

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

# Science of the Total Environment

journal homepage: [www.elsevier.com/locate/scitotenv](http://www.elsevier.com/locate/scitotenv)

## Methodology for the determination of normal background concentrations of contaminants in English soil

E. Louise Ander <sup>a,\*</sup>, Christopher C. Johnson <sup>a</sup>, Mark R. Cave <sup>a</sup>, Barbara Palumbo-Roe <sup>a</sup>,  
C. Paul Nathanail <sup>b</sup>, R. Murray Lark <sup>a</sup>

<sup>a</sup> British Geological Survey, Environmental Science Centre, Keyworth, Nottingham NG12 5GG, UK

<sup>b</sup> School of Geography, University of Nottingham, Nottingham NG7 2RD, UK

### HIGHLIGHTS

- Contaminated land Statutory Guidance refers to normal levels of soil contaminants.
- We represent normal levels by Normal Background Concentrations (NBCs).
- We explore the most suitable systematically sampled soil data for English soils.
- NBCs for soil contaminants presented not as a single national value but by domains.
- We present a robust statistical methodology for determining domain NBCs.

### ARTICLE INFO

#### Article history:

Received 4 October 2012

Received in revised form 24 January 2013

Accepted 1 March 2013

Available online 10 April 2013

#### Keywords:

Soil

Contaminated land

Background

Contaminant

Normal concentration

### ABSTRACT

The revised Environmental Protection Act Part 2A contaminated land Statutory Guidance (England and Wales) makes reference to 'normal' levels of contaminants in soil. The British Geological Survey has been commissioned by the United Kingdom Department for Environment, Food and Rural Affairs (Defra) to estimate contaminant levels in soil and to define what is meant by 'normal' for English soil. The Guidance states that 'normal' levels of contaminants are typical and widespread and arise from a combination of both natural and diffuse pollution contributions. Available systematically collected soil data sets for England are explored for inorganic contaminants (As, Cd, Cu, Hg, Ni and Pb) and benzo[*a*]pyrene (BaP). Spatial variability of contaminants is studied in the context of the underlying parent material, metalliferous mineralisation and associated mining activities, and the built (urban) environment, the latter being indicative of human activities such as industry and transportation. The most significant areas of elevated contaminant concentrations are identified as contaminant domains. Therefore, rather than estimating a single national contaminant range of concentrations, we assign an upper threshold value to contaminant domains. Our representation of this threshold is a Normal Background Concentration (NBC) defined as the upper 95% confidence limit of the 95th percentile for the soil results associated with a particular domain. Concentrations of a contaminant are considered to be typical and widespread for the identified contaminant domain up to (and including) the calculated NBC. A robust statistical methodology for determining NBCs is presented using inspection of data distribution plots and skewness testing, followed by an appropriate data transformation in order to reduce the effects of point source contamination.

© 2013 Natural Environment Research Council. Published by Elsevier B.V. All rights reserved.

### 1. Introduction

The work described here is part of the process to simplify the contaminated land regime for England and Wales (Defra, 2011). Part 2A of the Environmental Protection Act 1990, as amended in 2012, defines contaminated land that poses a Significant Possibility of Significant Harm (SPOSH). The United Kingdom Secretary of State for the environment issues Statutory Guidance (SG) in accordance

with section 78Y of the Environmental Protection Act 1990 to establish a legal framework for dealing with contaminated land (DETR, 2000). Revised SG was issued in April 2012 (Defra, 2012a). The intent of the SG is to explain how the contaminated land regime should be implemented. However, the original SG and a previous update to the SG (DETR, 2000; Defra, 2006), which were supposed to explain when land does (and does not) need to be remediated, did not resolve significant uncertainties. Contaminated land was still, in part, defined on the basis of the toxicologically vague concept of 'unacceptable intake'. Therefore, following a year of consultation, the SG was revised to be more useable for those working with contaminated

\* Corresponding author.

