Part 2A, Environmental Protection Act 1990

Technical Guidance Sheet (TGS) on normal levels of contaminants in English soils

Supplementary Information

ARSENIC (As)

Technical Guidance Sheet Supplementary Information TGS01s, July 2012.

Contents

defra
Department for Environment Food and Rural Affairs

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

List of Figures

List of Tables

Acknowledgments

This supplementary information for the arsenic Technical Guidance Sheet (TGS) is compiled with information derived mainly from the reports prepared for the Department for Environment Food and Rural Affairs (Defra) soil R&D project SP1008 by the British Geological Survey. This work has been led by Chris Johnson with assistance from Louise Ander, Mark Cave and Barbara Palumbo-Roe (all BGS, Keyworth) with additional contributions and comments from Murray Lark, Barry Rawlins, Don Appleton and Chris Vane (BGS Keyworth); Stephen Lofts (CEH Lancaster); and Paul Nathanial Land Quality Management Group, Nottingham. The authors also thank the Defra Soils Policy Team, the Project Steering Group and several Local Authority contaminated land officers who have given valuable advice to improve the content of this information sheet.

When referring to this document the following bibliographic reference should be made:

Defra, 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Arsenic – supplementary information. Technical Guidance Sheet No. TGS01s, July 2012. Department for Environment, Food and Rural Affairs (Defra), Soils R&D Project SP1008. Available on-line from Defra project SP1008 web page.

The arsenic Technical Guidance Sheet which this document supplements:

Defra, 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Arsenic. Technical Guidance Sheet No. TGS01, July 2012. Department for Environment, Food and Rural Affairs (Defra), Soils R&D Project SP1008. Available on-line from Defra project SP1008 web page.

Supplementary Information

Important soil sample and analytical information

Aggregate samples, soil depth and fraction size

Both the NSI (XRFS) and G-BASE data sets are derived from a soil sample that has been aggregated (composited) from a number of subsamples collected over the area of a site, rather than a single point sample. In the case of NSI this is 25 cores (subsamples) from a 20-m square (McGrath and Loveland 1992) whereas G-BASE is 5 cores, also from a 20-m square (Johnson *et al.* 2005; Fordyce *et al.* 2005). If a sample is collected as a single core, and the result is compared to the NBC, it is important to be aware that short-range variation (which can be substantial) for the single core sample will be potentially much greater than for the samples from which the NBC values are derived (Lark, 2012).

Soil samples used to calculate the As NBCs have been collected from the top 15 cm of the mineral soil profile (hence they are referred to as topsoils). When the sample is collected from a site covered with vegetation the surface organic layers (leaf litter) do not form part of the sample collected. Any recently deposited airborne particulates that have not yet migrated into the soil profile will not be sampled and surface organic material, which has the capacity to fix some contaminants from atmospheric deposition, is not included as part of the sample. In urban areas the top 15 cm will be expected to have been modified by historical urban land uses and, in rural agricultural areas, where relevant, will be within the ploughed horizon. Surveys targeting recent airborne pollution added to the soil will generally only collect from the top 2 cm of the profile in order to bias the soil results toward the airborne pollutant inputs. Such data has not been used in the NBC calculations.

Another consideration is the soil size fraction to be submitted for chemical analysis. The <2 mm fraction is widely used for soil analyses. However, other fractions are sometimes reported (*e.g.* <150 µm) in order to enhance some chemical contrasts and to reduce variability in the chemical results – coarser grains mean that a single "nugget" will give rise to greater variability in the analyses than will occur with a finer more homogenous material. The NBCs calculated here are exclusively based on the <2 mm soil fraction.

During the data exploration phase of this project the use of deeper soil samples (35 – 50 cm) was investigated and a comparison was made with topsoil samples collected from the same site (see Ander *et al.* 2011). For an area of central England approximately 13,000 sample sites have As determinations (by XRFS) for both top- (0 – 15 cm, <2 mm fraction) and deep soils (35 – 50 cm, <150 µm fraction) (see Appleton *et al.* 2008). These results have been plotted in Figure 1. Note that sampling from a fixed depth does not take into account any horizon development within the soil profile, the 0 – 15 cm and 35 – 50 cm depths are standard but for actual samples this may vary, for example, over a soil with a very shallow profile. There is ambiguity in the reporting of the surface point in a soil profile, for example, the G-BASE topsoil are cited as starting from a depth of 5 cm so as to take into account any organic litter, vegetation and root zone. As stated earlier, topsoils represent the top 15 cm of the mineral soil profile.

In Figure 1 the top- and deep soil results are generally similar as shown by the regression:

$$[\mathrm{As}_{\mathrm{topsoil}}] = 0.396 + (0.976 \times [\mathrm{As}_{\mathrm{deep\ soil}}]) \quad (n = 12{,}943;\ R^2 = 91\%,\ P<0.05)$$ ……*Equation 1*

However, when the data ratios (topsoil/deep soil) are compared, the data is negatively skewed implying that outliers from deep soil concentrations are greater than those found in the topsoil. These outliers are seen to have a systematic distribution that is closely related to the outcrop of specific geological formations. Because of the differing soil fraction analysed, and the fact that a very large number of topsoils were available for NBC determination, G-BASE deep soils have not been included in the calculations for NBCs. The good linear relationship shown above has meant that deep soils have been used as a predictor for As in the surface environment (Appleton *et al.* 2008).

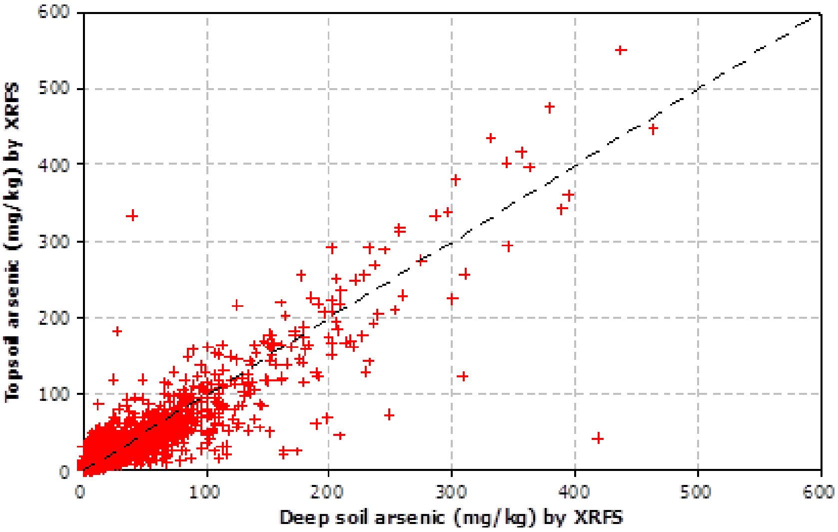


Figure 1: Comparison of deep and topsoil arsenic concentrations from the G-BASE project.

