



**British
Geological Survey**

NATURAL ENVIRONMENT RESEARCH COUNCIL



Normal background concentrations (NBCs) of contaminants in English soils: Final project report

Science Facilities Directorate

Commissioned Report CR/12/035^N

BRITISH GEOLOGICAL SURVEY

Science Facilities Directorate

COMMISSIONED REPORT CR/12/035^N

Normal background concentrations (NBCs) of contaminants in English soils: Final project report

C C Johnson, E L Ander, M R Cave and B Palumbo-Roe

The National Grid and other Ordnance Survey data © Crown Copyright and database rights 2012. Ordnance Survey Licence No. 100021290.

Keywords

soil, contaminated land, Statutory Guidance, contaminant.

Bibliographical reference

JOHNSON, CC, ANDER, EL, CAVE, MR and PALUMBO-ROE, B. 2012. Normal background concentrations (NBCs) of contaminants in English soils: Final project report. *British Geological Survey Commissioned Report*, CR/12/035. 40pp.

Copyright in materials derived from the British Geological Survey's work is owned by the Natural Environment Research Council (NERC) and/or the authority that commissioned the work. You may not copy or adapt this publication without first obtaining permission. Contact the BGS Intellectual Property Rights Section, British Geological Survey, Keyworth, e-mail ipr@bgs.ac.uk. You may quote extracts of a reasonable length without prior permission, provided a full acknowledgement is given of the source of the extract.

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of the Natural Environment Research Council.

British Geological Survey offices

BGS Central Enquiries Desk

Tel 0115 936 3143 Fax 0115 936 3276
email enquiries@bgs.ac.uk

Environmental Science Centre, Keyworth, Nottingham NG12 5GG

Tel 0115 936 3241 Fax 0115 936 3488
email sales@bgs.ac.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA

Tel 0131 667 1000 Fax 0131 668 2683
email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD

Tel 020 7589 4090 Fax 020 7584 8270
Tel 020 7942 5344/45 email bgs_london@bgs.ac.uk

Columbus House, Greenmeadow Springs, Tongwynlais, Cardiff CF15 7NE

Tel 029 2052 1962 Fax 029 2052 1963

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB

Tel 01491 838800 Fax 01491 692345

Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Belfast BT9 5BF

Tel 028 9038 8462 Fax 028 9038 8461

www.bgs.ac.uk/gsni/

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU

Tel 01793 411500 Fax 01793 411501
www.nerc.ac.uk

Website www.bgs.ac.uk

Shop online at www.geologyshop.com

Foreword

This is the final report for the Defra-funded Science and Research project SP1008 to establish normal background contaminant concentrations in the soils of England. The project was carried out between October 2011 and March 2012 and consisted of four work packages. Work package 1 (WP1) was concerned with a review of available contaminant data for the soils of England and Work package 2 (WP2) explored the principal available data sets. A robust statistical methodology for defining normal background concentrations of contaminants in soil was established as part of Work package 3 (WP3). The final deliverable has been a series of technical guidance sheets (TGSs) describing the determined normal background concentrations (NBCs) for a selected number of contaminants (WP4).

Acknowledgements

In addition to the authors named on this report, a number of other Natural Environment Research Council (NERC) staff have contributed to parts of this project, namely, Don Appleton, Murray Lark, Barry Rawlins, Chris Vane (all British Geological Survey) and Stephen Lofts (Centre for Ecology and Hydrology (CEH)). We are also grateful to the insight and comments from Paul Nathaniel of Land Quality Management, Nottingham regarding how the determination of normal background concentrations for contaminants fits within the revised Part 2A Contaminated Land Statutory Guidance. The Project also acknowledges the contribution of the Project "Steering Group" who have provided us with valuable feedback on our work to determine NBCs. Finally, we thank the Defra Soils Policy Team who have directed the Project to a successful conclusion.

We acknowledge the great efforts of those who have contributed to producing the soil data sets described in our metadata database (summarised in Appendix 2 of our WP1&2 report). We realise some information in this will need updating and we would appreciate any contributions that help us to improve this metadata.

The following licences have been issued with respect to some of the data sets investigated:

- Environmental Change Network (ECN): CEH – ECN:CJ10/11 – issued 6th December 2011 for one year
- Countryside Survey 10 km grid references for CS plots: CEH – issued 1st December 2011 for one year
- NSI Topsoil data – Data lease Code: L0212/00657 issued 12th December 2011 for one year (Cranfield ref: WU13058V)
- For the mineralisation and mining data layer based on the Department of Environment Ove Arup "Mining Instability in Britain" contract, copyright ownership of the "derived product" - the ArcView mining instability dataset - is vested in NERC/BGS (BGS IPR reference IPR35-10)

The Generalised Land Use Database Statistics (file 154329.xlsx) was used under the Open Government licence.

Contents

Foreword	i
Acknowledgements	i
Contents	ii
Summary	v
1 Introduction	1
2 Statutory Guidance and normal levels	4
3 Available contaminant data for English soils	5
3.1 Priority contaminants	5
3.2 Systematic sampling strategy	5
3.3 Spatially extensive	6
3.4 Quality assured	8
3.5 Total contaminant concentrations	8
3.6 Compatibility between data sets.....	9
3.7 Good resolution of site coordinates.....	9
4 Data Exploration	9
4.1 Spatial distributions	10
4.2 Data population distributions.....	14
5 Methodology for determining normal background concentrations (NBC)	16
6 NBCs for As, BaP, Cd, Cu, Hg, Ni and Pb	20
7 Concluding Remarks	22
Appendix 1 : Project Outputs	24
Abbreviations and Glossary	26
References	29

FIGURES

Figure 1: A summary of the Project's Work packages and their outputs	3
Figure 2: A map showing urban, semi-urban and rural areas of England defined using the Generalised Land Use Database	6
Figure 3: Map showing the distribution of samples used in the NBC determinations for As, Cd, Cu, Ni and Pb	7
Figure 4: Comparison of topsoil Pb concentrations in NSI soil samples done by two different analytical methods	8

Figure 5: Example of a contaminant interpolated geochemical image used to demonstrate contaminant variability in topsoils across England	11
Figure 6: Maps showing the contaminant domains (As, BaP, Cd, Cucont Hg, Ni, and Pb)	12
Figure 7: Example of a frequency distribution plot used in the data exploration – cumulative probability plot of topsoil As results categorised by domains	15
Figure 8: Example of a boxplot used to display the range of domain concentrations for As.....	15
Figure 9: Mercury cumulative probability plot	16
Figure 10:a Flow chart for the calculation of the NBC for a given contaminant – part I: Data gathering and exploration	18
Figure 10:b Flow chart for the calculation of the NBC for a given contaminant – part I: Data gathering and exploration	19

TABLES

Table 1: List of priority contaminants from Martin and Cowie (2008)	1
Table 2: A summary of the lead domain percentile classifications.	17
Table 3: Summary of domain normal background concentrations (NBCs) for the contaminants studied in this project.....	21

Summary

The British Geological Survey (BGS) has been commissioned by the Department for Environment, Food and Rural Affairs (Defra) to give guidance on what are normal levels of contaminants in English soils in support of the Part 2A Contaminated Land Statutory Guidance. This has initially been done by studying the distribution of four contaminants – arsenic, lead, benzo[*a*]pyrene (BaP) and asbestos – in topsoils from England. This work was extended to a further four contaminants (cadmium, copper, nickel and mercury) which enabled methodologies developed to be tested on a larger range of contaminants. The first phase of the Project gathered data sets that were: nationally extensive; systematically collected so a broad range of land uses were represented; and collected and analysed to demonstrably and acceptable levels of quality. Information on the soil contaminant concentrations in urban areas was of particular importance as the normal background is considered to be a combination of both natural and diffuse anthropogenic contributions to the soil. Issues of soil quality are most important in areas where these affect most people, namely, the urban environment. The two principal data sets used in this work are the BGS Geochemical Baseline Survey of the Environment (G-BASE) rural and urban topsoils (37,269 samples) and the English NSI (National Soil Inventory) topsoils (4,864 samples) reanalysed at the BGS laboratories by X-ray fluorescence spectrometry (XRFS) so both data sets were highly compatible. These two data sets provide results for most inorganic element contaminants, though results explored for mercury and BaP are drawn from a variety of different and much less extensive data sets.

The second phase of the Project explored the data focusing on contaminant spatial variability and population distributions. This investigated whether normal levels of contaminants could be quantified as normal background concentrations (NBCs). These are attributed to different regions of the country for each contaminant based on factors that were observed to contribute to higher concentrations in some areas, referred to as domains. The three most important contributing factors are: the underlying parent material upon which the soil has formed; non-ferrous metalliferous mineralisation and associated mining activity; and urbanisation. The data exploration determined that there was insufficient information about natural concentrations of asbestos in soils so it was not possible to determine NBCs for this contaminant.

A robust statistical methodology was developed in the third phase of the Project to deliver values for contaminant domain NBCs. This methodology is based on a study of a contaminant's population distribution. Using statistical measures (*e.g.* skewness coefficient and octile skew combined with inspection of population distribution plots) appropriate transformations are applied to the data to reduce the influence of outlying data points likely to have been caused by point source contamination. Percentiles for the domain data sets for each contaminant are generated along with calculations of percentile confidence intervals. The upper limit for a NBC has been defined as the upper 95% confidence limit of the 95th percentile. The NBCs can be used at national to regional scales as a guide as to what are normal levels and are meant for use in support of the Part 2A contaminated land regime. This methodology can be applied at more local scales by those who have their own sufficient and representative data sets.

For the contaminants for which NBCs could be determined, a series of technical guidance sheets (TGSs) have been written accompanied by supplementary information which includes many of the Project outputs.

1 Introduction

In April 2012 revised Part 2A Contaminated Land Statutory Guidance (SG) was issued by the Secretary of State for Environment, Food and Rural Affairs (Defra, 2012). This Guidance explains how the contaminated land regime should be implemented in England. Within the SG references are made to the “normal” presence and levels of contaminants in soils (see Section 2) and the British Geological Survey (BGS) has been commissioned by Defra to give guidance on what are normal concentrations of contaminants in English soils.

Eight contaminants were selected for study, namely, arsenic (As), asbestos, benzo[*a*]pyrene (BaP), cadmium (Cd), copper (Cu), mercury (Hg), nickel (Ni) and lead (Pb). These were chosen so as to represent contaminants produced from a variety of sources in the surface environment and that have variable quantity and quality of data. The use of the term “contaminant” is discussed in the SG (Defra, 2012, see Box 1). There are thousands of potential contaminants which might be present on various sites around England (although a smaller subset probably drives the risk on most sites) (Defra, 2008). A list of fifty six priority contaminants is presented by Martin and Cowie (2008) and this is shown in Table 1.

BOX 1 – Contaminant

“The terms “contaminant”, “pollutant” and “substance” as used in this Guidance have the same meaning – i.e. they all mean a substance relevant to the Part 2A regime which is in, on or under the land and which has the potential to cause significant harm to a relevant receptor, or to cause significant pollution of controlled waters. This Guidance mainly uses the term “contaminant” and associated terms such as “contaminant linkage”. However it recognises that some non-statutory technical guidance relevant to land contamination uses alternative terms such as “pollutant”, “substance” and associated terms in effect to mean the same thing.”

Statutory Guidance, page 4.