E-mail address: [land@bgs.ac.uk](mailto:land@bgs.ac.uk) (E.L. Ander).

land and remediation. Land is placed into one of four new categories to help decide when land is (and is not) contaminated (Fig. 1). Category 1 describes land which is clearly posing SPOSH, for example, because similar sites are known to have caused a significant problem in the past. Category 4 describes land that is clearly not posing SPOSH and should not be determined as statutory contaminated land. Categories 2 and 3 are land for which categorisation is less straightforward. Category 2 land poses SPOSH and is defined by a combination of expert opinion and scientific evidence that analogous conditions on other sites have or are highly likely to have caused significant harm. Category 3 land is that where the conditions for Category 2 are not met and the land is not contaminated. There is a fifth category not shown in Fig. 1, informally termed Category 0, which is land where significant harm to human health has occurred as a direct result of contamination, a category for which there is no example in England, Scotland or Wales.

The revised SG (Defra, 2012a) introduced the concept of 'normal' levels of contaminants in soil. Section 3 of the Guidance notes that 'normal' presence/levels of contaminants:

- should not be considered to cause land to qualify as contaminated land, unless there is a particular reason to consider otherwise (Section 3.22);
- may result from the natural presence of contaminants at levels that might be considered typical in a given area, and have not been shown to pose an unacceptable risk to health or the environment (Section 3.23(a)); and
- are caused by low level diffuse pollution, and common human activity other than specific industrial processes (Section 3.23(b)).

As part of its research programme to support Local Authorities implement the revised SG, a research project was instigated by Defra to investigate normal concentrations of contaminants in the soil of England. The question of "what is normal?" should be a prerequisite in any investigation of potentially contaminated soil, and the results of our research are just one of the several tools to be provided by Defra to support the revised SG.

It is important to emphasise that this work to determine 'normal' levels of contaminants in soil supports the English contaminated land regime rather than the planning regime. The contaminated land regime requires developers to show that land is safe, suitable for use and, after remediation, cannot be determined as statutory contaminated land. The regimes differ in the level of risk at which remediation is needed. In the case of planning, remediation is needed to ensure a site is suitable for its future intended use. Under the contaminated land regime, remediation is needed if the site, given its current use, is presenting such a high level of risk that if nothing is done, there is a significant possibility of significant harm such as death, disease or serious injury.

The research comprised four work packages with the aim of defining 'normal' level of contaminants in soil by determining Normal Background Concentrations (NBCs) for a selected number of contaminants:

- Assessment of available soil contaminant data for English soil;
- Exploration of relevant soil data sets;
- Development of a robust statistical methodology for determining NBCs; and
- Technical Guidance Sheets for the use of NBCs for selected contaminants.