Total and partial analytical determinations of element concentrations

There are established international procedures and standards for the determination of naturally occurring elements in the Earth's surface environment (Darnley *et al.* 1995). These procedures have been set up in order to develop a global database of chemical results that is compatible and of sufficient quality to be used for environmental and resource management. The analytical requirements to realise this objective includes: "*The total amount of each element present is the most fundamental (and reproducible) quantity in any sample, therefore direct measurement techniques, e.g. XRFS or neutron activation analysis (NAA), or total extraction procedures should be employed as a first priority*." The British Geological Survey has been one of the leading organisations in the development of this global geochemical database. Therefore, the vast majority of systematically collected soil sample data that are available for NBC calculations for English soils are total element concentrations determined by laboratory-based XRFS. Other analytical techniques that do not give total element concentrations are used to determine the nature of occurrence and speciation of an element within a sample. Methods to determine the bioaccessible fraction of As (Wragg *et al.* 2011), for example, will provide essential information for health risk assessments.

When using NBCs a common question will be "how should I interpret NBCs in the context of non-total analyses"? This was investigated as part of the data exploration phase of this project (Ander *et al.* 2011). Figure 2 shows a plot of As in topsoils collected by G-BASE (as part of the Tellus Project, Northern Ireland, 6,872 samples) which were analysed by both a total (XRFS) and partial (*aqua regia* followed by ICP-MS) method. There is a close linear relationship between the two methods with a systematic bias to higher concentrations by XRFS; this would be expected from this total measurement, unlike the acid digest which will leave a quantity of trace element bearing, residual material. There would also be an inherent, systematic bias expected between any two analytical measurement techniques.

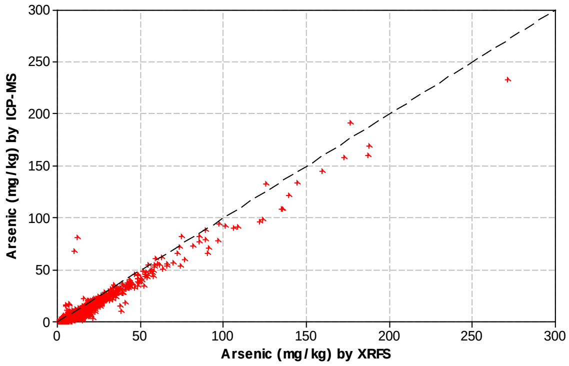


Figure 2: A comparison of topsoil As concentrations in Northern Ireland (Tellus Project) by XRFS and *aqua regia* digest ICP-MS

The regression equation for these data is:

$[As_{XRFS}] = 3.75 + (1.06 \times [As_{aqua\ regia}])$ ($n=6872$; $R^2 = 94\%$; and $P<0.05$)……*Equation 2*

A similar exercise has been done with the recently collected GEMAS project samples (see Ander *et al.* 2011). The 130 samples from England were analysed by both XRFS and *aqua regia* digest followed by ICP-MS analysis (Figure 3). Again there is a generally good linear relationship with lower concentrations recovered from the acid digest/ICP-MS than for the XRFS analysis. The regression equation is:

$[As_{XRFS}] = 2.58 + (1.13 \times [As_{aqua\ regia}])$ ($n=130$; $R^2 = 89\%$; and $P<0.05$) ……*Equation 3*

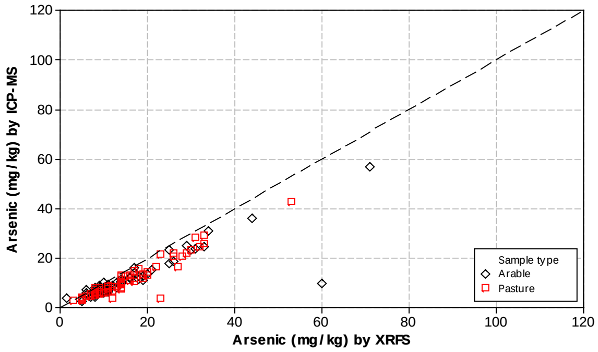
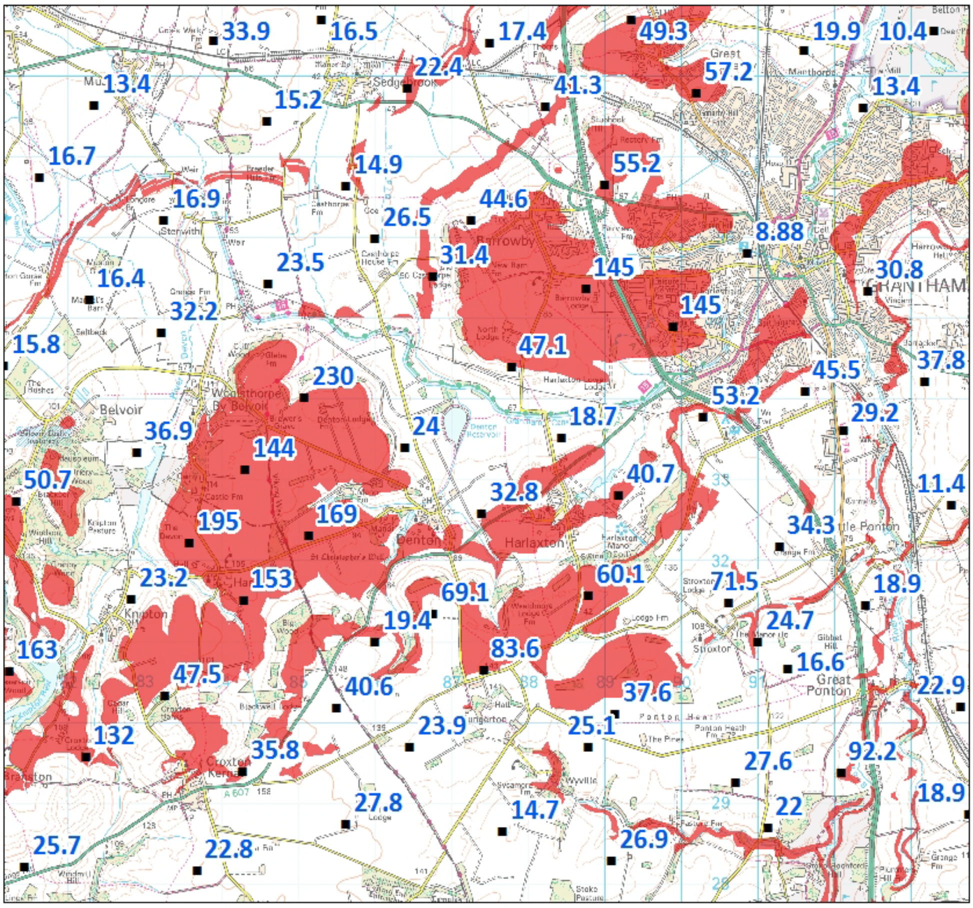


Figure 3: Comparison of GEMAS project topsoil As concentrations by XRFS and *aqua regia* digest ICP-MS in England

Using equations 1 and 2, a soil with 20 mg/kg As determined by *aqua regia* followed by ICP-MS could be estimated to have a total concentration of 25 mg/kg (rounded to nearest 1 mg/kg) for both equations. The regression equations are therefore a tool to estimate total concentration for *aqua regia*/ICP-MS determined samples, though its application must be done with awareness of the analytical error range, particularly at higher concentrations.

