinations. The clustering allowed the reduced sample batch to stay representative of the different geochemical groupings in the Northampton urban area as a whole (see the Results section for further details).

Soil pH and Organic Matter Content

Soil pH and organic matter (OM) content were determined on each of the original 281 urban Northampton soils as part of the G-BASE project. Both methods are described in full [14] but simply: soil pH was measured using a recognised standard protocol [15], $CaCl_2$ at 0.01M and at a soil:solution ratio of 1:10; and an estimate of OM was measured by calculating the loss on ignition at 450ºC.

Bioaccessible element concentrations

The bioaccessible As concentration of the <250 μm fraction of the fifty Northampton urban soils, selected using cluster analysis, was determined using the Unified BARGE Method (UBM). The UBM is an *in vitro* gastro-intestinal (GI) simulation developed to assess the potential availability of contaminants via accidental ingestion of soil and provide additional lines of evidence for human health risk assessments. The methodology is based on that first described [16] and was modified in order to ensure adequate conservatism and robustness whilst still being physiologically based and applicable to the different soil types found in a range of different countries. Details of the development of the UBM have been described by Wragg et al [17], its *in vivo* validation [18] and its application to As and other soil elements of concern by [19-22]. The UBM was applied to samples collected prior to validation of the methodology. As such the UBM applied to the Northampton samples is that detailed by [17; 23] and not subsequent versions which include:

• Fixation of the initial stomach pH at 1.2; and

• Using a centrifugation speed and time of 4500 g for 15 minutes.

The UBM was carried out at 37°C (body temperature), at a final soil:solution ratio of 1:100 (g mL^{-1}) with end over end rotation at 30 rpm. The simulation consisted of three stages: the mouth, stomach and small intestinal using simulated saliva and gastric fluids and bile plus duodenal fluids to represent the intestinal fluid. The extraction was carried out under fasting conditions using saliva, gastric and intestinal pHs of 6.5 ± 0.5, $1.2 - 1.4$ and 6.3 ± 0.5. Two individual extracts per test sample were produced for analysis. The first, called 'stomach' was sampled after completion of the saliva and stomach phases. The second called 'stomach & intestine' was collected after a second aliquot of the original soil sample was subjected to the saliva and stomach phases, followed by extraction under small intestinal conditions using simulated bile and duodenal fluids. Sample collection was undertaken by separation by centrifugation for 5 minutes at 3000 g, removal of a 1.0 mL aliquot of supernatant by pipette and preservation of the aliquot in 9.0 mL of 0.1M HNO$_3$ followed by refrigeration at 1 - 8°C until time of analysis. All extraction reagents were of analytical grade or higher and sourced from Sigma, Merck, J T Baker, or Fluka.

Analysis of the bioaccessibility extractions

The preserved UBM extractions were analysed for their bioaccessible As content using a Thermo Elemental ExCell quadrupole Inductively Couple Plasma – Mass Spectrometry (ICP-MS) instrument in combination with a Cetac ASX-510 autosampler, according to the operating conditions previously described by [24]. The instrument was calibrated at the

beginning of each analytical run using standards prepared from certified Claritas PPT® (Spex CertiPreP) multi-element standards in the range 0 to 50 $\mu g\ L^{-1}$. In addition, 10 $\mu g\ L^{-1}$ mixed element standards were inserted at regular intervals throughout the analysis run and used to correct for any drift in instrument sensitivity. Indium and rhenium were added to all solutions via a T-piece connection and used as internal standards to correct for any matrix suppression. Multi-element quality control (QC) standards were analysed after no more than every 20 unknown samples. All of the UBM extracts were diluted by a factor of ten with 1% HNO_3 prior to analysis. All element concentrations quoted have been converted into mg kg^{-1} extracted from the solid.

Bioaccessibility Quality Control

Within every batch, of a maximum of 10 unknown samples, one duplicate, one quality control soil and one blank were extracted. At present there are no certified reference materials for use in bioaccessibility studies. The QC of the UBM was monitored by carrying out replicate analyses of a guidance soil, BGS 102 [25], which has bioaccessibility reference values for a number of potentially harmful elements (PHE), which were derived in an international inter-laboratory trial of the UBM [17].

Near Infra-red Analysis

Analysis was carried out on the <2mm soil fraction. Optimum signal to noise was obtained with an average of 25 NIR reflectance scans for each replicate soil sample over a spectral range of 350-2500 cm^{-1} was obtained using an AgriSpec ASP350-2500. At the start of every analytical run and after every 20 readings, a white reference was run. Samples were scanned in duplicate with a 90° shift in orientation between measurements (to take into account inhomogeneity in the soil) with an average of the two spectra taken and converted from reflectance into absorbance units.

Data Modelling

NIR spectra

The NIR spectra were subjected to a self modelling mixture resolution (SMMR) process as previously described by Cave [26] which gave rise to five distinct spectral components (SCs) in the 281 spectra of the 2mm sieved soil samples. At this stage these represent the underlying spectra of soil components which combine together to give the measured spectra of the soils. NIR spectra of soils are relatively featureless with the information relating to soil properties associated with shifts in background. The physical meaning of the spectra are discussed in the results section.

Multiple linear regression modelling

The relationship between the bioaccessible As, as measured by the UBM, and the chemical composition of the soils was assessed using simple multiple linear regression (MLR) with bioaccessible As as the dependant variable and the total As concentration, major element composition (Na, Mg, Al, Si, Ti, Mn, and Fe), pH and spectral components as the explanatory variables. The statistical analysis was carried out in the R programming language [27] and associated packages. Although a very powerful technique for looking at relationships, MLR modelling has a number of pitfalls that can lead to misinterpretation of the results [28]. In order to ensure a robust parsimonious model the following procedure was adopted:

i) The optimum subset of the explanatory variables was obtained by an exhaustive search of all combinations of the variables picking the model with the lowest Bayesian Information Criterion (BIC) using the R programming language "leaps" package. The BIC is a measure of the goodness of fit of an estimated statistical model, describing the trade off between the precision and complexity of the model [29];

ii) The optimum model was checked to see if all the selected explanatory variables had significant coefficients (*i.e.* p values <0.05) discarding those which were not significant;

iii) An analysis of variance on the linear regression model was carried out to see if the variance explained by each of the remaining explanatory variables was significant discarding those that were not significant;

iv) Finally the model was checked to see that the residuals were approximately normally distributed and that there was no significant co-linearity in the model by checking the variance inflation factors of each explanatory variable [28].

Once the MLR model had been through the checking procedure, the predictive capability of the model was estimated using cross validation (CV). The training and testing data sets for the CV were obtained by four fold selection where the data are randomly split into 4

approximately equal data sets using the combined data from 3 of the sets to predict the As bioaccessibility on the 4th. The predicted and actual values are then compared to give a value for the standard error of prediction (SEP) defined as the standard deviation of the differences between the predicted values and the reference values for a set of samples [30]. This process was repeated 1000 times to give a distribution of SEP values which was used to assess the predictive capability of the model.