**Diagram showing the new Category 1-4 system (compared to current situation)**

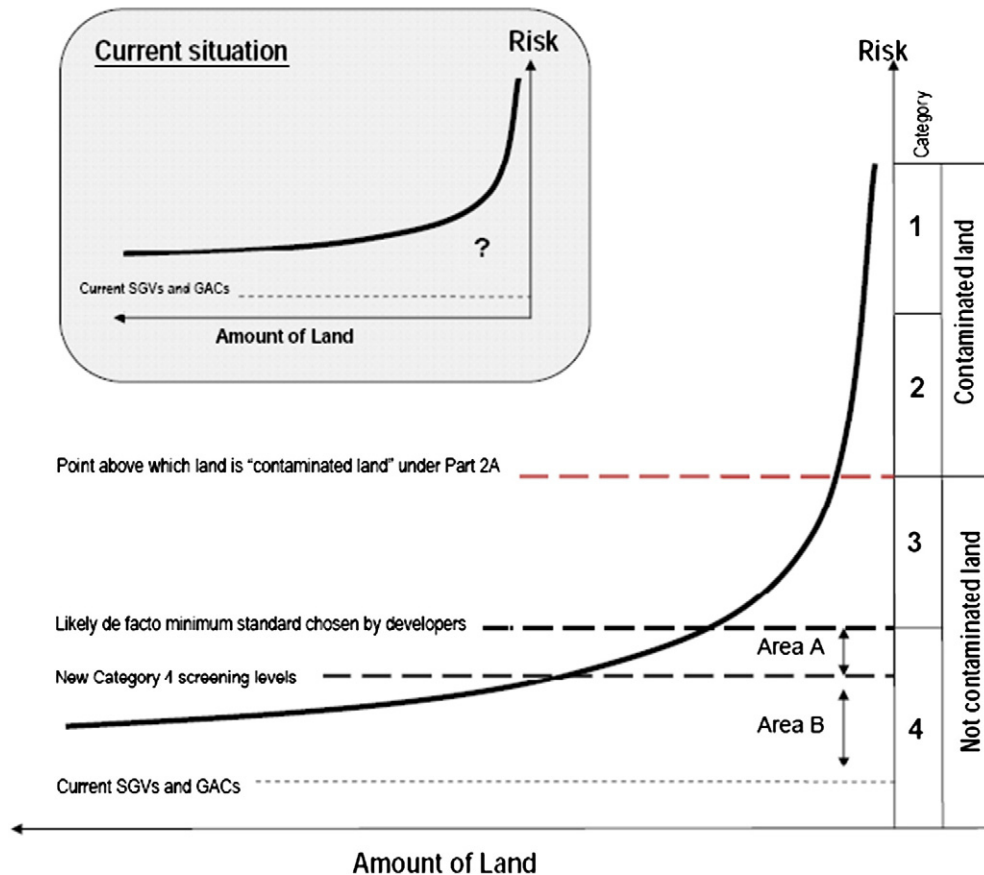


Fig. 1. The new four category system to help test when land is, and is not contaminated (from Defra, 2011).

This paper describes the first three of these work packages reported in more detail in technical reports (Ander et al., 2011, 2012; Cave et al., 2012; Johnson et al., 2012).

## 2. Contaminants

A detailed review of how a contaminant is defined is beyond the scope of this work which focuses on contaminants as defined in British legislation. A contaminant can be defined in many ways, here we use the term as defined in the Part 2A contaminated land Statutory Guidance (SG) (Defra, 2012a, page 4). In this the terms 'contaminant', 'pollutant' and 'substance' are used with the same meaning, that is, "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". Detailed clarification of what is meant as 'significant harm' and 'relevant receptor' is also given in the SG. Receptors to be protected from harm fall into three categories: human, ecological systems and property (e.g., buildings, crops and livestock).

Literature covering aspects of contamination give less robust definitions of a contaminant. For example, BS EN ISO 19258:2011 (BSI, 2011) defines a contaminant as "a substance or agent present in the soil as a result of human activity", and notes that there is no assumption in this definition that harm results from the presence of the contaminant. This definition does not include elements and substances derived purely from geological sources that are defined as contaminants but occur naturally in widespread and high concentrations (i.e. naturally occurring and contaminants). Cole and Jeffries (2009) in their report on using soil guideline values (SGVs) say "a contaminant is a substance that is in, on or under the land and has the potential to cause harm".

There are thousands of potential contaminants which might be present on various sites around England (although a smaller sub-set probably drives the risk on most sites) (Defra, 2008). The now withdrawn report on "Potential Contaminants for the Assessment of Land" (Defra and EA, 2002) identified the priority chemicals for the development of SGVs. This was based on the chemicals likely to be present in sufficient concentrations on affected sites that were considered to pose a risk. An updated priority chemical list is presented by Martin and Cowie (2008) (Table 1). This comprises fifty six chemicals, fourteen of which are chemical elements plus cyanides (an inorganic substance) and asbestos (mineralogically defined). The remaining contaminants can be classified as organic substances, and these in

the soil environment will be overwhelmingly (though not exclusively) associated with anthropogenic activity (radioactive elements are outside the remit of this investigation). Globally, there is generally good agreement as to what are priority contaminants, though there are national differences. In Finland, for example, the "Government Decree on the Assessment of Soil Contamination and Remediation" (Finnish Government, 2007) lists eleven inorganic elements: antimony (Sb), arsenic (As), mercury (Hg), cadmium (Cd), cobalt (Co), chromium (Cr), copper (Cu), lead (Pb), nickel (Ni), vanadium (V) and zinc (Zn). Two of these elements, Sb and Co, are not present in Table 1.