Scale and use of Normal Background Concentrations

NBCs have been determined for As using soils collected at a range of sampling densities, from 1 sample per 0.25 km^2 (G-BASE urban) through to 1 sample per 25 km^2 (NSI XRFS). The G-BASE urban samples provide a definition of the chemical surface environment to a much higher resolution than do the NSI (XRFS) samples. Thus G-BASE rural samples (collected at 1 per 2 km^2 sampling density) can show contaminant variability at a local area scale (1:50,000), as is shown by Figure 4. When investigating a sample result in the context of a NBC, it is important to ask whether localised variability (scales at less than 1:50,000), say within the Principal Domain, has been truly captured during the determination of the NBCs. Ander *et al.* 2011 describe that at a local scale both small coal seams and mineralised veins may not be mapped. Therefore, a high contaminant result should be attributed to a domain taking account of the localised underlying parent material features, even where this has an extent which is very discontinuous (e.g. Figure 4).



BGS © NERC. Contains Ordnance Survey data © Crown Copyright

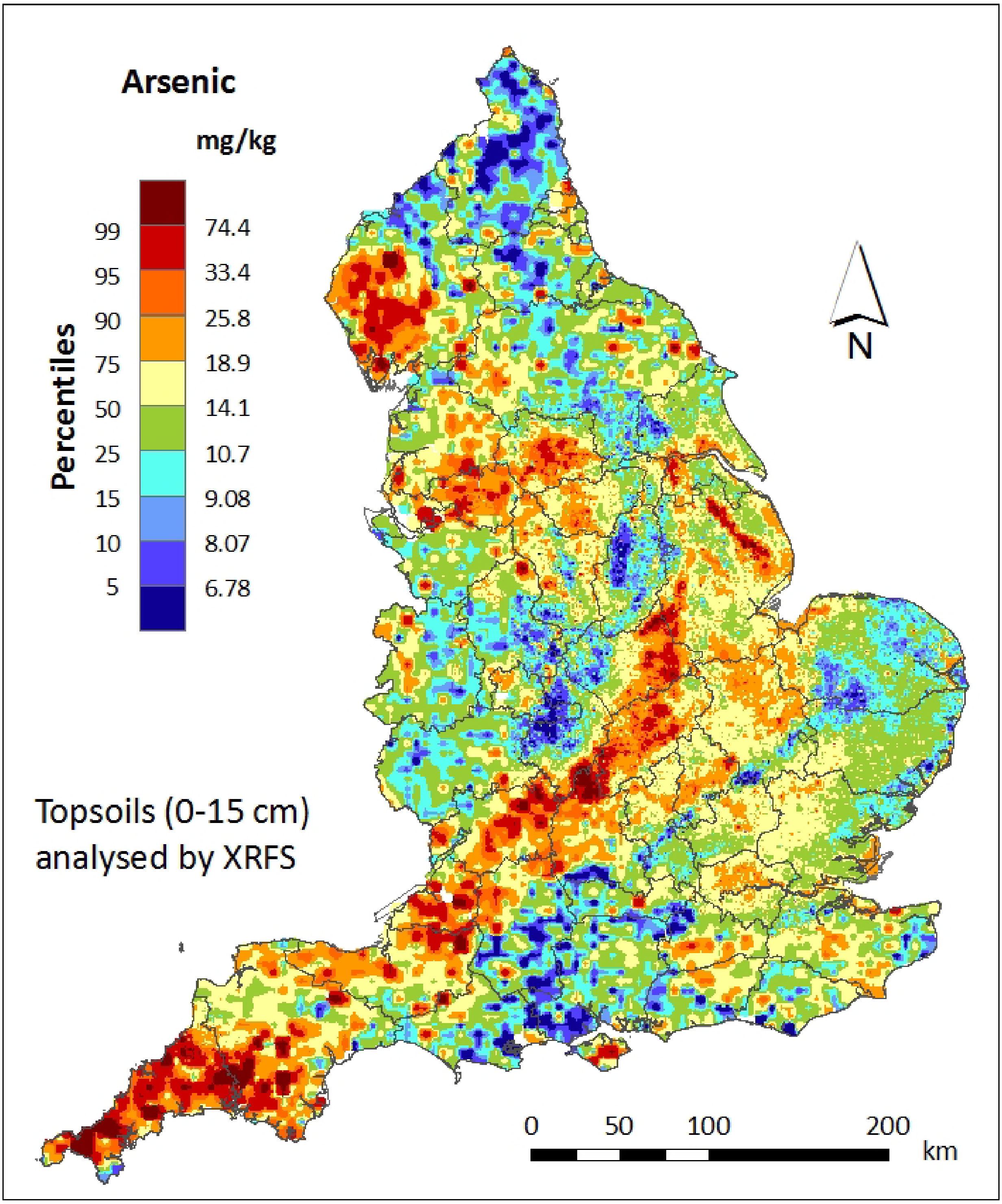
Figure 4: Example of topsoil As concentrations over and around ironstones (Grantham, Lincolnshire). Data in blue are As concentrations in mg/kg for sampling sites shown as black squares, ironstone outcrop area in red. The blue grid lines represent the 1 km OS grid squares.

Supplementary Information

Use of variograms

The domains that are defined for a particular contaminant correspond to major sources of variation in concentrations of that contaminant in soil, such as geological factors, urbanisation, mining or mineralisation. Concentrations of the contaminant vary within the domains, the procedure to define normal background concentrations (NBCs) quantifies this variation with robust statistics, from which the NBCs are computed. The spatial variation of a contaminant within a domain can be quantified by the variogram (Matheron, 1962). The variogram is a function that shows how the variation between observations of a variable at two sites depends on the distance in space between the sites. The variogram is half the mean squared difference between two observations plotted against the distance between them for all the results in a data set. Typically the variogram increases with distance until a plateau in the plot is reached at a value called the sill variance, which it reaches at a distance called the range. If the range is very short then this shows that the spatial variation is very intricate. If the range is longer then it may be feasible to map spatial variations from sample observations on a grid.

National map showing the distribution of arsenic in topsoils



Distribution of samples used in this interpolated map is shown in Figure 1 of the As technical guidance sheet.

Figure 5: National map of arsenic distribution in topsoils with county boundaries (using G-BASE and NSI (XRFS) results).

Supplementary Information

The national map of As distribution in topsoils (Figure 5) is shown here along with county boundaries to help with location at a regional scale. This map is given to demonstrate the variability in As across England and is also available to view on-line at the BGS project web page. The map has been generated from G-BASE and NSI (XRFS) topsoil data using 42,133 samples. Because central and eastern England have been sampled at a much higher density (by G-BASE), resolution of information in these areas is much higher. Figure 5 has been produced in ArcGIS v9.3 using the IDW option of the Spatial Analyst tool, cell size 1000 m and search radius 5000 m (inverse square option selected). The percentile classification is based on **all data** and differs from the domain data sets in which results are modelled to fit a normal distribution and the effect of outliers (representing point rather than diffuse pollution) have been reduced by normalisation of the data.