RESULTS

Table 1 gives the total As and the bioaccessible As concentration in the 50 selected soil samples

Quality control of bioaccessibility extractions

Results for the QC sample BGS 102, extracted as part of each bioaccessibility extraction batch, were assessed against previously reported consensus values (Wragg et al., 2009). Comparison of bioaccessibility results indicated that the As data obtained was in good agreement with the consensus values. For the 'stomach' and the 'stomach & intestine' compartments values of 4.7 ± 0.5 mg kg^{-1} and 3.3 ± 0.3 mg kg^{-1} were obtained respectively compared to guidance values of 4.5 ± 1.3 mg kg^{-1} and 5.4 ± 2.4 mg kg^{-1} respectively.

The mean repeatability of BGS 102 (n=15) for the 'stomach' and 'stomach & intestine' compartments was 9.9 % and 7.7 % respectively.

The mean repeatability for the duplicate soils (n=7) extracted from Northampton was 2.7 % and 0.3 % respectively in the 'stomach' and in the 'stomach & intestine' compartments of the UBM.

The method detection limits were calculated as five times the average blank measurement. The calculated detection limits for the two extraction phases equate to 0.5 mg kg^{-1} and 1.5 mg kg^{-1} in the 'stomach' and 'stomach & intestine' compartments of the UBM respectively. All blank UBM extractions returned values below the method detection limits.

Clustering of the total element soil data

The data were mean centred and scaled, and clustered using Euclidean distance and Ward's method linkage resulting in a dendrogram from which a visual inspection suggests the existence of 4 distinct clusters. Figure 3 shows the estimated probability density functions for 4 test elements in each of the four groupings, which illustrate the geochemistry of the four groups.

Group 4 (n=38) has the highest Fe content and represents samples with high iron oxide content. Group 3 (n=81) has high the highest average K and Mg and Figure 3 show it has the highest Rb content. All these elements are commonly associated with clay minerals suggesting these soils are more clay rich [31]. The high Si content of Group 1 (n=40) soils suggest that these are a more highly quartz dominated. Group 2 (n=116) soils do not have a particularly distinct geochemical signature compared to the others. Figure 3 shows that for Fe, Rb and Si, Group 2 soils are all in the mid range of the group distributions although a

comparison of Ca distributions shows it contains some high Ca samples compared to the other groups. Group 2 can therefore be considered as geochemically "average" soils for Northampton.

Fifty samples were chosen for further preparation and testing from these four groups based approximately on the number of samples found in each cluster (7 from cluster 1, 20 from cluster 2, 13 from cluster 3 and 10 from cluster 4).

Geochemistry of the soils

A simple review of potential relationships between the major elements, soil pH and total As was carried out using the Pearson Correlation Coefficient. The results are plotted pictorially in Figure 4 as a correlogram [32] where the bottom left triangle is made up of rectangles with forward hash shading representing positive correlation and backward hash shading representing negative correlation and the darkest gray values representing the highest negative or positive correlation value.

The top right triangle of pie charts represents the magnitude of the correlation, with positive correlations represented by the shaded portion of the pie in clockwise direction and negative correlations by the shaded part of the pie in an anti-clockwise direction with darkest gray shading representing highest correlations.

In addition, the correlations between variables in Figure 4 are re-ordered so that "similar" variables are positioned adjacently based on the principal component analysis of the correlation matrix loadings. The correlogram Figure 4 shows that As is correlated with Fe, Mn, and P suggesting that As is mostly associated with the iron oxide fraction in the soil [33].

There is a close association between Ti, Al, Mg and K which is suggestive of a clay mineral grouping and the correlation between Ca and pH is suggestive that the carbonate content of the soils is the main control on soil pH.

MLR modelling of bioaccessible As

The bioaccessible fraction of As was measured on the particle size range which is most likely to adhere to hands (<250 µm) and therefore be ingested through hand to mouth activity[34]. The total As in Northampton soil survey were, however, measured on the <2 mm fraction. If predictions of the bioaccessible As in the <250 µm fraction were to be made for the whole of the Northampton soils (without having to sieve and re-analyse all samples) then the model would need to predict the <250 µm bioaccessible As fraction using the total As in the <2 mm fraction as one of the predictor variables. In order for this to be possible it would be necessary for there to be a simple and predictable relationship between the two size fractions. Figure 5 shows the total As concentration in the two size fraction plotted against each other, showing there is a strong linear relationship with a slope of 0.95 and intercept not significantly different from zero. Since the <250 µm is a simple fraction of the <2 mm total As, the MLR model can proceed using the <2mm total As as a predictor variable for the <250 µm bioaccessible As for these Northampton soils.

Two MLR models were made using the predictor selection criteria set out in the Methods section:

1. Model 1 - used bioaccessible As as the dependent variable and the total As concentration, major element composition (K, Na, Mg, Al, Si, Ti, Mn,P and Fe), pH as the explanatory variables; and

2. Model 2 - in which spectral components 1-5 were used in addition to As, major elements and pH as the explanatory variables.

For Model 1 the first stage exhaustive search identified pH, Mg, P, Fe, Mn and As as the pedictor variables which gave the lowest BIC model. Subsequent testing showed that variance explained by Fe and Mn was not significant and, when these were removed, the Mg coefficient became insignificant. This model passed the residual and co-linearity checks, Table shows the summary statistics for Model 1. The predictor variables for Model 1 along with the major element groupings shown by the correlogram (Figure 4) provide a geochemically sound explanation for the bioacessible As fraction in the soil. The correlogram suggests that the majority of the total As is adsorbed to Fe oxides which is consistent with other studies of similar soils [35]. Although a negative intercept is difficult to assign to a physical meaning, the positive coefficients for As, pH and P composition can be explained as follows:

As – it is reasonable to suggest that the bioaccessible fraction is dependent on the total As in the soil.

P – phosphate, which is mostly associated with Fe oxides in soils developed over jurassic ironstones in the UK [36], adsorbs strongly to Fe oxides and displaces oxi-arsenic anions from Fe oxide surfaces [33]; therefore higher concentrations of P would point to the As being more mobile.

pH – If As is mostly held on Fe oxides, the adsorption Kd decreases with increasing pH [33], again leading to the As being more mobile at higher pH.

The predictive ability of Model 1 was assessed using four- fold cross validation with 1000 replications. The results are shown in Figure 6 with each of the points in the model showing the 1000 replications of its predicted value based on random selections of training data along with its median predicted value. For each of the 1000 replications a SEP was calculated with the 95 percentiles confidence limits on this value shown on the plot. The SEP is relatively stable suggesting that this is a reasonable model. Multiplying the SEP by 2 gives an approximate 95% confidence limit on the predictive ability of the model so that the bioaccessible As can be predicted with an accuracy of between 1.4 to 1.6 mg kg^{-1}.

For Model 2 the first stage exhaustive search identified pH, Mg, As, and spectral components 1, 2 and 4 as the predictor variables which gave the lowest BIC model. Subsequent testing showed that variance explained by spectral component (SC) 4 was not significant. This model passed the residual and co-linearity checks, Table shows the summary statistics for Model 2.

Model 2 gives a different view of the controlling factors on Bioacessibility than Model 1. Both pH and As are common between the models but Model 2 includes Mg and two spectral components. A geochemical explanation of Model 2 and how this fits in with Model 1 is as follows:

Model 2 (Table2) does not have a significant intercept, which is better for a physical explanation of the model.