In our data exploration and methodology development, eight contaminants – As, Cd, Cu, Hg, Ni, Pb, benzo(a)pyrene (BaP) and asbestos – were selected to represent the range of data availability and contributing sources, though the exploration shows that asbestos is a contaminant for which a NBC cannot be determined. The remaining substances are discussed in subsequent sections.

## 3. Background concentrations

A number of terms are used to convey the expected concentrations of a contaminant in soil. These include: normal, typical, baseline, ambient, characteristic, natural, background and widespread. There are some subtle differences between these terms, they can mean different things in different disciplines, and they can be confused with alternative uses. For example, a statistician would associate the word 'normal', in the context of defining the spread of a set of results, with the normal (or Gaussian) distribution. The terms normal, typical, characteristic and widespread are more or less synonymous. In this paper, we use 'normal' in the sense of the SG (Defra, 2012a) as set out here in Section 1.

In our work, we represent what is normal by use of Normal Background Concentrations (NBCs), so the term normal background excludes point contamination and encompasses typical variation around a domain average (Fig. 2).

The term background has a more complex and varied usage than the term normal and is discussed in detail by Matschullat et al. (2000); Reimann and Garrett (2005); and Reimann et al. (2005). It is used differently in different areas of science, for example:

- in exploration geochemistry the term background has been long-established and defines an area of normal element concentrations distinguished from anomalously high concentrations (that may indicate the presence of metalliferous mineralisation) by a threshold value;
- in environmental geochemistry background is a relative measure to distinguish between natural element or compound concentrations and anthropogenically-influenced concentrations in real sample collectives (Matschullat et al., 2000); and
- in the BSI (2011) guidance on soil background the content of a substance in a soil results from both natural geological and pedological processes and includes diffuse source inputs.

In the work described here, where the definition of normal is given in statutory guidance, the use of the term normal background is as defined by BS EN ISO 19258:2011 (BSI, 2011) and includes both geogenic (i.e., of geological origin) and anthropogenic diffuse pollution.

A further refinement in our use of NBCs is the realisation that, due to significant chemical variability in the underlying parent material on which soil is formed, different areas of the country will have different ranges of normal background concentrations. This will be particularly the case in areas overlying mineralisation or those regions subjected to a long history of anthropogenic activity, namely built-up or urban areas. Rather than having single national NBCs, we have used our data exploration of the different contaminants to identify areas, referred to as domains, which have characteristic NBCs.

The conceptual model that we follow for defining a contaminant concentration is summarised in Fig. 2. The concentration of a contaminant in

**Table 1**  
List of priority contaminants (from Martin and Cowie, 2008).

Inorganic	Organic	
Arsenic (As)	Acetone	Fenitrothion
Beryllium (Be)	Aldrin	Hexachlorobuta-1,3-diene
Cadmium (Cd)	Atrazine	Hexachlorocyclohexanes
Chromium (Cr)	Azinphos-methyl	Malathion
Copper (Cu)	Benzene	Naphthalene
Lead (Pb)	Benzo(a)pyrene (BaP)	Organolead compounds
Mercury (Hg)	Carbon disulphide	Organotin compounds
Molybdenum (Mo)	Carbon tetrachloride	Pentachlorophenol
Nickel (Ni)	Chloroform	Phenol
Selenium (Se)	Chlorobenzenes	Polychlorinated biphenyls (PCB)
Sulphur (S)	Chlorophenols	Polycyclic aromatic hydrocarbons (PAH)
Thallium (Tl)	Chlorotoluenes	Tetrachloroethane
Vanadium (V)	1,2-Dichloroethane	Tetrachloroethene
Zinc (Zn)	Dichlorvos	Toluene
Cyanide	DDT	Total petroleum hydrocarbons
Asbestos	Dieldrin	2,1,1,1-Trichloroethane
	Dioxins and furans	Trichloroethene
	Endosulfan	Trifluralin
	Ethylbenzene	Vinyl chloride
	Explosives	Xylenes