Inorganic	Organic		
arsenic (As)	acetone	dioxins and furans	tetrachloroethane
beryllium (Be)	aldrin	endosulfan	tetrachloroethene
cadmium (Cd)	atrazine	ethylbenzene	toluene
chromium (Cr)	azinphos-methyl	explosives	total petroleum hydrocarbons
copper (Cu)	benzene	fenitrothion	2, 1,1,1-trichloroethane
lead (Pb)	Benzo[<i>a</i>]pyrene (BaP)	hexachlorobuta-1,3-diene	trichloroethene
mercury (Hg)	carbon disulphide	hexachlorocyclohexanes	trifluralin
molybdenum (Mo)	carbon tetrachloride	malathion	vinyl chloride
nickel (Ni)	chloroform	naphthalene	xylenes
selenium (Se)	chlorobenzenes	organolead compounds	
sulphur (S)	chlorophenols	organotin compounds	
thallium (Tl)	chlorotoluenes	pentachlorophenol	
vanadium (V)	1,2-dichloroethane	phenol	
zinc (Zn)	dichlorvos	polychlorinated biphenyls (PCB)	
cyanide	DDT	polynuclear(cyclic) aromatic	
asbestos	dieldrin	hydrocarbons (PAH)	

Table 1: List of priority contaminants from Martin and Cowie (2008)

The BGS work to quantify what “normal” levels of contaminants are in the soils of England has followed the stages illustrated in Figure 1. Each stage is defined as a Work package and these have been described in detail in previous reports. Work packages 1 and 2 (WP1 and WP2) were the data gathering and data exploration phases of the Project and have been documented in Ander *et al.* (2011, 2012). Work package 1 sought out available contaminant data for English soils that was of reliable quality and has been systematically collected to represent a broad spectrum of land uses. The principal output from this is discovery metadata that summarises the available data sets (see Section 3).

Work package 2 explored the contaminant data, principally that from two major systematic national surveys. These were from the National Soil Inventory (NSI) for England (National Soil Research Institute (NSRI), Cranfield University, samples reanalysed by X-ray fluorescence spectrometry (XRFs) at BGS), and BGS Geochemical Baseline Survey of the Environment (G-BASE) rural and urban topsoil results. The data exploration was focused on contaminant distributions both across England and statistically when plotted and mapped by a variety of techniques. The objective of WP2 was to identify the main factors controlling the concentration and distribution of the selected contaminants in soils. Domains (see Box 2) were identified for each contaminant for attribution with NBCs and these are based on the three principal controls that contribute to contaminant concentrations in the soil – the soil parent material, urbanisation, and non-ferrous metalliferous mineralisation and associated mining/processing activities. The four initial contaminants selected for this project – As, Pb, BaP and asbestos – are investigated in detail in the WP1 and 2 report (Ander *et al.* 2011). Additional contaminants Cd, Cu, Hg and Ni were explored after the completion of WP1 and 2, and these are reported in Ander *et al.* (2012), the technical guidance sheets and supplementary information (see below).

Work package 3 (WP3) developed a robust statistical methodology for quantifying normal background concentrations (Cave *et al.* 2012) and generated percentile information, including percentile confidence intervals, for each contaminant domain. The upper limit for a NBC is defined as the upper 95% confidence limit of the 95th percentile (see Section 5). This is our representation of the “normal” levels referred to in the SG based on current available data. The data exploration of WP2 showed that there was insufficient data to establish NBCs for asbestos, so these were only determined for the seven remaining contaminants – As, BaP, Cd, Cu, Hg, Ni and Pb. For BaP there was insufficient data for English soils to calculate NBCs, so data from Wales and Scotland were also used.

BOX 2 – Domains

The term “domain” has been used by this project to identify areas of England to which high concentrations of a contaminant can be attributed as a result of readily distinguishable controlling factors. Such regions are defined by a boundary derived from a soil’s underlying parent material, an urbanisation index, or an area of non-ferrous metalliferous mineralisation with associated mining activities. The area remaining outside domains defined by these controlling factors is referred to as the Principal Domain.

The final Work package (WP4) produced a series of technical guidance sheets (TGSs) with supplementary information for each contaminant (see Appendix 1). These TGSs list the domain NBCs for respective contaminants and summarise the natural and diffuse anthropogenic factors contributing to a contaminant’s concentration and distribution. The use of NBCs is also introduced in the TGSs, and additional information generated from exploration of the available results for each contaminant (*e.g.* frequency distribution plots and summary statistics) are included in the TGS supplementary information.

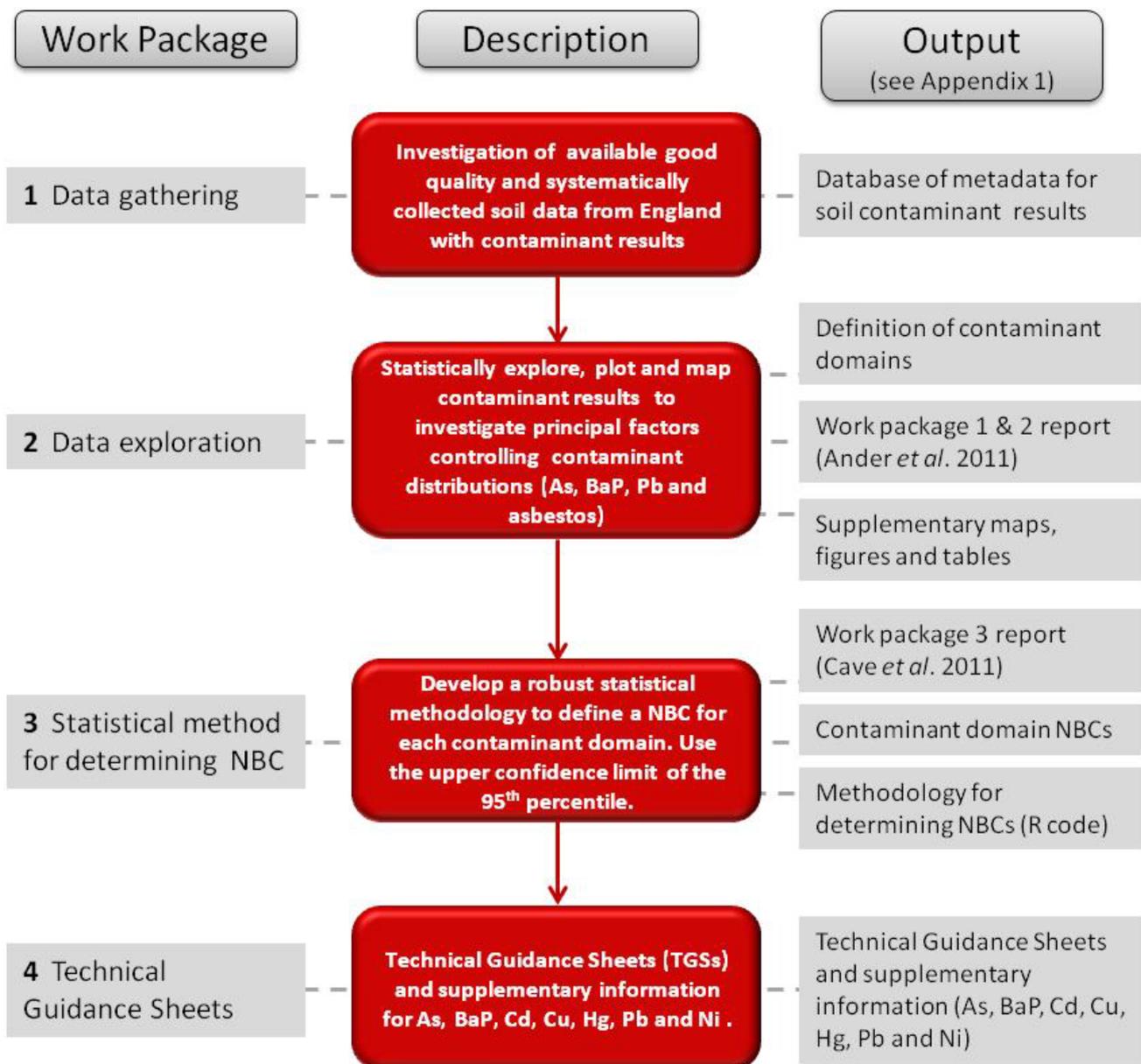


Figure 1: A summary of the Project's Work packages and their outputs

2 Statutory Guidance and normal levels

There are a number of similar terms used to convey the meaning of “normal” levels of contaminant concentrations including; widespread, typical, natural, background, baseline, ambient and characteristic. In the context of this work, the definition that matters as to what constitutes normal is that given in the SG (see Box 3).

BOX 3 – Statutory Guidance “normal” presence

“3.21 The Part 2A regime was introduced to help identify and deal with land which poses unacceptable levels of risk. It is not intended to apply to land with levels of contaminants in soil that are commonplace and widespread throughout England or parts of it, and for which in the very large majority of cases there is no reason to consider that there is an unacceptable risk.

3.22 Normal levels of contaminants in soil should not be considered to cause land to qualify as contaminated land, unless there is a particular reason to consider otherwise. Therefore, if it is established that land is at or close to normal levels of particular contaminants, it should usually not be considered further in relation to the Part 2A regime and the local authority should have regard to paragraphs 5.2 to 5.4 of this Guidance.

3.23 For the purpose of this Guidance, “normal” levels of contaminants in soil may result from:

(a) The natural presence of contaminants (e.g. caused by soil formation processes and underlying geology) at levels that might reasonably be considered typical in a given area and have not been shown to pose an unacceptable risk to health or the environment.

(b) The presence of contaminants caused by low level diffuse pollution, and common human activity other than specific industrial processes. For example, this would include diffuse pollution caused by historic use of leaded petrol and the presence of benzo[a]pyrene from vehicle exhausts, and the spreading of domestic ash in gardens at levels that might reasonably be considered typical.

3.24 In deciding whether land has normal levels of contaminants, the local authority should consider whether contamination is within the bounds of what might be considered typical or widespread: (a) locally, if there is sufficient information to make a reasonable consideration of what is normal within a local area; and/or (b) regionally or nationally in broadly similar circumstances, having due regard to similarity in terms of land use and other relevant factors such as soil type, hydrogeology, and the form of the contaminants.

3.25 The local authority should decide that normal levels of contaminants exist in relation to land where: (a) those levels are not significantly different to those likely to be typical or widespread within the authority’s area, or in other similar areas; and/or (b) those levels are common or usual in similar land use situations across England or parts of it; and (c) there is no specific reason to consider that those levels of contaminants are likely to pose an unacceptable risk.

3.26 It is possible that specific pieces of land at or slightly above normal levels of contamination with regard to specific substances may pose sufficient risk to be contaminated land, and that remediation of such land may bring significant net benefits. However, such cases are likely to be very unusual and the authority should take particular care to explain why the decision has been taken, and to ensure that it is supported by robust scientifically-based evidence.”

Defra (2012) Statutory Guidance, Sections 3.21-3.26.

An important part of the definition of “normal” is that it includes both the natural and diffuse anthropogenic contribution to a contaminants concentration in the soil. In this sense it is similar to the background content defined by ISO 19258:2011 (ISO, 2011) “*content of a substance in a soil resulting from both natural geological and pedological processes and including diffuse source inputs*”.

The work described here is our attempt to represent and quantify the term “normal” levels by defining normal background concentrations (NBCs) for contaminants in the soils of England (Section 5).

3 Available contaminant data for English soils

Work Package 1 investigated available contaminant data for English soils but also sought information for soils outside of England when the number of results available was small (*e.g.* BaP). For data sets to be applicable, results should:

1. Include priority contaminants;
2. Be associated with a systematic rather than a targeted sampling strategy so as to represent a broad range of land use types;
3. Be spatially extensive across England with a good sample density;
4. Be soils that have been collected and analysed to internationally recognised standards and have associated quality assurance;
5. Unambiguously define total concentrations of contaminants;
6. Be compatible with other available data sets; and
7. Provide good resolution of the sample site coordinates.

Information regarding available contaminant data sets for soils has been collated in a MS Access database and reported as an Appendix in Ander *et al.* 2011.