In Model 1 Mg was initially selected in the exhaustive search but was shown not to be significant, in Model 2, however, Mg appears to be significant but only accounts for about 4% of the variance (Table 2). The correlogram of the total element data in the soil (Figure 4) suggests that Mg is associated more closely with clay minerals. Combining this information, a positive coefficient for Mg may suggest that there is a small amount of bioacessible As associated with clays

The explanations for As and pH in Model 2 follow the same reasoning as in Model 1, although the variance explained by Total As in Model 2 has increased from 8.9 to 37.6%, but the proportion of variance remains the same for pH (22.6%) (Tables 1 and 2).

To try and understand the relationship of the spectral components SC1 and SC2 and the soil composition a new correlogram (Figure 7) with both major elements, pH and SC1 and SC2 has been plotted. The inclusion of the spectral data has split the Fe oxide group into Fe, Mn, SC1 and As and second group consiting of P and SC2. A plausible explanation of this could be that because NIR spectra can differentiate between Fe oxide minerals [33] that SC1 and SC2 are associated with different Fe oxide components in the soil. Table 2 shows that SC1 has a negative coefficient so increasing the amount of this component decreases the bioaccessible As and vice versa for SC2. Previous studies [35; 36] have shown that As associated with aged crystaline forms of Fe oxide has very low mobility in ingestion bioaccessibility tests but that As associated with more newly formed fine grained Fe oxide is more mobile.

Using this premise, SC1 could be assigned to a crystaline Fe oxide and SC2 to a fine grained Fe oxide, particularly as fine grained Fe oxides have a higher affinity for phosphates than the crystaline form [33]. The addition of the spectral components to Model 2 allows for differentiation of forms of Fe oxide as controlling factors of the As bioaccessibility.

The cross validated prediction of bioaccessible As for Model 2 is shown in Figure 8. The scatter of points about the line of equality is less than that for Model 1 (Figure 6) which is reflected in a lower SEP so that using Model 2 the bioaccessible As can be predicted with an accuracy of between 1.0 to 1.4 mg kg^{-1}.

Spatial Distribution of Bioaccessible As in Northampton

Using Model 1 and Model 2, the bioaccessible As (B-As) concentration for all of the soil samples collected from Northampton can be predicted. Parent material (PM) soil chemistry mapping [10] was used to map the spatial distribution of B-As in topsoils interpolated to a 100 m – PM grid with the value for each polygon being the average of the modelled B-As concentrations in the nearest four soil samples located on the same PM (Figure 9 and Figure 10). The modelled B-As data were not strongly skewed when grouped by PM so interpolation was carried out on un-transformed data. Both maps clearly demonstrate the geological controls on the bioaccessible As values. Higher bioaccessible As is associated principally with the Northampton Sandstone Formation with generally lower bioaccessible As associated with the Great Oolite Group limestones, clays, sandstones and siltstones, Sand and Gravel and Glacial Till. There appear to be some areas of higher B-As caused by dispersion of material derived from the Northampton Sand Formation ironstones onto the adjacent Upper Lias and also onto the alluvium in the Nene valley and its tributaries. The maps show that the estimated bioaccessible As across Northampton is relatively low compared to the total As. This is clearly illustrated by a plot of the distributions of total As and bioaccessible As is shown in Figure 11.

Whilst this study does not constitute a risk assessment, it clearly shows that, although the some of the soils in Northampton are high in total As (*e.g.* 40-100 mg kg^{-1}), the As is not very mobile giving rise to bioaccessible As concentrations ≤ 5 mg kg^{-1}.

CONCLUSION

The use of simple MLR modelling gives a clear insight into the geochemical controls on the bioaccessibility of As in soils in Northampton, that is fully consistent with previous studies. Overlaying interpolated maps of bioaccessible As over the geology suggests that the main controls are from the underlying geology and not from anthropogenic sources.

In addition to the information on the processes governing As bioaccessibility, the study also illustrates how a predictive model for what seems to be mostly naturally occurring bioaccessible As can be set up using the total concentration of As, soil pH and combinations of the major element compositions and or NIR spectra. Given suitable equipment *e.g.* hand held XRF, portable NIR spectrometers, and pH probes, all of which are now commercially available, it would be feasible to make these measurements from a field laboratory and hence be able to predict and screen for As bioaccessibility on-site. This would have to be followed up with laboratory testing for validation purposes, but could potentially allow a site investigation to obtain much more detailed spatial resolution on As bioaccessibility and hence assess potential contamination problems in a robust and cost effective manner.

References

[1] Ander, E.L., Cave, M.R., Johnson, C.C. and Palumbo-Roe, B., Normal background concentrations of contaminants in the soils of England. Available data and data exploration. British Geological Survey; 2012.

[2] Appleton, J.D., Rawlins, B.G. and Thornton, I. National-scale estimation of potentially harmful element ambient background concentrations in topsoil using parent material classified soil:stream-sediment relationships. Appl. Geochem., **2008**, *23*, 2596-2611.

[3] Rawlins, B.G., McGrath, S.P., Scheib, A.J., Breward, N., Cave, M., Lister, T.R., Ingham, M., Gowing, C. and Carter, S., The Advanced Soil Geochemical Atlas of England and Wales. . British Geological Survey; Nottingham 2012.

[4] Department for the Environment Food and Rural Affairs and the Environment Agency, Soil Guideline Values for Arsenic Contamination. Department for the Environment, Food and Rural Affairs; 2002.

[5] Rawlins, B.G., Webster, R. and Lister, T.R. The influence of parent material on topsoil geochemistry in eastern England. Earth Surf. Process. Landf., **2003**, *28*, 1839-1409.

[6] Kelley, M.E., Brauning, S.E., Schoof, R. and Ruby, M.V. *Assessing Oral Bioavailability of Metals in Soil*; Battelle Press; Columbus Richland, 2002.

[7] Cave, M.R., Wragg, J., Denys, S., Jondreville, C. and Feidt, C., Oral Bioavailability. In *Dealing with Contaminated Sites: From Theory towards Practical Application*; Swartjes, F. Ed.; Springer; 2011.

[8] Selinus, O. *Essentials of Medical Geology Impacts of the Natural Environment on Public Health*; Elsevier Academic Press, 2005.

[9] Paustenbach, D.J. The practice of exposure assessment: A state-of-the-art review (Reprinted from Principles and Methods of Toxicology, 4th edition, 2001). J. Toxicol. Env. Health-Pt b-Crit. Rev., **2000**, *3*, 179-291.

[10] Appleton, J.D. and Adlam, K.A.M. Geogenic control on soil chemistry in urban areas : a novel method for urban geochemical mapping using parent material classified data. Appl. Geochem., **2012**, *27*, 161-170.

[11] Cave, M.R., Wragg, J., Palumbo, B. and Klinck, B.A., Measurement of the Bioaccessibility of Arsenic in UK soils. Environment Agency; 2003.

[12] Johnson, C.C., Breward, N., Ander, E.L. and Ault, L. G-BASE: Baseline geochemical mapping of Great Britain and Northern Ireland. Geochemistry: Exploration, Environment, Analysis, **2005**, *5*, 347-357.