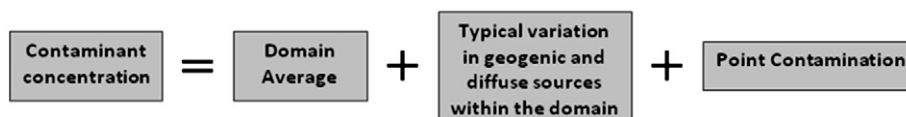


Fig. 2. Conceptual model of contaminant concentration.

soil varies about the domain mean due to geogenic and diffuse sources and, where present, point contamination. In the absence of point contamination, observed concentrations in soil within a domain represent the normal levels of concentration. Our aim is to identify contaminant domains and, to characterise the statistical distribution of NBCs with a statistical methodology that is robust to the effects of any point source contamination.

Methodologies for determining background concentrations are as numerous as there are definitions for background. There are important strategic, economic and legislative drivers for understanding and quantifying soil element/contaminant concentrations. From an economic and strategic point of view, the exploration and development of economic metalliferous mineral deposits by geochemical exploration have meant that much work has been done into developing methods to determine background levels. For more than sixty years, statistical methods have been used to distinguish between anomalous and background concentrations of the chemical elements in soil in order to locate buried mineralisation (e.g., Lovering et al., 1950; Hawkes and Bloom, 1955; Tennant and White, 1959; Lepeltier, 1969; Tidball et al., 1974; Sinclair, 1976, 1983, 1986). Some of these techniques are described in detail by Matschullat et al. (2000) with application examples and include: the Lepeltier method; relative cumulative frequency curves; normality of sample ranges; regression techniques; mode analysis; 4- $\sigma$  outlier test; interactive 2- $\sigma$  technique; and calculated distribution function.

Grunsky (2010), in a more recent review of interpreting geochemical survey data, discusses graphical methods for differentiating geochemical background from anomalies.

Legislation concerned with healthy and sustainable environments is also now a significant driver for information on background contaminant concentrations. For this purpose, documents such as British Standards Institution guidance on the determination of background values have been published. BS EN ISO 19258:2011 (BSI, 2011) covers the prerequisites of sampling, analysis and data handling and outlines some essentials of statistical evaluation of data.

Matschullat et al. (2000), Grunsky (2010) and BS EN ISO 19258:2011 illustrate the fact that the determination of background requires good quality concentration data and a statistical methodology to deliver estimates for background concentrations. The availability and robustness of available data sets for contaminant concentrations in English soil were the starting point for our work and are described in detail in Ander et al. (2011) and summarised here in Sections 4 and 5.

The majority of research to date has focused on methods for providing typical background concentrations of Potentially Harmful Elements (PHEs) in soil (e.g., Appleton, 1995; Appleton et al., 2008). Geochemists express the geochemical baseline (a spatially fluctuating chemical environment at a given point in time) in terms of the natural baseline. This is stable over a long period of time, provided that there are no sudden catastrophic events (such as flooding and marine transgression), and will be associated with an overprint of the anthropogenic baseline (with many contributing sources) that changes over a relatively short period of time (Johnson and Ander, 2008). An understanding of what constitutes the natural baseline enables the contribution of the anthropogenic component to be estimated. These approaches to determining 'backgrounds' are largely based on soil sampling and analyses over different parent material groups, which have been shown to exert a dominant control on topsoil chemistry in England (Rawlins et al., 2003).