3.1 PRIORITY CONTAMINANTS

A list of priority contaminants is given in Table 1. There is much data available for the inorganic elements on this list (with the exception of Hg) but far less information available on organic contaminants. Asbestos is unique in the list in that it is the mineral form of this substance and the concentration of minerals and the potential pathway into the lungs that is of prime concern. Other contaminants are measured by the elemental or substance concentrations in the soil. Lack of data on the occurrence of naturally occurring asbestos minerals in soil meant it was inappropriate to determine NBCs for asbestos. There are considerable gaps in knowledge regarding both the natural contribution of organic contaminants to soils and the levels of diffuse pollution of these elements (represented by systematically rather than targeted sampling) in urban and rural areas. For many of the organic contaminants listed in Table 1 the natural contribution to “normal” levels is thought to be minimal to non-existent and the diffuse anthropogenic pollution is likely to be of greatest concern.

3.2 SYSTEMATIC SAMPLING STRATEGY

In order to determine unbiased normal background concentrations for soils, it is important that the results used represent the broad spectrum of underlying parent material and land uses across England, *i.e.* the results of systematic rather than targeted surveys. Investigations of contaminated land will produce results that reflect the high contaminant levels generally associated with contaminated areas. Similarly, soil chemical data associated with metalliferous mineral exploration will be biased to known areas of mineralisation and have high levels of contaminants from the mineralisation being sought. The most appropriate data sets used in this work are the systematically collected national (*e.g.* NSI or G-BASE) or international soil surveys (*e.g.* FOREGS - Salminen *et al.* 2005 and GEMAS - Reimann *et al.* 2012). The British Geological Survey is the only UK organisation to have a national programme of systematically sampling urban areas (Fordyce *et al.* 2005; Flight and Scheib 2011; Box 4). This urban chemical data for

soils is of great value in the determination of NBCs. Urban/semi urban and industrial areas cover some 34% of the land area of England (Figure 2). It is where most of the population live so the health risks to humans and the environment are greatest. Most other national systematic surveys focus on just rural or agricultural areas and fail to capture the very important diffuse anthropogenic pollution of urban areas. This is particularly important in England because of the long history of industrialisation and urbanisation which has left a legacy of soils with elevated levels of many contaminants. In 2011 BGS released its London Earth soil chemical data representing the largest systematic urban sampling programme undertaken anywhere in the world.

BOX 4 – G-BASE urban sampling

The G-BASE project has amassed soil geochemical data from more than 20 cities in the United Kingdom, as part of its National Capability remit. A huge effort has been made to maintain standardised sampling, analytical and quality control methods, thus ensuring consistent data output. The project continues to seek to generate data on as wide a range of analytes as possible in order to remain relevant and to inform a diverse range of scientifically and legislatively driven needs for soil geochemical information, both now and in the future.

Flight and Scheib (2011).

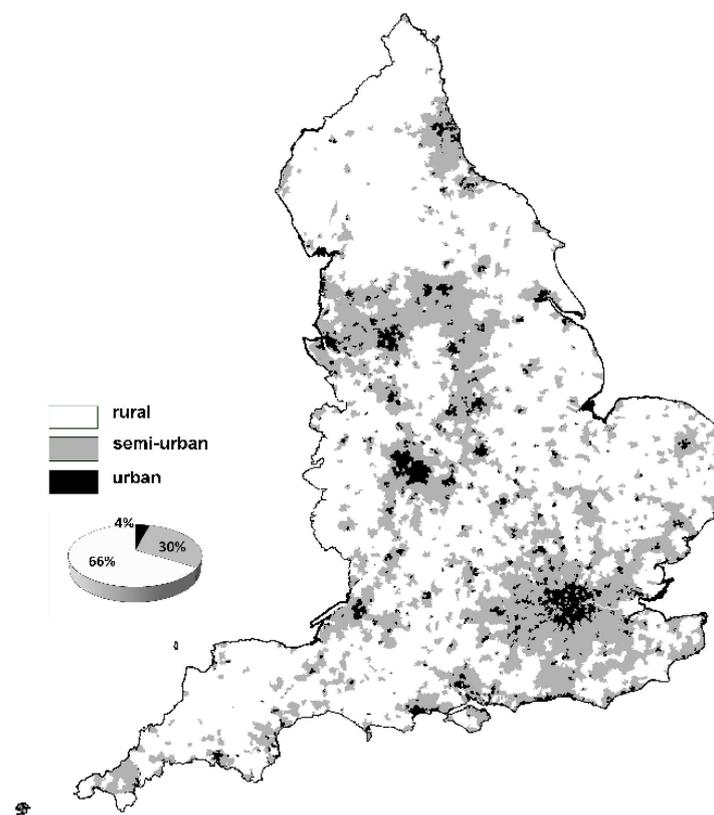


Figure 2: A map showing urban, semi-urban and rural areas of England defined using the Generalised Land Use Database (GLUD - Communities and Local Government, 2007). The method for defining an urbanisation index is described in Ander *et al.* (2011)

3.3 SPATIALLY EXTENSIVE

In order to determine spatially representative NBCs, data sets are required that have a good coverage of England. The best coverage (all of England), and at a sampling density of one soil site every 25 km², is the NSI data (Oliver *et al.*, 2002; Figure 3) targeted at rural areas. However, the only large data set to include both urban and rural soil results, and done using consistent

sampling and analytical methods is the G-BASE data set for which the urban and rural samples are collected at a density of one site per 0.25 km² and 2 km², respectively (Fordyce *et al.*, 2005; Johnson *et al.*, 2005; Figure 3). As both the NSI (4,864 soils) and G-BASE (37,269 soils) samples have been analysed in the BGS laboratories by XRFs, the NSI (XRFs) and G-BASE samples form the primary data sets used in this project for the inorganic elements (except Hg). Other data sets that cover all of England, albeit at low sampling densities, are the UK Soil and Herbage Pollutant Survey (Barraclough (2007) – UK-wide 156 urban/industrial and rural samples and including organic contaminant and Hg results); GEMAS (Reimann *et al.* 2012 – Europe-wide sampling of agricultural soils, including 131 samples from England); FOREGS (Salminen *et al.* 2005 – Europe-wide sampling to define the geochemical baseline across Europe, including 33 soil samples from England); and Countryside Survey (Emmet *et al.* 2010 – a monitoring programme for rural soils (76) in Great Britain and including organic contaminant and Hg results). These data sets are described in more detail in Ander *et al.* (2011). There is also an ongoing Europe-wide soil programme for which 684 soil samples from England have been collected for heavy metal analysis (LUCAS project, Montanarella *et al.* 2011). Sample chemical analyses are not yet complete. English soil results for systematically collected BaP and Hg data are not spatially extensive and those data sets mentioned above are supplemented with results from peer-reviewed publications (*e.g.* Cousins *et al.* 1997; Tipping *et al.* 2011).

It should be noted that whilst surveys such as the Countryside survey are monitoring contaminants over a period of time, most surveys are a single one off sampling collecting soils that represent the background at a specific point in time. NSI samples were collected between 1978 - 1983 and G-BASE topsoil samples have been collected since 1986.

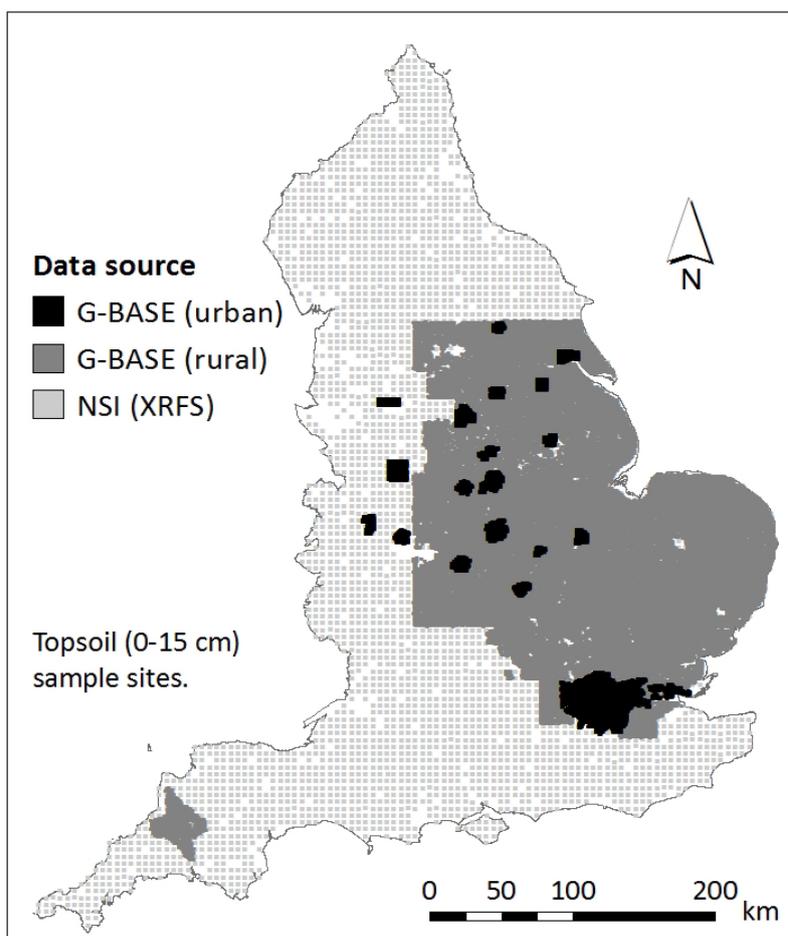


Figure 3: Map showing the distribution of samples used in the NBC determinations for As, Cd, Cu, Ni and Pb. NSI(XRFs) covers the whole of England at a sample density of 1:25 km². G-BASE sampling densities for rural and urban are 1:2 km² and 4:1 km², respectively.

3.4 QUALITY ASSURED

Only data that has documented methods for sampling and analysis, and has in place some measures for quality assurance, should be considered for NBC determinations. Procedures for soil sampling and chemical analysis vary between scientific disciplines. ISO 19258:2011 (ISO 2011) and the references therein provide some basic guidance on quality control for soil sampling and analysis. A generic top level document on “A global geochemical database for environment and resource management” (Darnley *et al.* 1995) has been effective in establishing common methodologies and quality control measures for the worldwide sampling of soils and other geological materials.

3.5 TOTAL CONTAMINANT CONCENTRATIONS

Many analytical measurements of contaminants are quoted as ‘total’ concentrations. For inorganic contaminants, determining true total concentrations usually involves direct analysis by non-destructive XRFs or neutron activation analysis (NAA). Destructive analytical methods involving sample digestion in an acid solution, for example, a combination of hydrofluoric-nitric-perchloric acid, gives “near total” results. Less vigorous reagents used in the extraction, such as *aqua regia*, will yield more readily available concentration of contaminants in soils.

BOX 5 – use of total analyses for soils

“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.”

Darnley *et al.* (1995). Section 6.3.2

For this project unambiguously total concentrations from XRFs analysis have been used. This conforms to Darnley *et al.*'s (1995) requirements (see Box 5) and furthermore, nearly all the large systematically and quality assured data sets available have been analysed by XRFs.

During the data exploration phase of this project, the relationship between XRFs analysis and other methods of determination has been investigated for those data sets where the same samples have been analysed by different analytical methods. This is reported in Ander *et al.* (2011, 2012) and in some of the contaminant TGS supplementary information. For elements investigated - As, Cd, Cu, Ni and Pb - there was seen to be a generally good linear relationship between XRFs and other strong acid extraction methods (*e.g.* Figure 4), though users of such an interpretation are cautioned that at some very high contaminant concentrations, depending on the source and form of contaminant, this relationship does not always hold true.