[13] British Standards Institute, Determination of Particle Size Distribution-Seiving Method. In *BS 933-1*; 1997.

[14] Johnson, C.C., Understanding the Quality of Chemical Data from the Urban Environment – Part 1: Quality Control Procedures. In *Mapping the Chemical Environment of Urban Areas*; Johnson, C.C., Demetriades, A., Locutura, J. and Ottesen, R.T. Eds.; Wiley-Blackwell; Oxford; 2011; 61-76.

[15] Rowell, D.L. *Soil Science, Methods and Applications*; Longman Scientific & Technical, 1994.

[16] Oomen, A.G., Hack, A., Minekus, M., Zeijdner, E., Cornelis, C., Schoeters, G., Verstraete, W., Van de Wiele, T., Wragg, J., Rompelberg, C.J.M., Sips, A. and Van Wijnen, J.H. Comparison of five in vitro digestion models to study the bioaccessibility of soil contaminants. Environ. Sci. Technol., **2002**, *36*, 3326-3334.

[17] Wragg, J., Cave, M.R., Basta, N., Brandon, E., Casteel, S., Denys, S.e.b., Gron, C., Oomen, A., Reimer, K., Tack, K. and Van de Wiele, T. An Inter-laboratory Trial of the Unified BARGE Bioaccessibility Method for Arsenic, Cadmium and Lead in Soil. Sci. Total Environ., **2011**, *409*, 4016-4030.

[18] Denys, S., Caboche, J., Tack, K., Rychen, G., Wragg, J., Cave, M., Jondreville, C. and Feidt, C. In Vivo Validation of the Unified BARGE Method to Assess the Bioaccessibility of Arsenic, Antimony, Cadmium, and Lead in Soils. Environ. Sci. Technol., **2012**, *46*, 6252-6260.

[19] Broadway, A., Cave, M.R., Wragg, J., Fordyce, F.M., Bewley, R.J.F., Graham, M.C., Ngwenya, B.T. and Farmer, J.G. Determination of the bioaccessibility of chromium in Glasgow soil and the implications for human health risk assessment Sci. Total Environ., **2010**, *409*, 267-277.

[20] Caboche, J., Denys, S., Feidt, C., Delalain, P., Tack, K. and Rychen, G. Modelling Pb bioaccessibility in soils contaminated by mining and smelting activities. J Environ Sci Health., Part A, **2010**, *45*, 1264-1274.

[21] Pelfrene , A., Waterlot, C., Mazzuca, M., Nisse, C., Bidar, G. and Francis, D. Assessing Cd, Pb, Zn human bioaccessibility in smeltercontaminated agricultural topsoils (northern France). Environ. Geochem. Health, **2011**.

[22] Roussel, H., Waterlot, C., Pelfrene, A., Pruvot, C., Mazzuca, M. and Douay, F. Cd, Pb and Zn oral bioaccessibility of urban soils contaminated in the past by atmospheric emissions from two lead and zinc smelters. Arch Environ Contam Toxicol, **2010**, *58*, 945-54.

[23] Wragg, J., Cave, M., Taylor, H., Basta, N., Brandon, E., Casteel, S., Denys, S., Gron, C., Oomen, A., Reimer, K., Tack, K. and Van de Wiele, T., Interlaboratory Trial of a Unified Bioaccessibility Procedure. In *Open*; British Geological Survey; Nottingham; 2009.

[24] Watts, M.J., Button, M., Brewer, T.S.D., Jenkin, G.R.T. and Harrington, C.F. Quantitative arsenic speciation in two species of earthworms from a former mini site. J. Environ. Monit., **2008**, *10*, 753-759.

[25] Wragg, J., Cave, M.R., Taylor, H., Basta, N., Brandon, E., Casteel, S., Gron, C., Oomen, A. and Van de Wiele, T., Inter-laboratory Trial of a Unified Bioaccessibility Procedure. British Geological Survey; 2009.

[26] Cave, M.R., Taylor, H. and Wragg, J. Estimation of the Bioaccessible Arsenic Fraction in Soils using Near Infra Red Spectroscopy. J Environ Sci Health., Part A, **2007**, *42*, 1293-1301.

[27] R Development Core Team, R: A language and environment for statistical computing.; R Foundation for Statistical Computing; Vienna, Austria.; 2011.

[28] Draper, N. and Smith, H. *Applied Regession Analysis*; John Wiley, 1998.

[29] Schwarz, G.E. Estimating the dimension of a model. Annals of Statistics, **1978**, *6*, 461-464.

[30] Efron, B. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. Journal of the American Statistical Association, **1983**, *78*, 316-331.

[31] Deer, W.A., Howie, R.A. and Zussman, J. *An Introduction to the Rock Forming Minerals,* 2nd Ed.; Prentice Hall; New Jersey, 1992.

[32] Friendly, M. Corrgrams: Exploratory displays for correlation matrices. American Statistician, **2002**, *56*, 316-324.

[33] Cornell, R.M. and Schwertmann, U. *The Iron Oxides - Structure Proprties, Reactions, occurences and Uses*; VCH Publishers; Weinheim, 1996.

[34] Duggan, M.J., Inskip, M.J., Rundle, S.A. and Moorcroft, J.S. Lead in Playground Dust and on the Hands of Schoolchildren. Sci. Total Environ., **1985**, *44*, 65-79.

[35] Wragg, J., Cave, M. and Nathanail, P. A Study of the relationship between arsenic bioaccessibility and its solid-phase distribution in soils from Wellingborough, UK. J Environ Sci Health., Part A, **2007**, *42*, 1303 - 1315.

[36] Palumbo-Roe, B., Cave, M.R., Klinck, B.A., Wragg, J., Taylor, H., O'Donnell, K. and Shaw, R.A. Bioaccessibility of arsenic in soils developed over Jurassic ironstones in eastern England. Environ. Geochem. Health, **2005**, *27*, 121-130.

List of Figure Captions

Figure 1 Arsenic concentration (mg kg^{-1}) in England and Wales topsoils [3] divided into percentile ranges along with the location of Northampton in England.

Figure 2 Simplified geological map of Northampton

Figure 3 Probability Density functions of selected elements in the 4 clusters

Figure 4 Correlogram of the major elements, pH and As composition of the Northampton Soils

Figure 5 Relationship between As in the <2 mm and <250 μm size fractions

Figure 6 Cross Validation plot for Model 1 comparing the modelled and measured bioaccessible As showing the line of equivalence

Figure 7 Correlogram of major elements and NIR spectral components

Figure 8 Cross Validation plot for Model 2 comparing the modelled and measured bioaccessible As showing the line of equivalence