The map shown in Figure 5 uses soils to represent the geochemical baseline. Other national/regional scale geochemical atlases for soils are those of McGrath and Loveland (1992) (NSI *aqua regia* data) and Rawlins *et al.* (2012) (NSI XRFS data). On a continental scale, soil contaminant maps are available from the EuroGeoSurveys FOREGS and GEMAS projects (Salminen *et al.* 2005; Reimann *et al.* 2012). A preferred way of representing the geochemical baseline at a national/regional scale is to use stream sediments. The fine stream sediment in a drainage channel is representative of material washed down the drainage catchment to the sampling site in the stream and so gives a much better regional average of the chemical environment than is given by soils. The G-BASE project also collects stream sediments at a sampling density of approximately one sample per 2 km^2 and results for England have been presented in a series of atlases (*e.g.* Lake District (BGS 1992) and NE England (BGS 1996)) and these can be used to further demonstrate element variability across the surface environment of England. For the more recently sampled parts of England, the G-BASE project has also determined a large range of elements in stream waters (*e.g.* Environmental Geochemical Atlas of Central and Eastern England). Comparing the element concentrations and distributions of different sample types collected from the same locality can provide useful information about the mobility of a chemical element in that area.

A stream sediment atlas for England and Wales was also completed by Webb *et al.* (1978) (Wolfson Geochemical Atlas) and more recently low density sampling has produced continental scale geochemical baselines for Europe based on a number of sampling media, including stream sediments, stream waters and soils (Salminen *et al.* 2005) and the ongoing GEMAS project (Reimann *et al.* 2012). Appleton *et al.* (2008) using the BGS soil and stream sediment results and the Wolfson data, have produced national-scale estimation of potentially harmful element ambient background concentrations in topsoil using a parent material classified soil:stream-sediment relationships. These geochemical data have been used to derive geospatial data sets that identify the estimated concentrations (mg/kg) of As, Cd, Cr, Ni and Pb in soils.

Descriptive statistics for arsenic in topsoil data

Arsenic Domain percentile classifications

Arsenic data for soils has been gathered from data sets as described in the As TGS and classified according to the most important domains as detailed by Ander *et al.* (2011). A percentile of a data distribution (in this case the distribution of As in soil for a given domain) is the value of a variable below which a certain percentage of observations fall. The 95th percentile, for example, is the value below which 95% of the observations may be found, *i.e.* it encompasses the majority of the data. The contaminant concentrations in the soil for a given domain are a subset of the total population of all possible soil concentrations and therefore any percentile calculation will only be an approximation of the true value. The uncertainty on the percentile increases as the number of samples used to calculate it decreases. Lower and upper limits can be statistically estimated for each percentile giving a confidence interval for that percentile. **The As NBC for each domain is defined as the upper 95% confidence limit of the 95th percentile for the As topsoil concentrations that fall within that domain** (Cave *et al.* 2012). A summary of domain percentiles with their upper and lower limits is given in Table 1.

Percentile	**Ironstone Domain (437)**			**Mineralisation Domain (187)**			**Principal Domain (41,509)**		
	lower	middle	upper	lower	middle	upper	lower	middle	upper
50	46	50	54	39	46	54	14	14	14
55	51	55	60	43	51	61	15	15	15
60	57	62	67	47	56	70	16	16	16
65	63	69	74	51	63	80	17	17	17
70	71	77	84	56	71	91	18	18	18
75	80	88	96	62	80	110	20	20	20
80	91	100	110	69	92	130	21	21	22
85	110	120	130	78	110	150	23	24	24
90	130	150	160	92	130	200	26	27	27
95	170	200	**220**	120	180	**290**	31	32	**32**

Figure in brackets represents the number of samples used in the domain calculation

Table 1: A summary of the arsenic domain percentile classifications. Domain NBCs shown in bold red, concentrations in mg/kg.

Descriptive statistics arsenic topsoil data set

Table 2 shows descriptive statistics for all the topsoil As results from the G-BASE and NSI (XRFS) data sets. The cities and towns in Table 2(c) are those that have been systematically sampled by the G-BASE project – these are shown in Figure 1 of the As TGS. Some of these data sets have associated reports that can be downloaded by clicking on the location place marker on the map at http://www.bgs.ac.uk/gbase/urban.html. Other data sets for other English cities may exist but they are not made publicly available and are not sampled and analysed to a nationally consistent standard.

(a) All data	**Number**	**Mean**	**Minimum**	**25th percentile**	**Median**	**75th percentile**	**Maximum**	**Skewness**
G-BASE(urban+rural) + NSI (XRFS)	42133	17.6	<0.5	10.7	14.1	18.9	15100	186
(b) Data set type	**Number**	**Mean**	**Minimum**	**25th percentile**	**Median**	**75th percentile**	**Maximum**	**Skewness**
All NSI(XRFS)	4864	18.5	<0.5	10.8	14.4	19.9	536	10
G-BASE (rural)	23686	17.5	<0.5	10.1	13.4	18.1	15100	143
G-BASE (urban)	13583	17.5	1.2	12.1	15.1	19.5	1010	29
Eastern England (G-BASE)	23222	16.2	<0.5	9.99	13.2	17.9	555	12
Tamar catchment (G-BASE)	464	80.4	6.9	16.5	22.3	42.3	15100	21
(c) Urban (G-BASE)	**Number**	**Mean**	**Minimum**	**25th percentile**	**Median**	**75th percentile**	**Maximum**	**Skewness**
Corby	133	23.4	10.7	16.9	19.5	23.9	90	3
Coventry	390	9.86	2.03	7.07	9.08	11.1	105	9
Derby	275	15.8	5.66	11.3	13.5	16.7	63.5	3
Doncaster	279	15.3	2.03	10.1	13.1	17.1	74.6	3
Hull	407	24.1	3.04	15.1	20.2	26.2	207	6
Leicester	652	14.1	4.25	9.89	13.2	17.3	84.2	3
Lincoln	215	15.3	4.05	8.08	11.1	21.2	65.5	2
London (GLA area)	6494	17.1	1.2	12.9	15.5	18.9	161	5
Manchester (part of)	300	28.1	2.53	15.8	20.3	28.4	1010	16
Mansfield	257	13.9	3.04	7.07	11.1	16.1	71.5	3
Northampton	275	34.3	8.48	22.8	30.4	42.6	107	1
Nottingham	636	14.4	5.05	11.1	13.1	16.1	87.7	4
Peterborough	272	18	7.47	14.5	17.2	21.1	34.8	1
Scunthorpe	196	26.4	3.04	11.1	19.2	31.2	191	3
Sheffield	575	25.8	4.05	17.1	22.2	29.2	241	5
South Essex towns	715	14.5	4.62	11.4	13.4	16.2	82.3	4
Stoke-on-Trent	745	16.3	2.03	11.1	14.1	18.1	137	5
Telford	292	12	5.05	8.08	10.1	13.9	54.4	3
Wolverhampton	284	19.8	3.54	13.4	16.6	22.5	158	3
York	191	11.6	3.04	8.08	10.1	13.1	93.7	7

Table 2: Descriptive statistics of underlying primary data sets for As in all topsoils. These are classified by the various data set subgroups of the original projects (total concentrations (XRFS) in mg/kg) (based on Ander *et al*. 2011, with results cited to 3 significant figures).