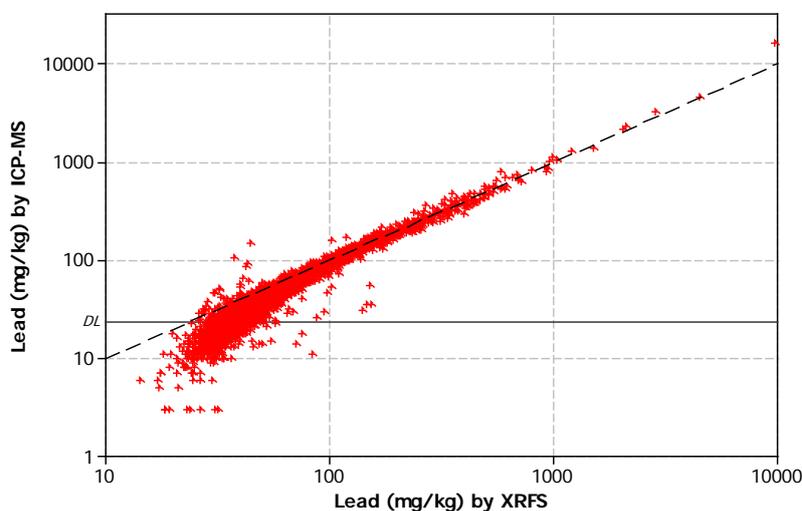


Figure 4: Comparison of topsoil Pb concentrations in NSI soil samples done by two different analytical methods - XRFs and *aqua regia* digest followed by ICP-MS.

DL = detection limit.

3.6 COMPATIBILITY BETWEEN DATA SETS

A major consideration in deciding what data sets to use in the calculation of contaminant NBCs is ensuring that the data has been generated by samples collected in an acceptably consistent manner and that the analytical methodology is compatible. Although there are some sampling procedural differences between NSI and G-BASE projects (*e.g.* NSI samples are aggregated from 25 subsamples collected within a 20 m square, G-BASE samples are aggregated from 5 subsamples over a similar area), both data sets are deemed to be similar and treated as a single data set. Samples collected using substantially different sampling methodologies and/or different analytical methods, where no data levelling has been done, are deemed incompatible and are not used to calculate NBCs. However, for the contaminants where there is relatively little data (*i.e.* BaP and Hg), there is not so much scope for insisting on complete data compatibility and, as is the case for Hg results, data derived from different analytical methods are used creating further uncertainties in the data interpretations (see Section 4.2).

3.7 GOOD RESOLUTION OF SITE COORDINATES

During the data exploration phase (WP2) it has been necessary to work with the sample data in a GIS environment to capture the sample points into polygons from landscape layers such as urban areas, parent material polygons and areas of non-ferrous metalliferous mineralisation. In order to capture the point data the spatial resolution has to be better than ± 100 m, especially when the landscape data polygons are mapped at scales of 1:50,000. A number of the data sets, as a result of sample site access agreements, publish data with coordinate data degraded so locations of the sites cannot be reasonably resolved on large scale maps. The UK Soil and Heritage data and Countryside Survey are published with degraded coordinates (to the nearest 10 km) and this greatly limits how the data can be used to attribute domains. Both the NSI and G-BASE data sets are published with good coordinate spatial resolution and this greatly enhances the usefulness of these data sets.

4 Data Exploration

Data exploration for As, Cd, Cu, Ni and Pb has used the NSI (XRFS) and G-BASE topsoil results (upper 15 cm of the soil profile). This represents a substantial data set of approximately 42,000 samples covering all of England (Figure 3) and includes some 13,500 topsoils from the G-BASE urban programme. These data sets were also supplemented by other less densely sampled surveys to provide information about how contaminant concentrations varied with depth and between different analytical methods (Ander *et al.* 2011). For Hg and BaP there were far fewer results to use in the data exploration (1,638 and 403 respectively), but there were still sufficient for statistical analysis and NBC determination. BaP NBCs are determined for domains covering Great Britain rather than just England. As described earlier, for BaP and Hg more minor national data sets and peer-reviewed publications were used to provide results for England, Scotland and Wales to use in the data exploration. The two principal software applications used for the data exploration were: Minitab v16 statistical and plotting software and ESRI ArcGIS v9.3. The results of this phase of the project (WP2) are reported in detail for As, Pb and BaP in Ander *et al.* (2011) and for all the contaminants investigated summarised in the series of TGSs and associated supplementary information (see Appendix 1). In the earlier phase of the project it was recognised that there was insufficient information available on naturally occurring asbestos minerals and this contaminant was not explored any further.

BOX 6 – ISO 19258:2011 General considerations

“Probability distributions of substances in soils are rarely normal. They are often positively skewed but not necessarily log-normal. The estimation of the required number of samples to assess variability of such a distribution can then be equal to the number of samples necessary to draw a representative histogram or to calculate representative percentile. To this end, a minimum number of 30 samples is recommended.” (Section 5.4.1.2).

“The pedo-geochemical or background content can be regarded as a statistical population. The objective of the data processing is to represent and characterize this population, using a variate of n individual values. Therefore, the data processing is done in order to

a) select the data corresponding to pedo-geochemical or background content from the whole data set (e.g. test for outliers), and

b) characterize the distribution in such a way that a frequency or probability can be associated, at least approximately, to any of the selected values.” (Section 5.5.1.1).

(ISO 2011)

There are numerous texts and published examples of the exploration of chemical data for soils. The ISO 19258:2011 (ISO 2011), for example, provides some basic guidance on the statistical analysis of “pedo-geochemical” data (see Box 6). ISO 19258:2011 also gives useful guidance on the viable minimum number of samples that should be used - a minimum number of 30 samples is recommended (Box 6). A more detailed account on methods for the statistical analysis of geochemical data is presented by Reimann *et al.* (2008). Oliver *et al.*'s (2002) statistical and geostatistical analysis of the NSI *aqua regia* data is a good example of a thorough exploration of a large soil data set.

Two important components of the data exploration are: how the geospatial distribution relates to possible sources and controlling factors of contaminant distribution in the soil; and, a study of the data population distributions.

4.1 SPATIAL DISTRIBUTIONS

Where there is sufficient contaminant data (As, Cd, Cu, Ni and Pb), national maps can be plotted by interpolating between sample sites to give geochemical images such as that shown in Figure 5. These gridded images can be classified and coloured in several ways and in the TGSs and supplementary information are produced as a simple monochrome image and a full colour image with a greater number of classes. These maps give an excellent visualisation of a contaminant's variability in soils across England caused mainly by three main factors:

1. The single most important controlling factor is the underlying parent material (principally geology) which provides the geogenic component of a contaminant's natural background. England has a very varied geology both in the age of rocks and the rock types, and this contributes to a significant variability in the natural contaminant concentrations.
2. A further geogenic component is when a contaminant is enriched in soils because of mineralisation in the underlying rocks. This may also be associated with an anthropogenic component caused by mining related activities.
3. England has a long history of urbanisation and industrialisation and, whilst there are now many environmental safeguards in place, there is a legacy of pollution in our cities and towns. This is represented by both point source and diffuse anthropogenic inputs to the natural background.

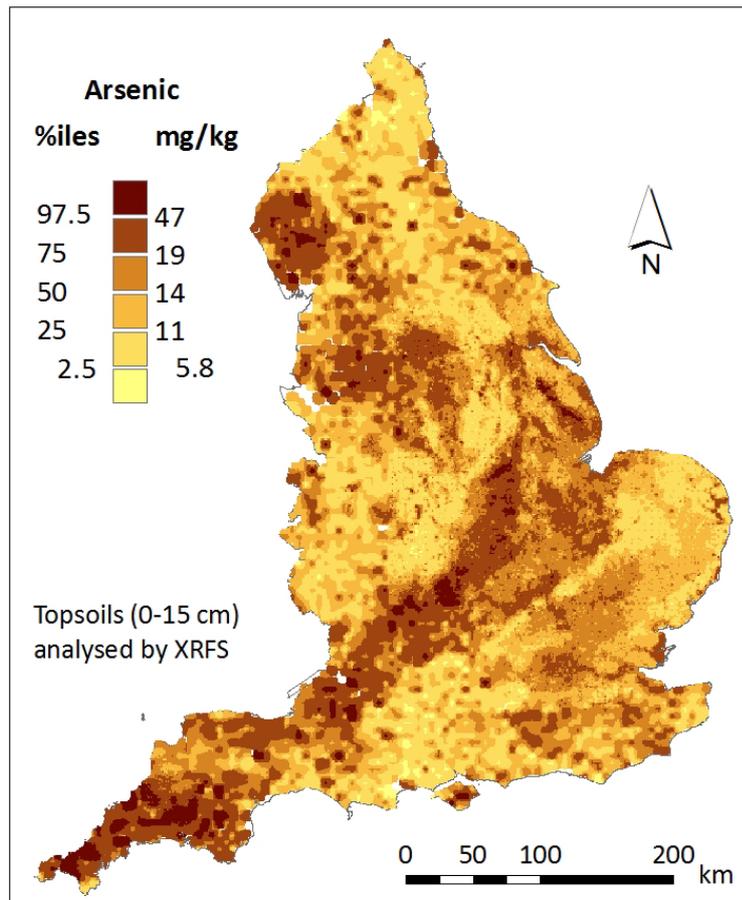


Figure 5: Example of a contaminant interpolated geochemical image used to demonstrate contaminant variability in topsoils across England. (Inverse distance squared weighting option, cell size 1000 m and search radius 5000 m. %iles = percentiles)

Mercury and BaP have far less data points to use in the data exploration, thus sites are widely spaced and so it is not a valid exercise to create interpolated maps for these elements. Instead of interpolated images, classified point maps are plotted to illustrate the spatial variability in these contaminants (see TGSs for Hg and BaP).

As there is spatial variability seen in the contaminant distributions, the approach of this project is not to apply a single national NBC but to delineate the principal areas of the country where the three important contaminant controlling factors listed above can be identified, and, as appropriate, a NBC calculated for each of them. An important part of the data exploration is therefore to identify contaminant domains across England, particularly those where the contaminant concentrations are highest. This has been achieved by a statistical process called k-means cluster analysis and by plotting the results as maps that highlight the regions where the high contaminant results occur (see Ander *et al.* 2011). The results can then be interpreted in the GIS environment using layers of information that show areas of urbanisation, non-ferrous metalliferous mineralisation and mining, and the underlying parent material. The various contaminant domains determined are shown in Figure 6.