Alternative approaches have been investigated, based on associations with particle size fractions across England and Wales (Zhao et al., 2007) or globally based on statistical relationships with total soil iron or manganese (Hamon et al., 2004). The approach proposed by Appleton et al. (2008) is to estimate typical background concentrations from a statistical measure (e.g., the geometric mean) based on existing soil analyses within soil parent material polygons. The background concentrations for a particular PHE are mapped using delineations of the parent material polygons. Oliver et al. (2002) present a very thorough statistical analysis of the National Soil Inventory (NSI) results for England and Wales, and such work gives an appreciation of the concentration ranges for complete data sets.

In the United Kingdom, there are examples of previous work to define 'average' trace element concentrations in soil. Archer and Hodgson (1987) report normal ranges – values between twice the log-derived standard deviation above and below the mean (i.e., approximately 95% of the data) for soil in England and Wales. Paterson et al. (2003) describe background levels of contaminants in Scottish soil, citing minimum and maximum trace element ranges with the median and lower and upper quartiles. The fundamental weakness of all these previous studies is that the results cover mainly agricultural or rural soil, and have no results for soil from urban areas. These use a simple statistical approach, essentially to define data ranges.

A review of worldwide national approaches to determining background concentrations was outside the scope of this work, but it is worth noting some examples here from Italy and Finland. In Italy, APAT-ISS (2006) gives government guidance for the determination of background values of metals and metalloids in Italian soil. This national guidance uses the BSI (2011) definition of natural background concentration. The stepwise approach for deriving background values involves the collection of data, the statistical analysis of the data and the determination of the background value. The selection of the sampling sites follows the typological approach (based on parent material, soil type and land use), choosing sites within homogeneous areas. The statistical analysis is carried out on data sets, each representative of homogeneous typologies. The descriptive statistics for data distribution include the minimum, maximum, median, percentile, standard deviation, skewness, kurtosis and graphic representations, such as boxplots, histograms and percentage cumulative frequency plots. The guideline describes in detail a series of statistical tests to identify outliers and to define the distribution type of the data (normal, lognormal, gamma, non-parametric distribution). The background value is defined as the 95th percentile of the population. In Finland, a Government Decree on the Assessment of Soil Contamination and Remediation Needs (214/2007) (Finnish Government Decree, 2007) became legislation on 1 June 2007 (Tarvainen and Jarva, 2011). The decree defines a geochemical baseline as being the natural geochemical background concentration and superimposed diffuse anthropogenic input of elements in the topsoil. Backgrounds are assessed on a local investigation of the geochemical baseline rather than on national values, and the upper limit of geochemical baseline variation for element X ( $BL_X$ ) is estimated as follows:

$$BL_X = P_{75} + 1.5(P_{75} - P_{25})$$

where  $P_{75}$  is the 75th percentile of element X concentrations and  $P_{25}$  is the 25th percentile of element X concentrations.

An important point made by many of the accounts looking at methodologies for background determinations (e.g., Matschullat et al., 2000; Reimann and Garrett, 2005; Tarvainen and Jarva, 2011) is that estimations are very dependent on location and scale. The domain approach and spatial exploration of the contaminant data, as described in Section 5, are important parts in understanding normal levels of contaminant concentrations in soil. The approach used is dependent on the data sets being able to capture contaminant variability at the regional scale at which domains are defined, both in rural (Johnson et al., 2005) and urban (Fordyce et al., 2005) areas. This is discussed in the following section.