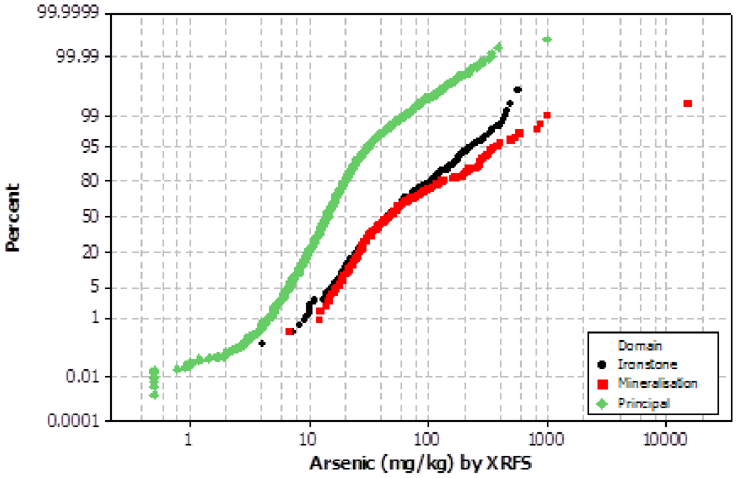


Figure 6: Cumulative probability plot of topsoil As results categorised by domain.

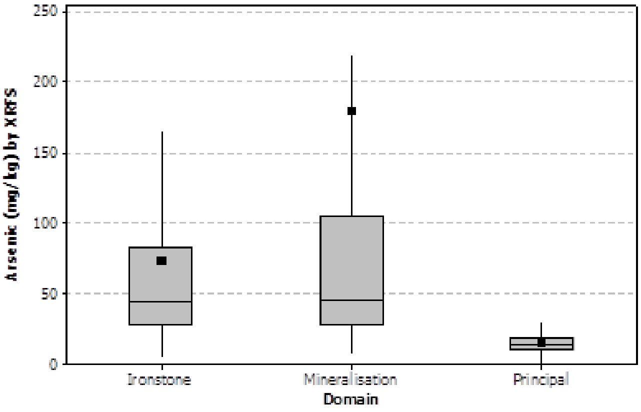


Figure 7: Boxplot of as topsoil results attributed to domains.

Figure 6 and Figure 7 show the frequency distribution of results for soils over the three domains defined for As. These plots can be used in conjunction with any new results plotted in a similar way to compare distributions within the defined domains. The box of the boxplot represents the interquartile range (Q1, Q3), with the median (Q2) as a line within the box. The point symbol shows the mean value. The upper whisker = Q3 + 1.5(Q3-Q1); lower whisker = Q1-1.5(Q3-Q1).

Archer and Hodgson (1987) carried out a study of total and extractable trace element contents of agricultural soils (from a depth of 15 cm) in England and Wales, including As. “Total” As analyses were done by AAS following a digestion using perchloric and nitric acids. They defined the normal range for trace element contents as that between twice the log-derived standard deviation above and below the mean; approximately 95% of the data range. For 222 agricultural topsoils they determined a As median of 10.4 mg/kg and a “normal” range of 2.3-53 mg/kg.

Supplementary Information

Landscape data used to define contaminant domains

Rather than seeking to define a single arsenic NBC for the whole of England, the project has, through its data exploration (Ander *et al.* 2011), determined the most significant domains that can be defined in order to capture the most significant controls on areas of high As concentration in soils. For As these have been identified as soils developed on an ironstone parent material and some mineralised/metalliferous mining areas. These domains have been defined using some key datasets within a GIS environment, namely: the BGS Soil-Parent Material Model (SPMM) (Lawley, 2009) and a revised and digitally updated version of the Ove Arup (1990) Department of the Environment (DoE) Metalliferous Mining and Mineralisation data set.

Soil parent material

The Soil-Parent Material Model[1] (SPMM) has been developed by BGS, using as its basis the mapped boundaries of the national 1:50,000 superficial and bedrock geological data (DigMapGB-50[2]), and is used within a GIS environment. Soil 'Parent Material' is the first recognisably geological material found beneath a soil profile, and is the lithology on which that soil has developed. Soils thus inherit many properties, including chemical composition, from this material.

In the SPMM the geological data have been combined into one layer of information which indicates the rock/sediment formation mapped as directly underlying soil. Where this is a superficial deposit (such as alluvium, glacial deposits, peat), the data set also maintains the record of the solid geological formation first encountered beneath this surface sediment; such information is of benefit where the underlying solid geology imparts chemical (or other) characteristics into the overlying superficial deposits, and thus the soil. The information, which has historically routinely been attributed to the mapped digital polygons in DigMapGB, largely comprises lithological and chronological information. Augmenting this in the SPMM is additional information on texture, mineralogy and lithology, which is attributed in a hierarchical classification system. In the context of the present study this means that a higher level of aggregated characteristics can easily be applied to soil geochemical data than is possible solely using DigMapGB; for instance, retrieving all formations which are classed as 'ironstones' (irrespective of their formal name) and confers benefits from using the SPMM.

The scale of mapping for the soil parent material is also relevant – 1:50,000 is the scale at which much of the systematic geochemical soil sampling has been undertaken, and gives the user a reasonable feel for the degree of uncertainty on the data. Where geographical information is provided at other common scales, such as 1:250,000 or 1:625,000, the boundaries and number of polygons are simplified and aggregated in order to provide generalised information at the national-scale. More detailed mapping, such as 1:10,000, is not available in a consistent format or as part of the SPMM data, and would imply greater certainty in sample locations and polygon boundaries than is appropriate from the data. Soil series mapping is available at a national-scale (see e.g. NSRI NATMAP[3]) but this is not systematically mapped at 1:50,000 and would require attribution with the latest geological mapping data in order to retrieve information on key formations, and so has not been used in this study.

---

[1] http://www.bgs.ac.uk/products/onshore/soilPMM.html

[2] http://www.bgs.ac.uk/products/digitalmaps/digmapgb_50.html

[3] http://www.landis.org.uk/data/natmap.cfm

Metalliferous mining and mineralisation

The data set which has been examined in this project is that of non-ferrous Metalliferous Mineralisation and Mining database, originally produced in hard-copy by Ove Arup (1990) for DoE (Department of Environment), but which has been 'cleaned' and turned into a polygon layer by BGS. The data for England has been further attributed for this project by giving a name to the major ore fields allowing soil sample sites and geochemical data to be joined to the ore fields and separately characterised for typical soil concentrations. This mapping is generalised to 0.5 km grid squares, which is a suitable level of spatial resolution for this type of data. Therefore, it should be expected that not every occurrence of mineralisation/mining has been captured within this GIS layer. Where soil chemical data is encountered that is located outside a given mineralisation domain, but of a concentration expected for that contaminant within the local mineralisation domain, and lies over the parent material which is known to be affected by mineralisation in that ore field, then that high soil concentration could relate to natural processes, or historical mining.