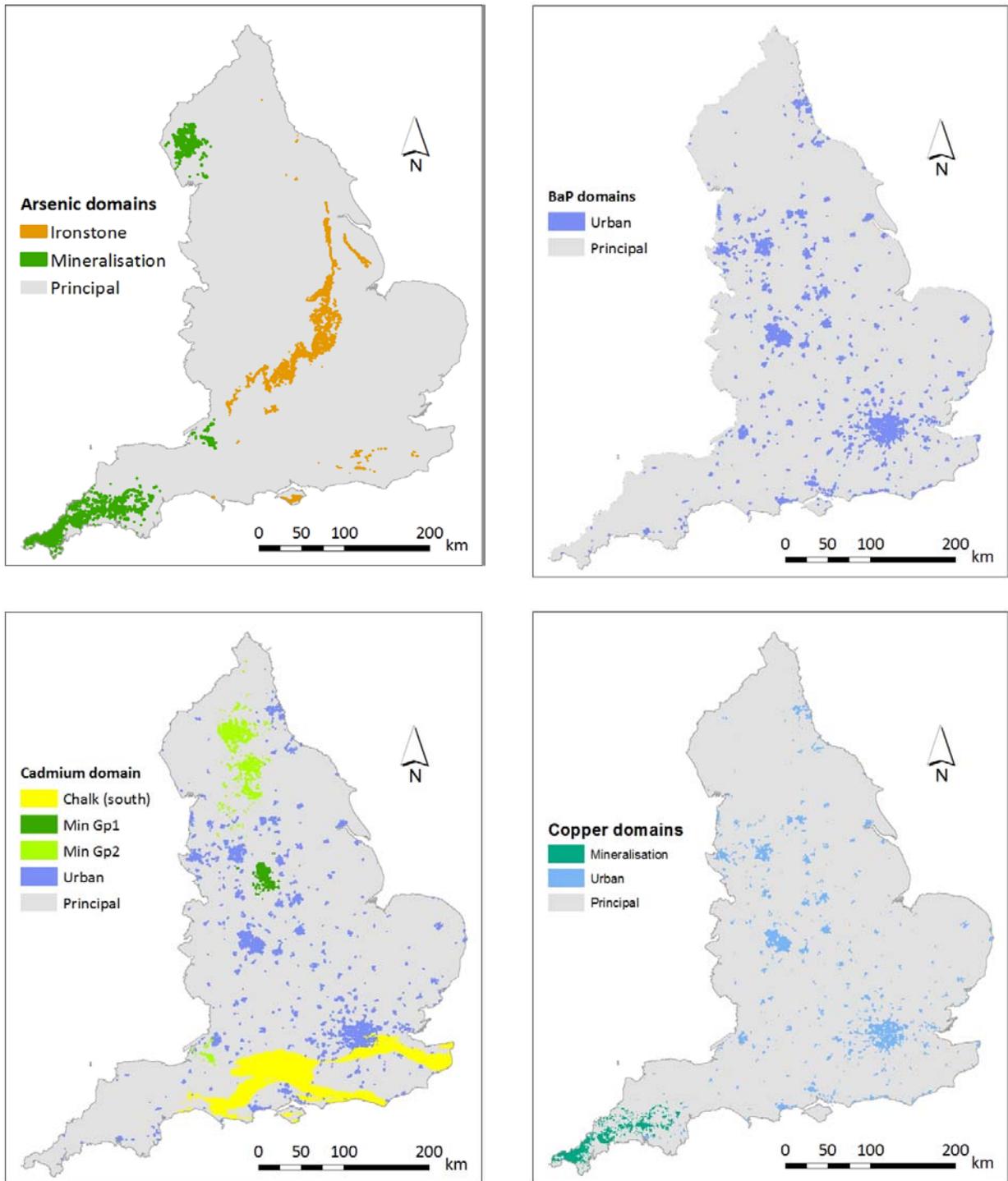


Figure 6: Maps showing the contaminant domains (As, BaP, Cd, Cucont Hg, Ni, and Pb)

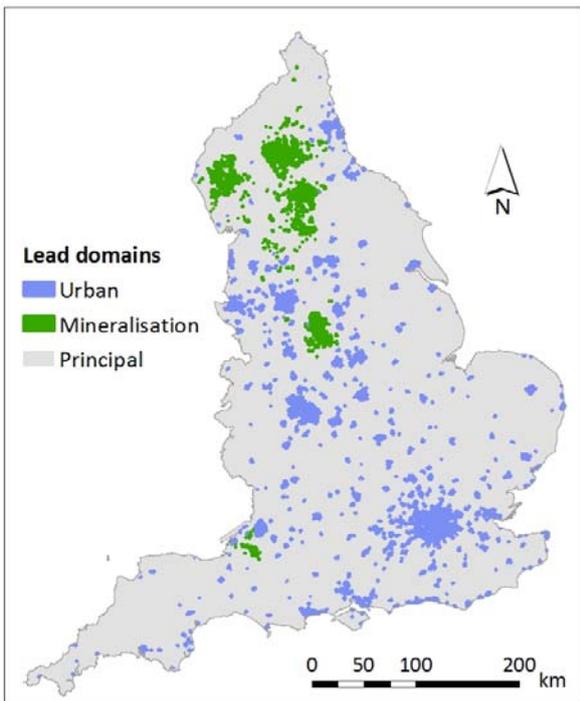
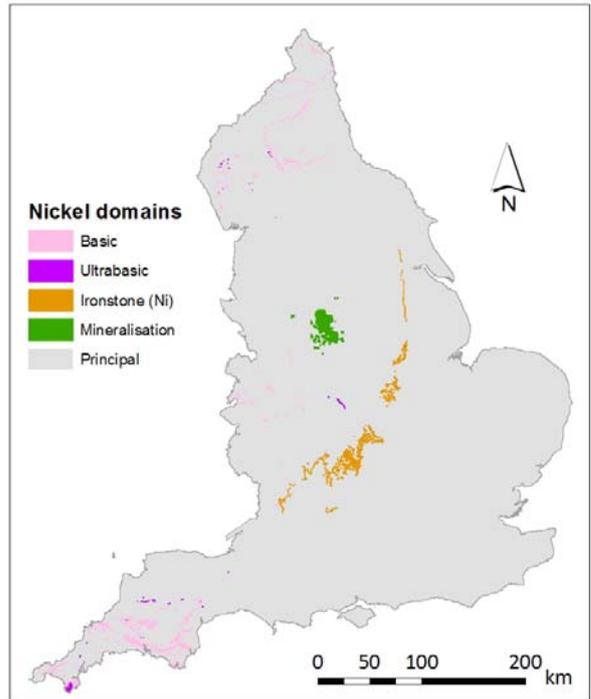
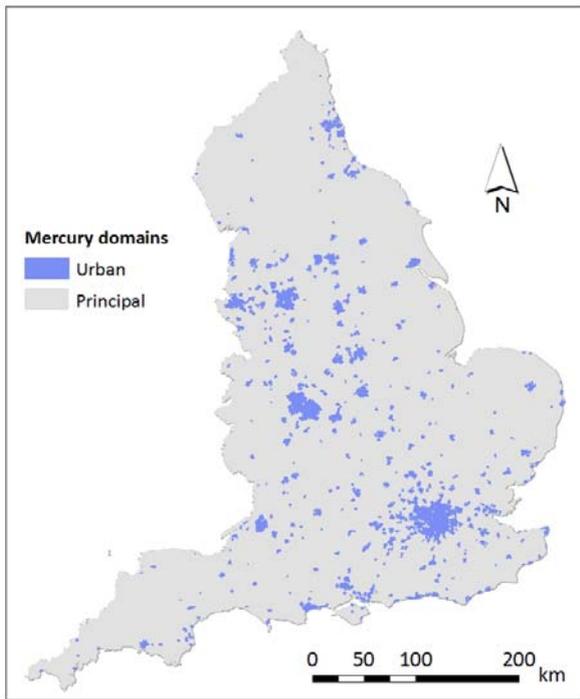


Figure 6: (continued). Maps showing the contaminant domains (Hg, Ni and Pb)

4.1.1 Urban domain

Land use is an important factor in controlling anthropogenic contaminant contributions to the soil. Certain activities, for example, metallurgical industries, areas of high traffic volume and coal burning, have been, and in some instances continue to be, responsible for raising contaminant levels in the environment. At the scale of this work we have not considered any one specific land use but included industrial activities and the built environment under the general classification of urban. In order to delineate this, an index of urbanisation has been defined (see Ander *et al.* 2011) using the Generalised Land Use Database 2005 statistics for England (Communities and Local Government, 2007). Based on the index, a map of urban, semi-urban and rural areas has been created (Figure 2) and the delineated urban areas used to define an urban domain for many of the contaminants. Those contaminants, where only the urban area is identified as the important controlling factor on high contaminant concentrations, will have just two domains associated with them – the Urban Domain and the Principal Domain (*i.e.* non-urban areas).

4.1.2 Mineralisation domain

A significant contributor to high levels of many inorganic contaminants in soil is non-ferrous mineralisation and associated mining activities, referred to in this project as the mineralisation domain. Note that this domain does not include ferrous or coal mining which are also noted to have significant controls on contaminant levels in soils. Such mining is generally related to specific rock strata rather than mineral veins that can cut across a variety of rock types and so are investigated by looking at underlying parent material (Section 4.1.3). In the GIS environment the Project has used the non-ferrous Metalliferous Mineralisation and Mining database, originally produced in hard-copy by Ove Arup (1990) for DoE (Department of Environment), and made into a digital product by BGS.

4.1.3 Underlying parent material

The Soil-Parent Material Model (SPMM) (Lawley, 2011) has been developed by BGS, using as its basis the mapped boundaries of the national 1:50,000 superficial and bedrock geological data and is used within a GIS environment by this project to identify the most influential contributors to high contaminant concentrations. 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 BGS 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 into the soil. In the SPMM there is additional information on texture, mineralogy and lithology not present in the geological mapping data, which is attributed in a hierarchical classification system.

4.2 DATA POPULATION DISTRIBUTIONS

Data population distributions are a fundamental part of the methodology for determining NBCs (Section 5) but are also an important part of the data exploration. Plots of the data frequency distributions (*e.g.* cumulative probability plots such as those illustrated in Figure 7) identify important features concerning the modelling of the data population such as whether a

distribution is skewed, lognormally distributed, truncated by lower detections limits or composed of multiple populations. It is particularly useful to look at classified frequency distribution plots so different data population structures can be compared. Another useful plot for exploratory data analysis is the boxplot (Figure 8). Frequency distribution plots and boxplots are included in each contaminant's TGS supplementary information.

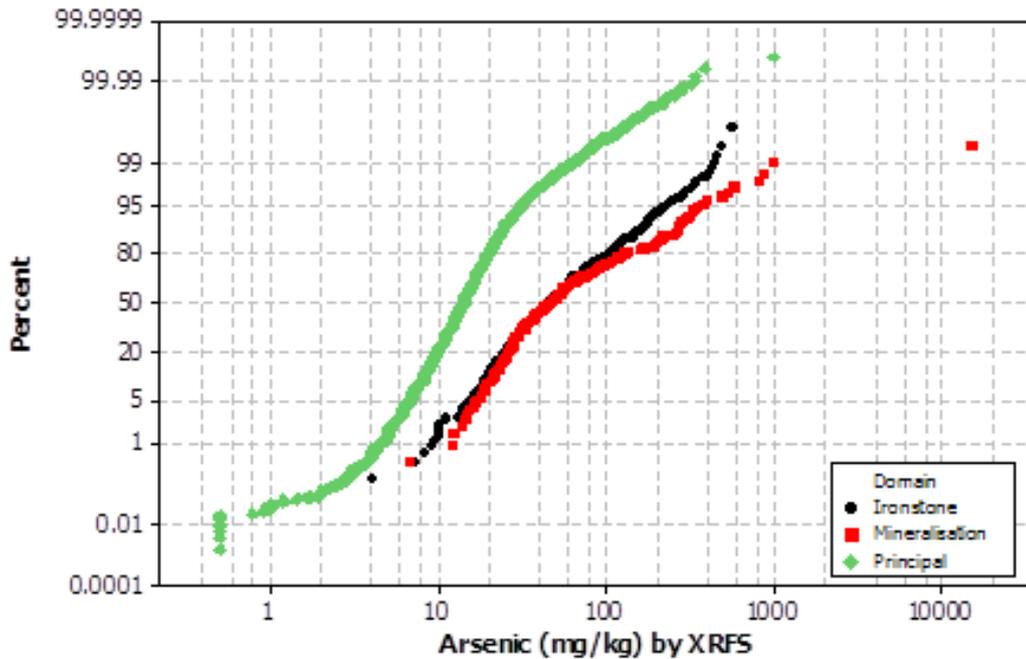


Figure 7: Example of a frequency distribution plot used in the data exploration – cumulative probability plot of topsoil As results categorised by domains