Figure 9 Predicted As Bioaccessibility in Northampton using Model 1

Figure 10 Predicted As Bioaccessibility in Northampton using Model 2

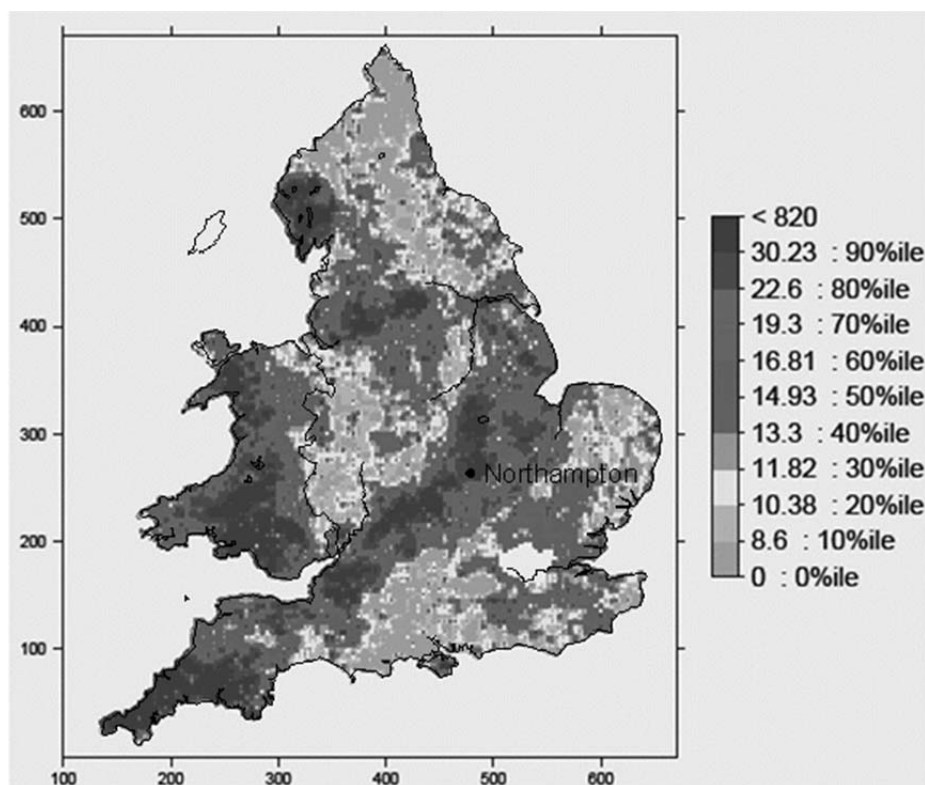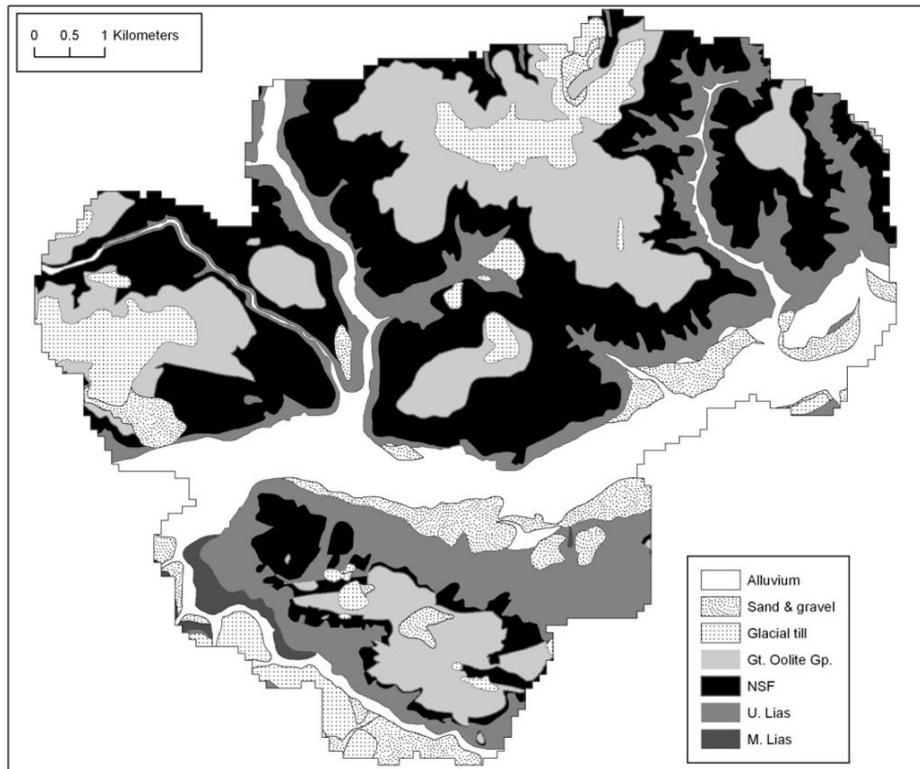Figure 11 Density Distribution of