Supplementary Information

Summary of statistical procedure to determine NBCs

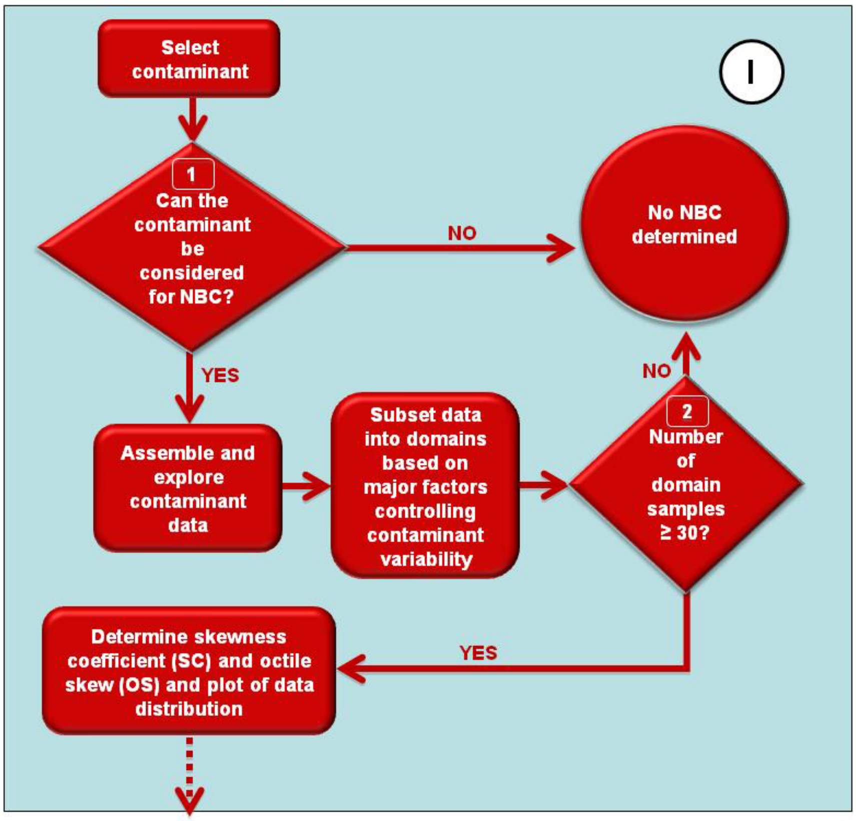


Figure 8: Flow chart for the calculation of the NBC for a given contaminant domain (OS and SC are octile skew and skewness coefficient, respectively. MAD = median absolute deviation). See text for explanation, continued overleaf.

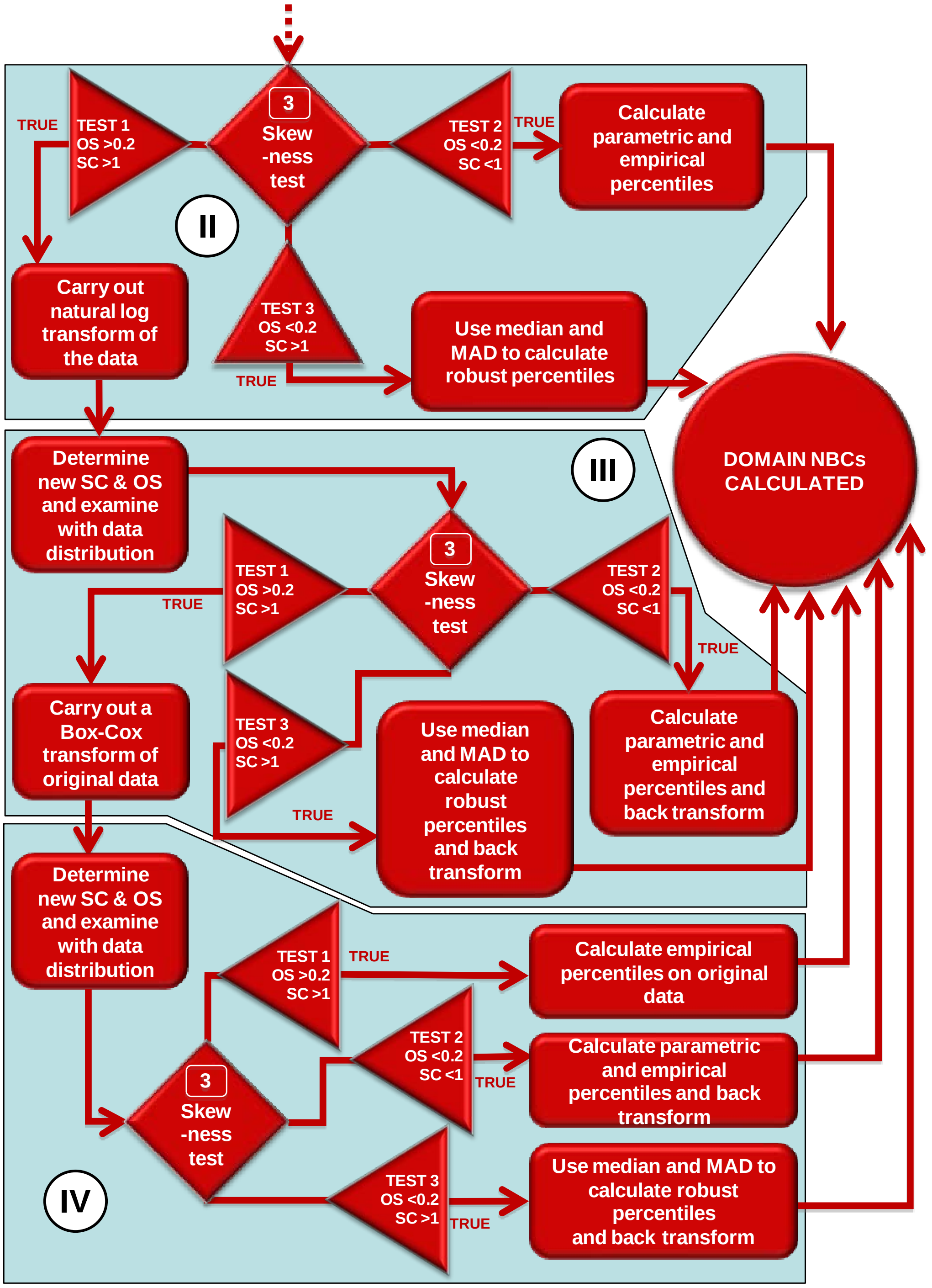


Figure 8 continued. Flow chart for the calculation of the NBC for a given contaminant domain (OS and SC are octile skew and skewness coefficient, respectively. MAD = median absolute deviation). See text for explanation.