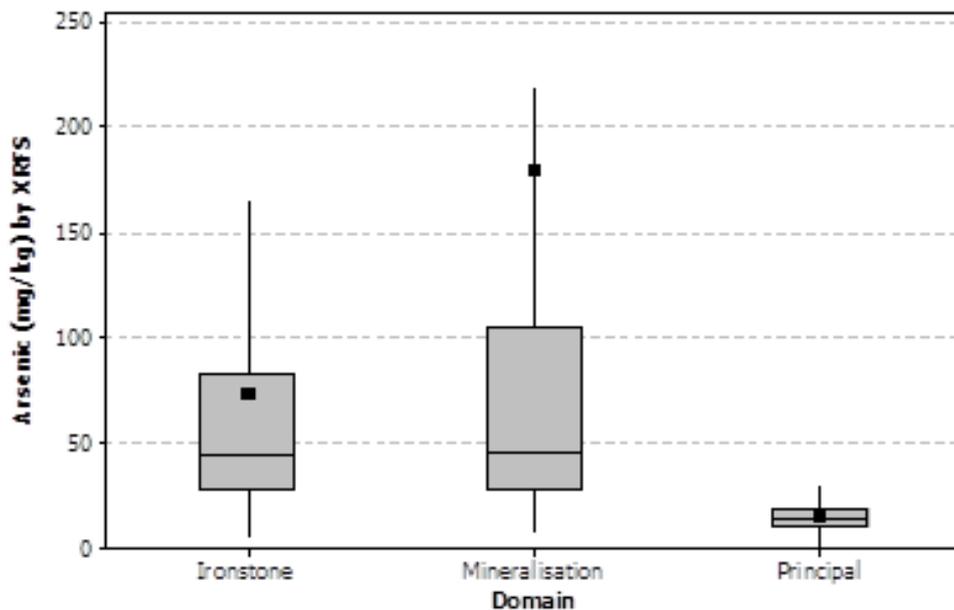


Figure 8: Example of a boxplot used to display the range of domain concentrations for As. The box represents the interquartile range (Q1, Q3), with the median (Q2) as a line within the box. The point symbol shows the mean value. The upper whisker = $Q3 + 1.5(Q3-Q1)$; lower whisker = $Q1 - 1.5(Q3-Q1)$

Whilst Figure 7 shows population distributions for compatible data sets (NSI (XRFS) and G-BASE), Figure 9 shows the distribution plot for Hg results from a variety of different sources. There are clearly significant differences in the population distributions which may be due to variations in the analytical methods used. At lower concentrations there are clearly differing detection limits (shown by the vertical line of data) and data sets analysed at the same laboratory (*i.e.* FOREGS_A, G-BASE London and GEMAS – see Figure 9) appear more compatible. The analytical uncertainty for the Hg (and BaP) results has to be borne in mind when using the NBCs for these contaminants.

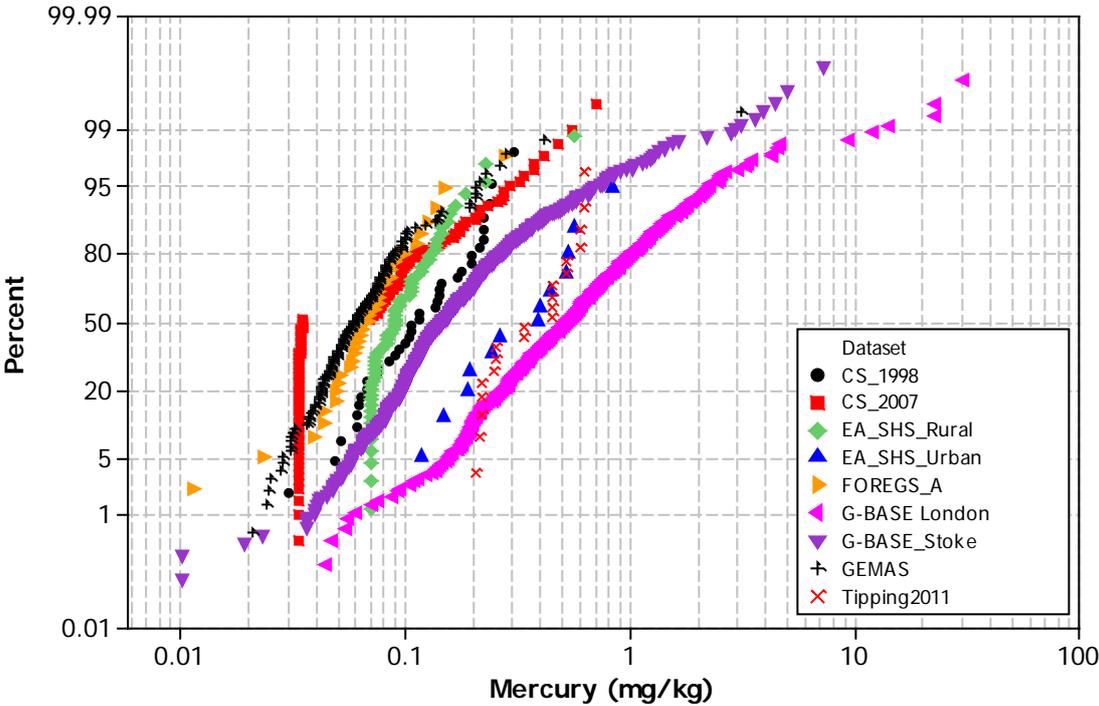


Figure 9: Mercury cumulative probability plot

5 Methodology for determining normal background concentrations (NBC)

The data exploration phase (WP2) prepares the contaminant data ready for the NBC calculation and identifies the areas where elevated background concentrations, through natural and diffuse anthropogenic processes, can be characterised as sufficiently different on a regional/national scale to warrant separate NBC determinations. For all the contaminants studied there has been at least one domain identified as having typically higher background concentrations, the area of England outside any defined domain is termed the Principal Domain. Therefore, there are always at least two domains to which NBCs have been attributed.

The sample points in any particular domain are captured in the GIS environment within the polygon(s) defining that domain to form a subset of the data ready for NBC determination.

Approaches to determining normal background are discussed in Cave *et al.* 2012. The starting point for this is the assumption that the contaminant data conform to a linear mixed model of the form:

$$Z = \text{fixed effects} + \text{continuous random variation} + \text{point contamination}$$

where Z is a random variable that represents the contaminant data derived from WP2. The fixed effects are sources of variation in the observed concentrations that are attributable to geogenic sources or diffuse anthropogenic activities. The identification of domains in WP2 should capture this, with the domain mean representing the fixed effect. The continuous random variation (typically Gaussian (normal) or log-Gaussian) represents the typical variation arising from geogenic or diffuse anthropogenic sources within the defined domains. The point contamination is assumed to introduce outlying values into the data. The equation above can be re-written informally as:

$$\text{concentration} = \text{domain average} + \begin{matrix} \text{typical variations in} \\ \text{geogenic and diffuse} \\ \text{sources within the} \\ \text{domain} \end{matrix} + \text{point contamination}$$

For any contaminant the first two terms (domain average + typical variations) give rise to the normal range of values or normal variation of the contaminant. The objective of the procedure is to characterise this normal variation in terms of a statistical distribution.

The Project has developed a robust statistical methodology for determining NBCs that is based on exploration of data distributions (by testing the distribution skewness) and applying data transformations (see Cave *et al.* 2012). The procedure is summarised in Figure 10a and Figure 10b, and the actual analysis is performed using the R programming language (R Development Core Team, 2011 – see Appendix 1 regarding availability of scripts) starting with domain data subsets for a particular contaminant input to the R programme as a comma separated variable file.

Percentile	Urban Domain (7,529)			Mineralisation Domain (347)			Principal Domain (34,257)		
	lower	middle	upper	lower	middle	upper	lower	middle	upper
50	170	170	180	300	330	370	43	43	44
55	190	190	200	340	380	430	46	47	47
60	210	220	220	380	440	490	50	50	51
65	240	250	250	440	500	570	55	55	55
70	280	280	290	510	580	660	60	60	61
75	320	320	330	590	680	780	67	67	68
80	370	380	390	700	820	950	76	76	77
85	440	450	460	850	1000	1200	89	90	91
90	550	570	580	1100	1300	1600	110	110	120
95	770	790	820	1600	1900	2400	170	170	180

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

Table 2: A summary of the lead domain percentile classifications (reported to 2 significant figures, all domains are based on the calculation of parametric percentiles). Domain NBCs shown in bold red. Concentrations in mg/kg.

The output from each NBC statistical analysis procedure is a table of percentile values (50th to 95th in steps of 5) calculated in three different ways (see Cave *et al.* 2012) and the methodology flow chart (Figure 10b) indicates which percentile calculation should be used. Associated with each percentile is an upper and lower confidence limit, this information is produced as a table in the supplementary information accompanying each contaminant TGS. An example of such a table (for Pb Domains) is shown in Table 2. Note that this procedure for determining domain NBCs is not an automated process and requires frequency distribution plots to be used in conjunction with the skewness coefficients to determine the appropriate data transformations.

Figure 10a (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. 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 (Figure 10b - Parts II – IV).

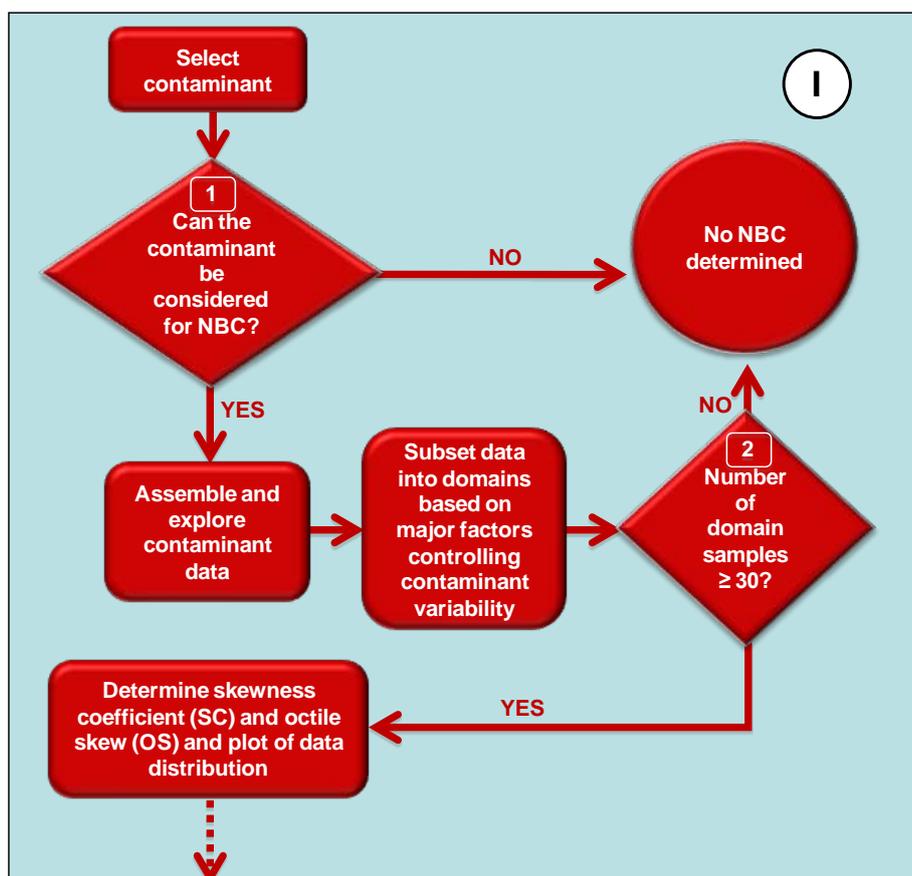


Figure 10:a Flow chart for the calculation of the NBC for a given contaminant – part I: Data gathering and exploration (see text for explanation)