total As and bioaccessible As in Northampton
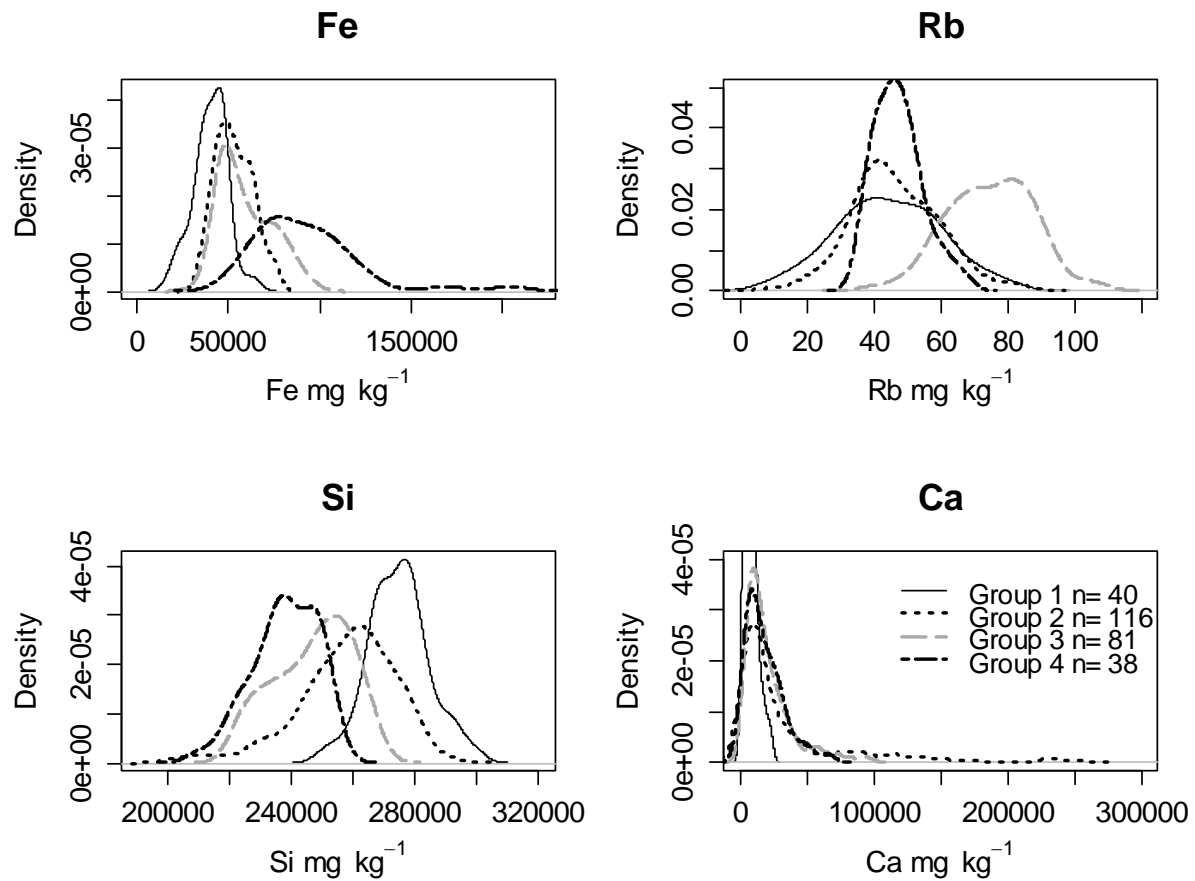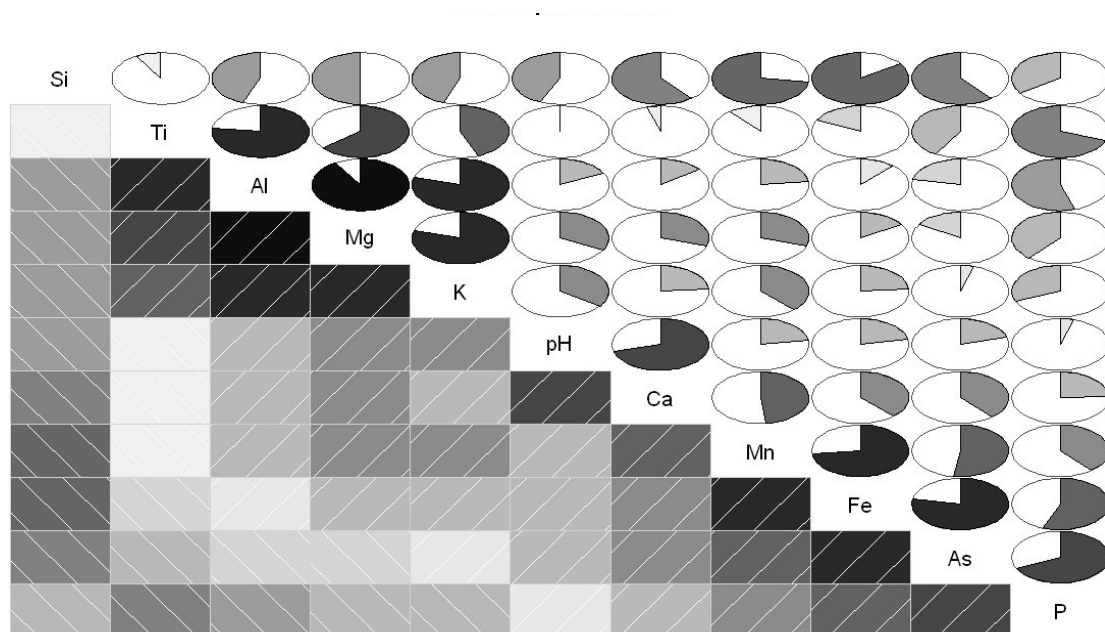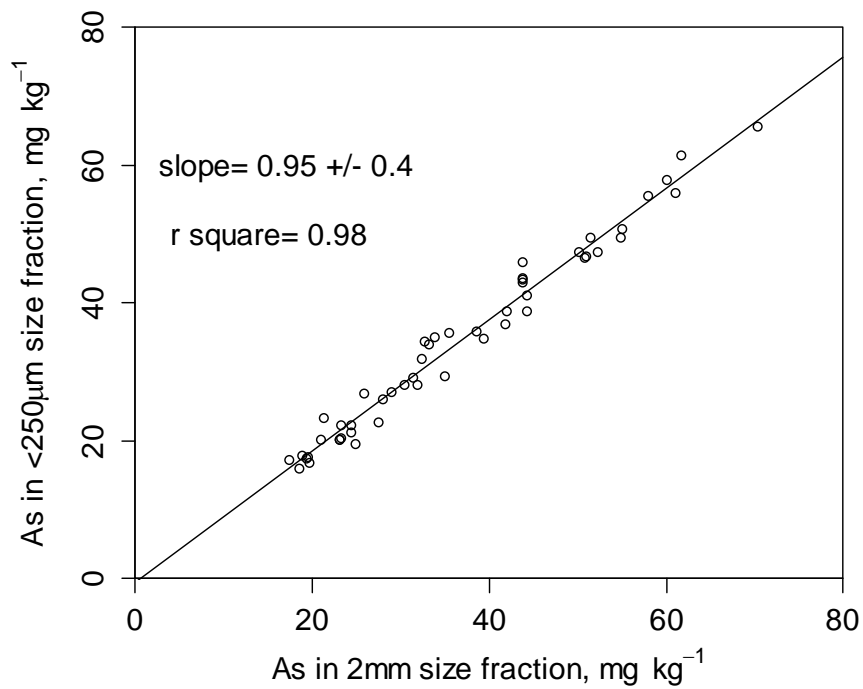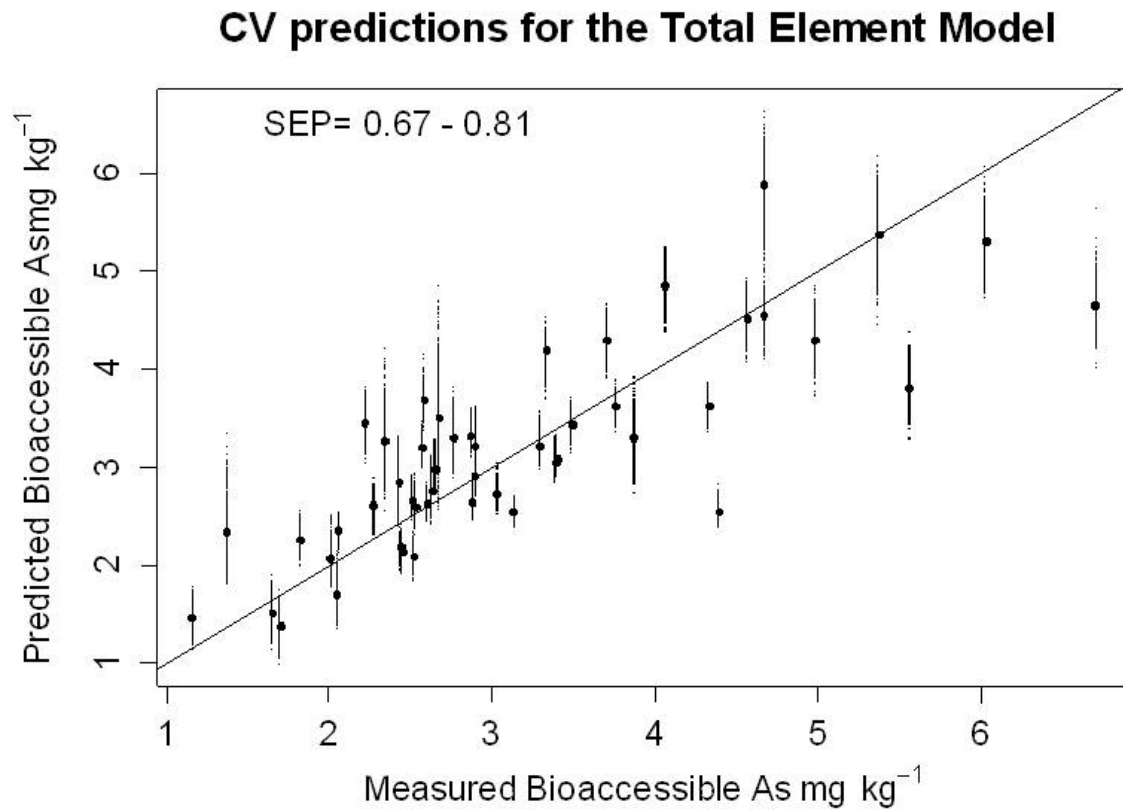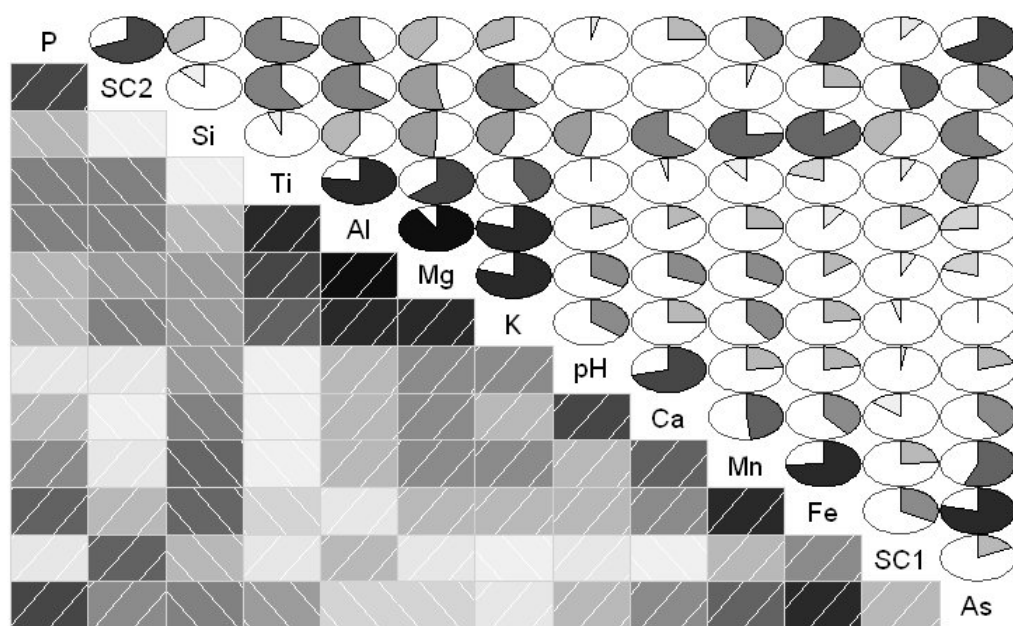
Figures