Figure 8 summarises the statistical procedure used to determine contaminant NBCs (see Cave *et al.* 2012). Part I essentially represents the data gathering and exploration phase of the project (WP1&2) in which domain areas are identified. Question 1 asks if the contaminant is suitable for a NBC. Asbestos and manufactured organic contaminants with no natural origin, for example, fail this question. The data exploration (Ander *et al.*, 2011) identifies the areas (domains) where there are clearly identifiable controls on high concentrations of a specified contaminant. The contaminant data set is then subdivided into domain data sets. In Question 2 (Figure 8), a minimum of 30 results are considered necessary to determine a NBC (see Cave *et al*., 2012). Once the data has been subsetted into domains, then skewness testing and inspection of frequency distribution plots can be done to select the appropriate data transform and method of calculating percentiles (Parts II – IV). Question 3, the skewness test, has three possible outcomes. TEST 1 (OS > 0.2 and SC >1) is true if the data distribution is skewed and not suitable for fitting to a Gaussian model and the data need to be transformed to using either a logarithmic or Box-Cox transform. If TEST 2 (OS < 0.2 and SC <1) is true then the data are consistent with the assumption of a Gaussian distribution and the parametric percentiles are fitted based on the mean and standard deviation of the data. Finally, TEST 3 (OS < 0.2 and SC >1) means the data show a mostly symmetrical distribution but with potential outliers. Here the data are consistent with the assumption of a Gaussian distribution and the parametric percentiles are fitted using median and the median absolute deviation (MAD) in place of the mean and standard deviation as these measures are robust to outliers.

Access to data and information resources used to calculate NBCs

Project Reports and information

These resources are available from the BGS project web page [4] and include:

Data Exploration Reports (BGS report No. CR/11/145 and CR/012/041); Methodology Report (BGS report No. CR/12/003); Final Project Report (BGS report No. CR/12/035); Technical Guidance Sheets and supplementary information; MS Access Database summary of available data; Project Bibliography (Endnote bibliography); R code scripts used to determine NBCs; and GIS Resources served as WMS files (Domain polygons; the urbanisation index polygons defined from GLUD database; and the national contaminant interpolated image maps).

Web map services (WMS) are an industry standard protocol for serving georeferenced images across the web. They were developed and first published by the Open Geospatial Consortium (OGC) in 2000. Since this date WMS have had a steady uptake and are being increasingly used in traditional desktop based GIS, web-based GIS systems (including Google Earth), and the latest Smartphone 'apps'. BGS holds the data on their servers and publish it openly via the BGS project web page.

Principal contaminant data sets for England

Intellectual Property Rights for the raw soil data sets resides with the organisations responsible for those data sets. In the case of the G-BASE and NSI (XRFS) data is made freely available subject to certain licensing terms and conditions. For large data sets there will also be a data handling fee. Further information regarding access to the G-BASE and NSI (XRFS) soil data is given at the BGS project web page and enquiries should be sent to enquiries@bgs.ac.uk .

Other data sets providing information on soil chemistry are summarised in Appendix 2 of Ander *et al.* (2011) and this includes contact and web site links.

Soil parent material

The BGS Soil-Parent Material Model is described on a BGS web page (SPPM) [5] and this contains information regarding further information and pricing.

Land use data including metalliferous mining and mineralisation

The Generalised Land Use Database (GLUD) Statistics for England 2005 is available for free from the Communities and Local Government website. [6] Users interested in the detailed maps at land parcel level who hold the appropriate public sector licence to use OS MasterMap® can request to see the GLUD data at this large scale level (gis@communities.gsi.gov.uk).

The Ove Arup Mineralisation and mines data updated and modified by BGS is available from BGS subject to terms and conditions (see the BGS project web page).

---

[4] http://www.bgs.ac.uk/gbase/NBCDefraProject.html
[5] http://www.bgs.ac.uk/products/onshore/soilPMM.html
[6] http://www.communities.gov.uk/publications/planningandbuilding/generalisedlanduse

Supplementary Information

Further Reading

The following is a list of bibliographic references that provide more detailed information regarding the distribution and behaviour of arsenic in the surface environment. Some of these references are referred to in this supplementary information section.

Ander, E.L., Cave, M.R., Johnson, C.C. and Palumbo-Roe, B. 2011. Normal background concentrations of contaminants in the soils of England. Available data and data exploration. *British Geological Survey Commissioned Report*, CR/11/145. 124pp.

Ander, E.L., Cave, M.R., Johnson, C.C. and Palumbo-Roe, B. 2012. Normal background concentrations of contaminants in the soils of England. Results of the data exploration for Cu, Ni, Cd and Hg. *British Geological Survey Commissioned Report*, CR/12/041. 88pp.

Abrahams, P.W. and Thornton, I. 1987. Distribution and extent of land contamination by arsenic and associated metals in mining regions of southwestern England. *Transactions of the Institution of Mining and Metallurgy Section B-Applied Earth Science*, 96, B1-B8.

Appleton, J.D., Rawlins, B.G. and Thornton, I. 2008. National-scale estimation of potentially harmful element ambient background concentrations in topsoil using parent material classified soil:stream-sediment relationships. *Applied Geochemistry*, 23(9), 2596-2611.

Archer, F.C. and Hodgson, I.H. 1987. Total and extractable trace element contents of soils in England and Wales. *Journal of Soil Science*, 38, 421-431.

Bevins, R.E., Young, B., Mason, J.S., Manning, D.A.C. and Symes, R.F. (2010), *Mineralization of England and Wales*, Geological Conservation Review Series, No. 36, Joint Nature Conservation Committee, Peterborough, 598 pages, illustrations, A4 hardback, ISBN 978 1 86107 566 6. Available on-line at: http://www.jncc.gov.uk/page-3298 last accessed 1st May 2012.

BGS. 1992. Regional geochemistry of the Lake District and adjacent areas. British Geological Survey Regional Geochemistry Atlas Series. BGS, Keyworth, Nottingham, UK. ISBN 0 85272 225 7.

BGS. 1996. Regional geochemistry of north-east England. British Geological Survey Regional Geochemistry Atlas Series. BGS, Keyworth, Nottingham, UK. ISBN 0 85272 255 9.

Bowell, R.J. 1994. Sorption of Arsenic by Iron-Oxides and Oxyhydroxides in Soils. *Applied Geochemistry,* **9**(3), 279-286.

Breward, N. 2007. Arsenic and presumed resistate trace element geochemistry of the Lincolnshire (UK) sedimentary ironstones, revealed by a regional geochemical survey using soil, water and stream sediment sampling. *Applied Geochemistry*, 22 (9), 1970-1993.

Cave, M.R., Wragg, J., Palumbo, B. and Klinck, B.A. 2002. Measurement of the bioaccessibility of Arsenic in UK Soils. BGS-EA, P5-062/TR02.

Cave, M.R., Johnson, C.C., Ander, E.L. and Palumbo-Roe, B. 2012. Methodology for the determination of normal background contaminant concentrations in English soils. *British Geological Survey Commissioned Report*, CR/12/003. 56pp.

Communities and Local Government. 2007. Generalised Land Use Database Statistics for England 2005. Department for Communities and Local Government. Product Code 06CSRG04342. February 2007. Available online: http://www.communities.gov.uk/publications/planningandbuilding/generalisedlanduse last accessed 1st May 2012.