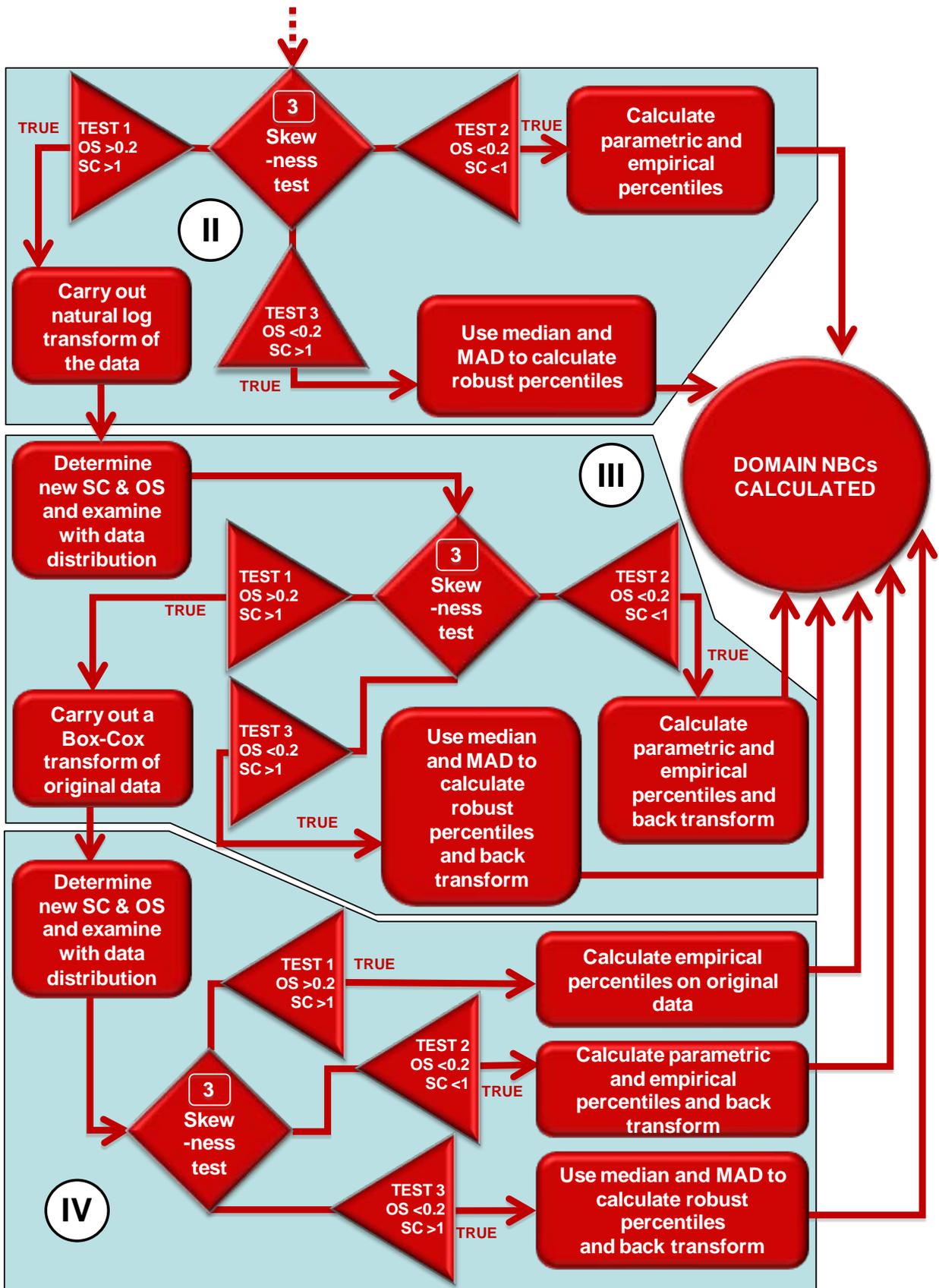


Figure 10b: Flow chart for the calculation of the NBC for a given contaminant – parts II - IV: Skewness testing and transformations (see text for explanation). MAD is the mean absolute deviation.

In Figure 10b, 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.

Having arrived at some robustly defined percentile values that have been derived taking into account any skewness or outlying data in the data set, a decision has to be made as to what result to use to represent the upper limit of normal background concentrations (see Cave *et al.* 2012, Sections 2.1.4 and 2.1.6).

This Project has defined the upper limit of normal background concentrations as the upper 95% confidence limit of the 95th percentile (taking into account data transformations). When defining the NBC we must consider what is typical and widespread (words used in the SG to describe “normal” levels – see Box 3). The median (50th percentile) is a measure of the central tendency, but on its own cannot be used to define what is typical or widespread as it only represents half the data. A percentile that encompasses most of the data is required and, although rather arbitrary in choice, the 95th captures a high proportion of the data and is likely to exclude the very high results that would be associated with point source data. It has also been used in other approaches to define background concentrations for environmental purposes (*e.g.* in Italy, APAT-ISS, 2006). Any uncertainty in the 95th percentile value is fully captured by taking the upper limit of the 95% confidence limit. In the determination of geochemical background, of great economic value in the search for ore bodies, the mean (\bar{x}) plus 2σ (σ = standard deviation) is a commonly used (Matschullat *et al.* 2000), which represents *c.* 97% of the data for normalised distributions.

6 NBCs for As, BaP, Cd, Cu, Hg, Ni and Pb

A summary of domain NBCs for the selected contaminants is given in Table 3. Those for As, Cd, Cu, Ni and Pb are derived from a very large data set, though when subdivided into domains, some domain NBCs are only based on a small number of samples (*e.g.* Ni Basic and Ultrabasic Domains which have a very small spatial extent). The importance of domains for which there are only a limited number of soil results can be confirmed using contaminant results for sample media other than soils, such as the BGS high density stream sediment sampling (Johnson *et al.*, 2005) or estimated derived soil values (Appleton *et al.*, 2008). The calculation of NBCs for these domains, and for BaP and Hg generally, can be improved with increased number of samples. These represent normal levels of contaminants at a national to regional scale. Within the defined domains there is still contaminant variability and, at a local scale it may be possible to define smaller more localised domains. Using the methodology described in Section 5, and providing there are sufficient systematically collected and analysed soil samples that are representative of an area, more localised contaminant NBCs could be calculated by those tasked with considering normal levels at a local scale.

As	DOMAIN				
	Principal	Mineralisation	Ironstone		
	NBC	32	290	220	
<i>N</i>	<i>41,509</i>	<i>187</i>	<i>437</i>		
BaP	DOMAIN (Great Britain)				
	Principal	Urban			
	NBC	0.5	3.6		
<i>N</i>	<i>371</i>	<i>32</i>			
Cd	DOMAIN				
	Principal	Min. Grp. 1	Min. Grp. 2	Urban	Chalk South
	NBC	1.0	17	2.9	2.1
<i>N</i>	<i>4,418</i>	<i>224</i>	<i>95</i>	<i>9,308</i>	<i>265</i>
Cu	DOMAIN				
	Principal	Mineralisation	Urban		
	NBC	62	340	190	
<i>N</i>	<i>34,504</i>	<i>153</i>	<i>7,475</i>		
Hg	DOMAIN				
	Principal	Urban			
	NBC	0.5	1.9		
<i>N</i>	<i>1,126</i>	<i>512</i>			
Ni	DOMAIN				
	Principal	Ironstone (Ni)	Peak District	Basic	Ultrabasic
	NBC	42	230	120	*
<i>N</i>	<i>41,768</i>	<i>117</i>	<i>221</i>	<i>23</i>	<i>4</i>
Pb	DOMAIN				
	Principal	Mineralisation	Urban		
	NBC	180	2,400	820	
<i>N</i>	<i>34,257</i>	<i>347</i>	<i>7,529</i>		

Table 3: Summary of domain normal background concentrations (NBCs) for the contaminants studied in this project. All concentrations in mg/kg, N is the number of samples used in the NBC calculation. The Ni Basic and Ultrabasic Domains are shown here, but as they are each defined by less than 30 samples, the NBCs are not calculated.

7 Concluding Remarks

1. There are large amounts of soil data covering England containing inorganic element contaminant results that has enabled this project to calculate NBCs for contaminant domains determined in the data exploration phase of the project. Gaps in knowledge relate mainly to organic contaminants, though the calculations of NBCs for these can still be made for England by utilising results from other parts of Great Britain, as has been done for BaP.
2. Once the contaminant data sets have been subdivided into domains, some areas are defined by a very small number of samples, particularly where only the lower density NSI soil results are available. Such domain NBCs could be improved by the collection of further samples (*e.g.* basic and ultrabasic rock areas for Ni, and Cu mineralised areas (Lake District)). However, the more pressing need for further information is likely to be driven by contaminants for which the geogenic and diffuse contributions are significant in terms of human health risks and these will primarily be in urban areas.
3. Where soil data are sparse, sample media other than soils, such as drainage sediments, can be used to help define the presence of high, widespread and typical levels of some inorganic contaminants in the surface environment. The BGS high density stream sediment sampling of Great Britain and Northern Ireland (Johnson *et al.*, 2005) will be completed by 2013 and can provide supplementary information to define contaminant domains for the areas not covered by the high density G-BASE soil sampling.
4. There are a lot of soil data generated by various sectors (especially universities and contaminated land workers) that is not captured to any central point. If the data has been collected and analysed to acceptable standards, and represents systematic rather than targeted sampling, it should be collated and made available so it can be used by others to improve assessments as to what normal levels of contaminants are. Those that are tasked with doing national surveys generating contaminant results for soil also need to make results more readily and freely available from a single national soil portal. The research value of many soil samples is greatly reduced because they are collected with restrictions on the site location resolution due to site access agreements.
5. Contaminant information for soils systematically collected across urban areas has provided extremely useful data for this work. Urban areas are those that have been most impacted by human activity leaving a potential legacy of contamination in areas where most of the population live and work. Many of the cities and towns of England still need to be systematically sampled.
6. The statistical methodology used in this work can be used by others either to determine NBCs not investigated here or calculated at a more local scale where there is systematically collected soil results of appropriate quality available. As more data becomes available this needs to be added to the knowledge base and NBCs recalculated to greater levels of confidence, particularly in those areas where there are knowledge gaps.
7. Reports and other resources from this project are made readily available from Defra and BGS web sites (see Appendix 1).

8. Normal background concentrations for the contaminant domains are our best effort to define what is the upper limit of “normal” levels of contaminants in soil as described by the Part 2A contaminated land Statutory Guidance. They are not a planning or risk assessment tool and must be used in the context of the SG in the manner described in the TGSs.
9. We define the NBC as the upper 95% confidence limit of the 95th percentile. Other percentiles and their confidence limits are listed in the TGS supplementary information, should others wish to consider our definition of “normal” levels in the context of other statistical information.

Appendix 1 : Project Outputs

Below project outputs and resources are listed with information about how and where they can be accessed

Project reports and associated outputs

Reports are available from the Defra project SP1008 web page¹ and include:

Data Exploration Report (BGS report No. CR/11/145); Methodology Report (BGS report No. CR/12/003); Final Project Report (BGS report No. CR/12/035); Cu, Ni, Cd and Hg data exploration (BGS report No. CR/12/041); and, Technical Guidance Sheets and supplementary information for As, BaP, Cd, Cu, Hg, Ni and Pb.

The BGS reports will also be made available through the NERC Open Research Archive ([NORA](#)).

Other project outputs and resources are available at:

www.bgs.ac.uk/gbase/NBCDefraProject.html .

These include:

- MS Access database with meta data on available soil data sets for England that have contaminant information;
- Project Bibliography (Endnote bibliography);
- R code scripts used to determine NBCs; and
- GIS Resources served as WMS files:
 - Domain polygons
 - Urbanisation index polygons defined from GLUD database
 - 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'.

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 data, this is made freely available subject to certain licensing terms and conditions. For large data sets there will also be a data handling fee. Enquiries should be sent to enquiries@bgs.ac.uk . The new NSI (XRFS) raw data has restricted terms of use until 2014.