Figure 1

Figure 2

Figure 3

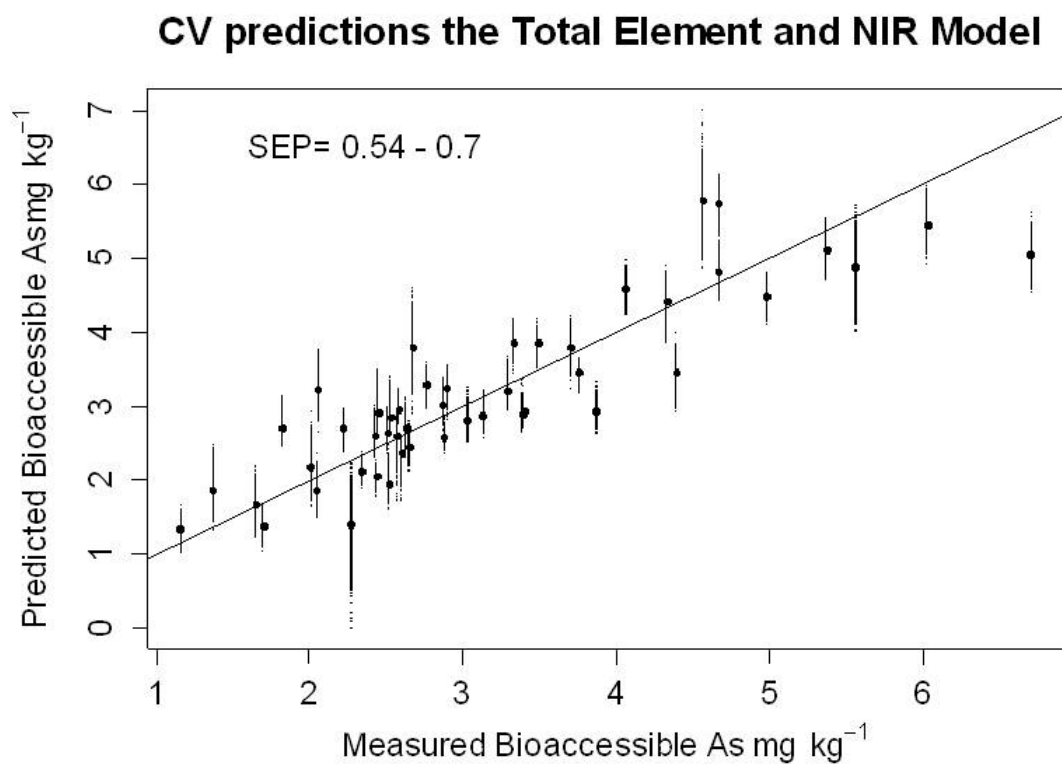Figure 4

Figure 5
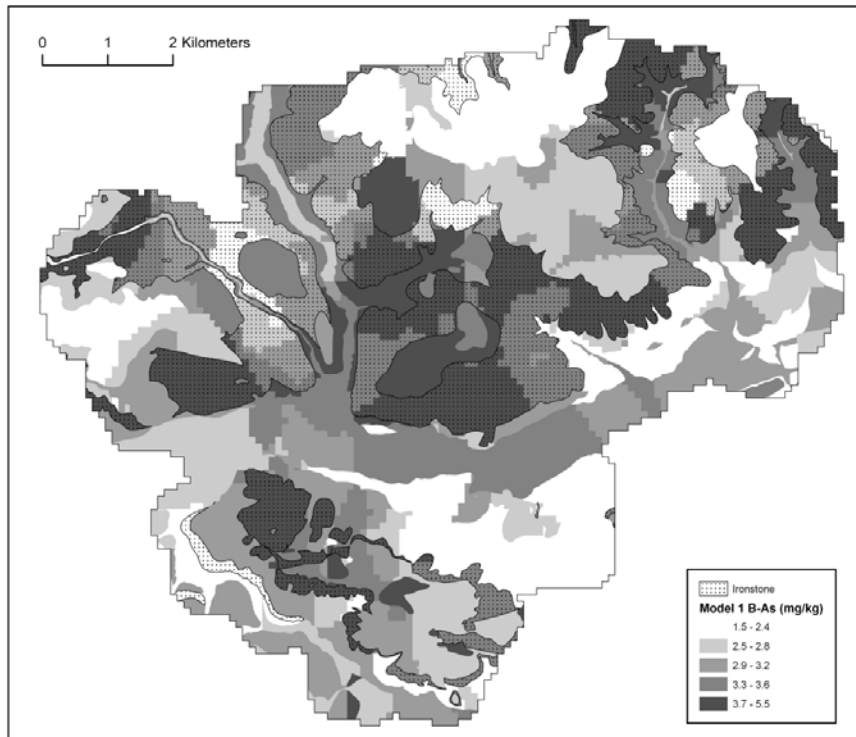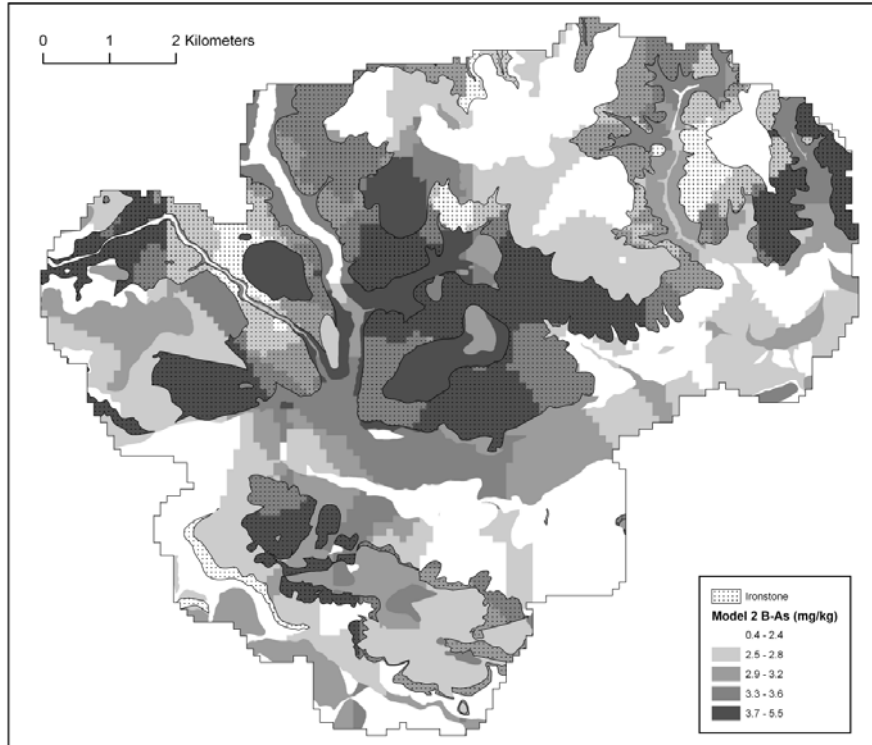

Figure 6

Figure 7

Figure 8
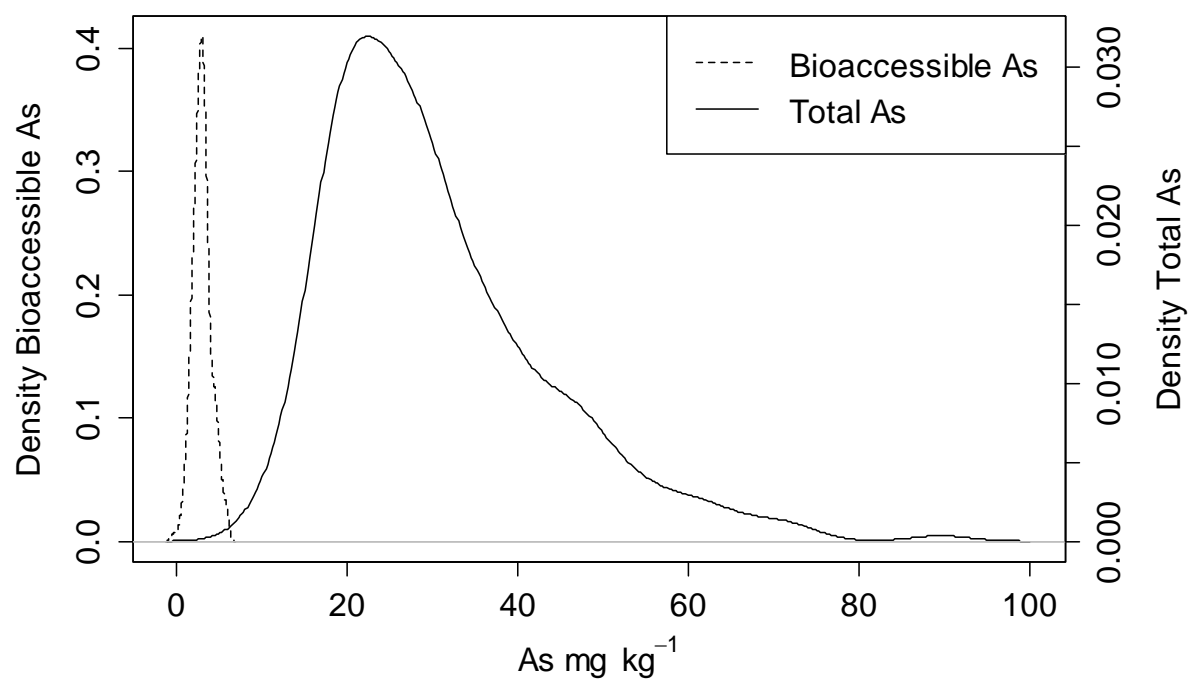
Figure 9

Figure 10

Figure 11

Tables

Table 1 Total and bioaccessible As concentrations for the 50 selected soils

Total As	Bioaccessible As	Total As	Bioaccessible As
31.9	2.6	43.8	5.6
23.1	2.5	51.5	6.7
51.0	3.3	33.8	2.0
27.9	3.3	61.7	4.7
24.3	2.4	31.3	3.1
55.0	11.0	20.9	1.7
23.3	2.3	21.3	2.7
43.8	2.2	35.4	3.9
54.8	3.7	38.6	2.4
41.8	2.7	19.4	1.4
39.3	3.8	19.3	1.8
44.3	5.0	24.4	3.0
23.3	2.9	17.4	1.2
28.9	3.4	30.4	2.9
52.2	2.9	43.8	4.3
60.0	4.1	61.0	5.4
57.9	4.7	70.4	6.0
33.1	2.3	19.6	1.7
18.8	2.5	23.0	2.1
43.8	2.6	50.8	4.6
44.3	3.5	32.3	4.4
42.0	2.9	24.9	2.5
35.0	2.6	25.8	2.1
27.4	3.4	18.5	2.5
50.1	2.8	32.7	2.6

Table 2 Summary statistics for MLR model 1

Coefficient	Value	Standard Error	P value	% Variance explained
Intercept	-3.35	0.88	<0.05	-
pH	0.65	0.14	<0.05	22.6
P	0.00050	0.00015	<0.05	40.4
As	0.036	0.0096	<0.05	8.9

R square = 0.72

Table 3 Summary statistics for MLR model 2

Coefficient	Value	Standard Error	P value	% Variance explained
Intercept	-0.56	1.22	<0.64	-
pH	0.41	0.12	<0.05	22.6
As	0.05	0.01	<0.05	37.6
Mg	0.00	0.00	<0.05	4.9
SC1	-21.1	3.6	<0.05	2.5
SC2	16.2	2.5	<0.05	16.5

R square = 0.84