Darnley, A.G., Bjorklund, A., Bolviken, B., Gustavsson, N., Koval, P.V., Plant, J.A., Steenfelt, A., Tauchid, M. and Xuejing, X. 1995. A global geochemical database for environmental and resource management. *UNESCO publishing*, 19.

Davis, A., Ruby, M.V., Bloom, M., Schoof, R., Freeman, G. and Bergstom, P.D. 1996. Mineralogic constraints on the bioavailability of arsenic in smelter-impacted soils. *Environmental Science & Technology,* **30**, 392-399.

Fordyce, F.M., Brown, S.E., Ander, E.L., Rawlins, B.G., O'Donnell, K.E., Lister, T.R., Breward, N. and Johnson, C.C. 2005. GSUE: Urban geochemical mapping in Great Britain: *Geochemistry: Exploration-Environment-Analysis*, 5(4), 325-336.

Goa, S. and Burau, R.G. 1997. Environmental factors affecting rates of arsine evolution from mineralisation of arsenicals in soil. *Journal of Environmental Quality*, 26, 753-763.

Johnson, C.C., Breward, N., Ander, E.L. and Ault, L. 2005. G-BASE: Baseline geochemical mapping of Great Britain and Northern Ireland. *Geochemistry: Exploration, Environment, Analysis*, 5(4), 347-357.

Johnson, C.C., Ander, E.L., Cave, M.R. and Palumbo-Roe, B. 2012. Normal Background Concentrations of contaminants in English soil: Final project report. *British Geological Survey Commissioned Report*, CR/12/035. 40pp.

Kabata-Pendias, A. and Mukherjee, A.B. 2007. *Trace Elements from soil to Human*. Springer-Verlag, Berlin. 550pp.

Klinck, B., Palumbo, B., Cave, M.R., and Wragg, J. 2005. Arsenic dispersal and bioaccessibility in mine contaminated soils: a case study from an abandoned arsenic mine in Devon, UK. British Geological Survey Research Report No. RR/04/003, British Geological Survey, Keyworth, Nottingham, UK. Available on-line from NORA last accessed 1st May 2012.

Lark, R.M. 2012. Some considerations on aggregate sample supports for soil inventory and monitoring. *European Journal of Soil Science*, February 2012, 63, 86-95.

Lawley, R. 2011. The Soil-Parent Material Database: A User Guide. *British Geological Survey Open Report*, OR/08/034. 53pp. Available on-line from NORA last accessed 1st May 2012.

Masscheleyn, P.H., DeLaune, R.D. and Patrick, W.H. 1991. Effect of Redox Potential and pH on Arsenic Speciation and Solubility in a Contaminated Soil. *Environmental Science & Technology,* 25, 1414-1419.

Matheron, G. 1962. Traité de géostatistique appliquée. Tome 1, Editions Technip, Paris, 334pp.

McGrath, S P, and Loveland, P J. 1992. The Soil Geochemical Atlas of England and Wales, Blackie Academic and Professional, Glasgow.

Meunier, L., Walker, S.R., Wragg, J., Parsons, M.B., Koch, I., Jamieson, H.E. and Reimer, K.J. 2010. Effects of Soil Composition and Mineralogy on the Bioaccessibility of Arsenic from Tailings and Soil in Gold Mine Districts of Nova Scotia. *Environmental Science & Technology,* 44, 2667-2674.

Martin, I., De Burca, R. and Morgan, H. 2009. Soil Guideline Values for inorganic arsenic in soil. Science Report SC050021/arsenic SGV. Environment Agency, May 2009. Available on-line: http://www.environment-agency.gov.uk/static/documents/Research/SCHO0409BPVY-e-e.pdf last accessed 1st May 2012.

Nico, P S, Ruby, M V, Lowney, Y W, and Holm, S E. 2006. Chemical speciation and bioaccessibility of arsenic and chromium in chromated copper arsenate-treated wood and soils. *Environmental Science & Technology*, Vol. 40, 402-408.

Ordnance Survey, 2011. OS MasterMap Topography Layer. OS MasterMap® Topography Layer. http://www.ordnancesurvey.co.uk/oswebsite/products/os-mastermap/index.html last accessed 1st May 2012.

Ove Arup (1990). Mining Instabililty in Britain. Unpublished reports by Ove Arup for the Department of the Environment (DoE), UK. Associated data sets modified by BGS into GIS format (see http://www.bgs.ac.uk/gbase/NBCDefraProject.html)

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

Palumbo-Roe, B, and Klinck, B. 2007. Bioaccessibility of arsenic in mine waste-contaminated soils: A case study from an abandoned arsenic mine in SW England (UK). *Journal of Environmental Science and Health Part A*, Vol. 42, 1251 - 1261.

Rawlins, B.G., McGrath, S.P., Scheib, A.J., Cave, M.R., Breward, N., Lister, T.R., Ingham, M., Gowing, C.J.B., and Carter, S. 2012. The Advanced Soil Geochemical Atlas of England and Wales. British Geological Survey, Keyworth, Nottingham. Published as an electronic book. Published as an electronic book. Available on-line at: www.bgs.ac.uk/gbase/advsoilatlasEW.html last accessed 1st May 2012. Arsenic, pages 26-29.

Supplementary Information

Reimann, C., Matschullat, J., Birke, M. and Salminen, R. 2009. Arsenic distribution in the environment: The effects of scale. *Applied Geochemistry*, 24(7), 1147-1167.

Reimann, C., de Caritat, P., GEMAS Project Team and NGSA Project Team. 2012. New soil composition data for Europe and Australia: demonstrating comparability, identifying continental-scale processes and learning lessons for global geochemical mapping. *Science of The Total Environment*, 416, 239-252.

Salminen, R. (chief ed.) *et al*. 2005. Geochemical Atlas of Europe. Part 1 - Background Information, Methodology and Maps. Geological Survey of Finland, Otamedia Oy, Espoo, 525pp.

Smedley, P.L. and Kinniburgh, D.G. 2002. A review of the source, behaviour and distribution of arsenic in natural waters. *Applied Geochemistry,* 17, 517-568.

Thornton, I. 1996. Sources and pathways of arsenic in the geochemical environment: health implications. In: *Environmental Geochemistry and Health*. Appleton, J.D., Fuge, R. and McCall, G.J.H (editors). Geological Society Special Publication No 113, 153-161.

Webb, J S, Thornton, I, Thompson, M, Howarth, R J, and Lowenstein, P L. 1978. *The Wolfson geochemical atlas of England and Wales*. Clarendon Press, Oxford.

Wragg, J., Cave, M.R. and Nathanail, P. 2007. A Study of the relationship between arsenic bioaccessibility and its solid-phase distribution in soils from Wellingborough, UK. *Journal of Environmental Science and Health, Part A* 42, 1303 - 1315.

Xu, J. and Thornton, I. 1985. Arsenic in garden soils and vegetable crops in Cornwall, England: implications for human health. *Environmental Geochemistry and Health*, 7, 131-133.

Yang, J K, Barnett, M O, Jardine, P M, and Brooks, S C. 2003. Factors controlling the bioaccessibility of arsenic(V) and lead(II) in soil. *Soil & Sediment Contamination*, Vol. 12, 165-179.