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

1

<http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=17768&FromSearch=Y&Publisher=1&SearchText=sp1008&SortString=ProjectCode&SortOrder=Asc&Paging=10#Description>

Soil parent material

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

Land use data including non-ferrous 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](#). 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).

Abbreviations and Glossary

GLOSSARY

The terms used in this report (alphabetical order). Those marked with an ‘*’ are defined in the Statutory Guidance (Defra, 2012).

<i>anthropogenic</i>	having an origin associated with human activity
<i>background</i>	see "normal background"
<i>baseline</i>	in geochemistry used to describe the spatial distribution of an element. It is usually defined by extrapolating between sample sites to model the varying spatial distribution of a chemical element
<i>basic rock</i>	a low quartz (silica) igneous rock
<i>boxplot</i>	a graphical representation of a range of values where the length of the box represents the spread of values. Sometimes called a box-and-whisker plot as lines (whiskers) are added to the box to further demonstrate a range of results
<i>built environment</i>	human-made environment associated with activity from the scale of personal shelter and buildings to neighbourhoods and cities including their supporting infrastructure, such as transport networks
<i>contaminant*</i>	a substance relevant to the Part 2A regime (Defra, 2012) which is in, on or under the land and which has the potential to cause significant harm to a relevant receptor, or to cause significant pollution of controlled waters. Synonymous with terms "pollutant" and "substance"
<i>contaminated land*</i>	is any land which appears to the local authority in whose area it is situated to be in such a condition, by reason of substances in, on or under the land that – (a) significant harm is being caused or there is a significant possibility of such harm being caused; or (b) significant pollution of controlled waters is being caused, or there is a significant possibility of such pollution being caused
<i>domain</i>	a region defined by a boundary derived from a soil's underlying parent material, an urbanisation index, or an area of non-ferrous metalliferous mineralisation/mining and is used in this report to nationally characterise significantly different areas of NBCs
<i>diffuse pollution</i>	pollution from widespread human activities with no one discrete source. These activities may be recent or have been carried out in the past but cannot be tied down to a specific location or source. Examples of diffuse pollution include atmospheric deposition of contaminants arising from industry, domestic and industrial coal combustion and traffic exhaust, and disposal of domestic coal ash.

<i>Gaussian</i>	a term used in statistics, the Gaussian (or normal) distribution is a continuous probability distribution that has a bell-shaped probability density function, known informally as the bell curve
<i>geochemistry</i>	is the study of the distribution of chemical elements in and at the earth's surface
<i>geogenic</i>	pertaining to a geological origin
<i>Guidance*</i>	see Statutory Guidance
<i>interquartile range</i>	a statistical term being a measure of dispersion equal to the difference between the upper and lower quartiles
<i>mean</i>	a term used in statistics to quantify the average of a group of numbers determined by calculating the total of all the numbers and dividing this by the number of values used in the calculation
<i>median</i>	a term used in statistics for the middle result in a sorted list of results. There are therefore 50% of the results below this value
<i>normal*</i>	a term used to describe contaminant concentrations when they are seen as typical and encompasses contributions both from geogenic sources and diffuse anthropogenic pollution. Normal Background Concentrations (NBCs) are an expression of normal contaminant levels and should not be considered to cause land to qualify as contaminated land, unless there is a particular reason to consider otherwise
<i>Part 2A*</i>	means Part 2A of the Environmental Protection Act 1990 (as amended)
<i>pedological</i>	relating to soils and the processes associated with them
<i>percentile</i>	a statistical term based on ranking numbers in order and assigning a value based on below which a certain percentage of observations fall. It is therefore a non-parametric measure
<i>pollutant</i>	see "contaminant"
<i>quartile</i>	quartiles are a set of values being the three points that divide a data set into four equal groups, each representing a fourth of the population being sampled
<i>rural</i>	areas not categorised as industrial or urban. Predominantly agricultural land or undeveloped countryside. In this report quantitatively defined using an urbanisation index
<i>semi-urban</i>	A classification of land use defined in this report by the urbanisation index and is intermediary between urban and rural land use
<i>skewness</i>	a statistical term that is a measure of the asymmetry of the probability distribution of a population of results

- standard deviation** a statistical term that is a measure of how spread out numbers are. It is quantified by taking the square root of the variance
- Statutory Guidance*** is issued by the Secretary of State for Environment, Food and Rural Affairs in accordance with section 78YA of the Environmental Protection Act 1990 ("the 1990 Act"). It is intended to explain how local authorities should implement the regime, including how they should go about deciding whether land is contaminated land in the legal sense of the term.
- topsoil** is generally a sample of mineral soil collected from the top 30 cm of a soil profile and so represents a combination of both geogenic and anthropogenic processes that will influence the soil's chemistry
- ultrabasic rock** see basic rock. 'Ultra' implies that it is very basic
- urban area** from a physical scientist's perspective an urban area is a built environment (Johnson and Ander, 2008), *i.e.* an area with significant anthropogenic modification. The UK Soil and Herbage Pollutant Survey (Wood *et al.*, 2007), for example, defines urban as being an area which is $\geq 90\%$ built-up
- urbanisation** see "urban area"
- urbanisation index** in this report it is an estimate used to indicate the degree to which an area has been subjected to anthropogenic influence and is derived from the GLUD statistics and is the ratio of built space to open space using the Census Area Statistical Wards (CASW)

ABBREVIATIONS

- | | |
|--|--|
| BaP benzo[<i>a</i>]pyrene | GLUD Generalised Land Use Database |
| BGS British Geological Survey | IPR intellectual property rights |
| CEH Centre for Ecology and Hydrology | ISO International Standards Organisation |
| CS Countryside Survey | NAA neutron activation analysis |
| Defra Department for Environment, Food and Rural Affairs | NBC normal background concentration |
| DoE Department of Environment | NERC Natural Environment Research Council |
| EA Environment Agency | NSI National Soil Inventory |
| FOREGS FORum of European Geological Surveys | NSRI National Soil Resources Institute |
| G-BASE Geochemical BAseline Survey of the Environment | SG Statutory Guidance |
| GEMAS GEochemical Mapping of Agricultural and grazing Soils | SGV Soil Guideline Value |
| GIS geographical information system | SPMM Soil-Parent Material Model |
| | TGS Technical Guidance Sheet |
| | WP work package |
| | XRFS X-ray fluorescence spectrometry |

References

- APAT-ISS. 2006. Protocollo Operativo per la determinazione dei valori di fondo di metalli/metalloidi nei suoli dei siti d'interesse nazionale. Revisione 0. Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici and Istituto Superiore di Sanita.
- 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, EL, Cave, MR, Johnson, CC 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.
- Appleton, JD, Rawlins, BG 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.
- Barracough, D., 2007. UK Soil and Herbage Pollutant Survey. Introduction and Summary. Environment Agency (EA), UK.
- Cave, MR, Johnson, CC, Ander, EL 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. 41pp.
- Communities and Local Government. 2007. Generalised Land Use Database Statistics for England 2005. Department for Communities and Local Government. Product Code 06CSR04342. February 2007. Available online: <http://www.communities.gov.uk/publications/planningandbuilding/generalisedlanduse> last accessed 16th August 2012.
- Cousins, I.T., Kreibich, H., Hudson, L.E., Lead, W.A. and Jones, K.C., 1997. PAHs in soils: contemporary UK data and evidence for potential contamination problems caused by exposure of samples to laboratory air. *Science of The Total Environment*, 203(2), 141-156.
- 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.
- Defra. 2008. Guidance on the Legal Definition of Contaminated Land, Department for Environment, Food and Rural Affairs (Defra, UK)
- Defra, 2012. Environmental Protection Act 1990: Part 2A Contaminated Land Statutory Guidance. April 2012. Department for Environment, Food and Rural Affairs (Defra). HM Government. Available on-line at: <http://www.defra.gov.uk/environment/quality/land/> last accessed 16th August 2012.
- Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D.J., Brittain, S.A., Frogbrook, Z.L., Hughes, S., Lawlor, A.J., Poskitt, J.M., Potter, E., Robinson, D.A., Scott, A., Wood, C.M. and Woods, C. 2010. Countryside Survey. Soils Report from 2007. Centre for Ecology & Hydrology CS Technical Report No. 9/07. 194pp.
- Flight, D.M.A. and Scheib, A., 2011. Soil geochemical baselines in UK urban centres:: the G-BASE project. In: C.C. Johnson, A. Demetriades, J. Locutura and R.T. Ottesen (Editors), *Mapping the Chemical Environment of Urban Areas*. John Wiley & Sons Ltd., Chichester, UK, 186-206.
- 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.
- ISO 2011. Soil quality: Guidance on the determination of background values. ISO 19258:2011. International Organisation for Standardisation. 2011 revision of ISO 19258:2005.
- Johnson, C.C. and Ander, E.L. 2008. Urban Geochemical Mapping Studies: How and why we do them. *Environmental Geochemistry and Health*, 30, 511-530.
- 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.

- Lawley, R. 2011. The Soil-Parent Material Database: A User Guide. *British Geological Survey Open Report*, OR/08/034. 53pp. Available on-line at : <http://nora.nerc.ac.uk/8048/1/OR08034.pdf> last accessed 16th August 2012.
- Martin, I. and Cowie, C. 2008. Compilation of data for priority organic pollutants for derivation of Soil Guideline Values. Environment Agency, UK SC050021/SR7. 171pp.
- Matschullat, J., Ottenstein, R. and Reimann, C. 2000. Geochemical background - can we calculate it? *Environmental Geology*, 39, 990 - 1000.
- Montanarella, L., Tóth, G. and Jones, D.G., 2011. Soil component in the 2009 LUCAS Survey. In: G. Tóth and T. Németh (Editors), The international conference on 'Land Quality and Land Use Information in the European Union' 26-27 May 2011 in Keszthely, Hungary. European Commission, Joint Research Centre, Institute for Environment and Sustainability, Keszthely, Hungary, 209-218.
- Oliver, M.A., Loveland, P.J., Frogbrook, Z.L., Webster, R. and McGrath, S.P., 2002. Statistical and Geostatistical Analysis of the National Soil Inventory of England and Wales. Project SP0124, Defra (formerly MAFF) Soil Programme Technical Report.
- Ove Arup. 1990. Mining Instability in Britain. Department of Environment contract. Hardcopy maps digitised and adapted by the British Geological Survey.
- R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.r-project.org/> last accessed 16th August 2012.
- Reimann, C., Filzmoser, P., Garrett, R.G. and Dutter, R., 2008. *Statistical Data Analysis Explained. Applied Environmental Statistics with R*. Wiley.
- Reimann, C., de Caritat, P., GEMAS Project Team and NGS 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., Batista, M. J., Bidovec, M., Demetriades, A., De Vivo, B., De Vos, W., Duris, M., Gilucis, A., Gregorauskiene, V., Halamic, J., Heitzman, P., Lima, A., Jordan, G., Klaver, G., Klein, P., Lis, J., Locotura, J., Marsina, K., Mazreku, A., O'Connor, P. J., Olsson, S. A., Ottesen, T., Petersell, V., Plant, J. A., Reeder, S., Salpeteur, I., Sandstrom, H., Siewers, U., Steenfelt, A., and Tarvainen, T., 2005. *Geochemical atlas of Europe. Part 1, Background information, methodology and maps*. Geological Survey of Finland, Espoo, Finland, 526pp.
- Tipping, E., Poskitt, J.M., Lawlor, A.J., Wadsworth, R.A., Norris, D.A. and Hall, J.R., 2011. Mercury in United Kingdom topsoils; concentrations, pools, and Critical Limit exceedances. *Environmental Pollution*, 159, 3721 - 3729.
- Wood, M.D., Copplestone, D. and Crook, P. 2007. UK Soil and Herbage Pollutant Survey (UKSHS). Chemical and radiometric sample collection methods. Report No.2, UK Environment Agency. UK