#### 4. Available data sets

The distribution of chemical elements within soil is of interest to many scientific disciplines. Soil science is concerned with soil as a natural resource and in its management. Its chemical, physical and biological properties need to be mapped and understood. Geochemists are also interested in mapping the behaviour and distribution of chemical elements at the Earth's surface for a variety of sample types, and soil is the most ubiquitous material for this purpose. Indeed, as with many scientific challenges in the management and exploitation of resources, and concerns regarding the environmental and health impacts of changes to the chemical surface environment, many of the traditional sciences now work together under the umbrella of environmental science. As a result, reviewing the existing soil chemical data for England has involved a broad cross-section of scientific disciplines. Furthermore, the investigation of available data has not just been restricted to England. In instances where contaminant information is very limited, for example the organic contaminants, data sets covering areas outside England have been used to supplement sparse information.

Information on the soil data sets for English soil with chemical results (origin, methods, date of sampling, quality control, etc.) has been stored in a relational MS Access database and summarised in the appendix of Ander et al. (2011). Different data sets are more appropriate than others for particular contaminants, and others, whilst not spatially extensive, may provide valuable data that the larger data sets do not contain. The most useful data sets are those that:

- Include results for priority contaminants by analytical methods with suitable lower limits of quantification;
- Are associated with a systematic rather than a targeted sampling strategy so as to represent a broad range of land use types;
- Are spatially extensive across England with a good sample density;
- Are soil samples that have been collected and analysed to internationally recognised standards and have associated quality assurance;
- Unambiguously define total concentrations of contaminants;
- Are compatible with other available data sets; and
- Provide good resolution of the sample site coordinates.

Such data sets are generally those that have been generated by national/international baseline surveys, conducted by publicly funded organisations producing unbiased data.

##### 4.1. Primary soil chemical data for England

Three primary data sets satisfy the majority of criteria listed above, and are the best data available for exploring most of the inorganic contaminants. These are the G-BASE rural, G-BASE urban and NSI (XRFS) topsoil data sets (Fig. 3). The Geochemical Baseline Survey of the Environment (G-BASE) is the British Geological Survey's (BGS) systematic geochemical baseline programme (Johnson et al., 2005; Fordyce et al., 2005; Flight and Scheib, 2011). The National Soil Inventory (NSI) was part of the National Soil Map Project (1978–1983) (Oliver et al., 2002), and samples have been analysed by several analytical methods yielding two geochemical atlases (McGrath and

Loveland, 1992; Rawlins et al., 2012). The NSI samples currently come under the custodianship of the National Soil Resources Institute, Cranfield University, UK.

These soil data sets use topsoil (c. 0–15 cm) so as to be representative of both the anthropogenic and parent material contribution of the contaminant. Surface soil (0–2 cm), whilst being the most useful for health risk assessments and representative of airborne contributions to the soil (e.g. Ottesen et al., 2008) is likely to contain a greater proportion of organic litter. The surface soils may be better for targeted surveys of specific land uses (e.g. playground or forestry soils) but they suffer from problems of representativity (sampling depths tend to be indicative) and are not widely used for systematic surveys (Johnson et al., 2011). Deeper soil (>30 cm), whilst capturing some historical anthropogenic contamination, is likely to over-represent the parent material contribution. G-BASE and NSI samples are composite samples, made up from sub-samples (5 and 25 sub-samples, respectively) collected from a 20-m square using a soil auger.

The G-BASE rural and urban data are sampled in a consistent manner, and have been determined for some fifty chemical elements by laboratory-based X-ray fluorescence spectrometry (XRFS), including contaminants As, Cd, Cr, Cu, Pb, Mo, Ni, Se, S, Ti, V and Zn. The high density of sampling (one site every two kilometre squares for rural and four sites every kilometre square in urban areas), enables interpretations to be made down to a local area scale. The combined G-BASE rural and urban data (c. 47,000 topsoil samples) gives a data set of ten orders of magnitude bigger than the next largest data set (NSI). G-BASE is the only programme to have systematically mapped the chemical baseline of urban areas, and the 'London Earth' sub-project of G-BASE has provided chemical information for the capital city, representing the largest urban geochemical mapping project in the world. The G-BASE soil baseline for all of England currently covers mainly central and eastern England (see Fig. 3). However, the

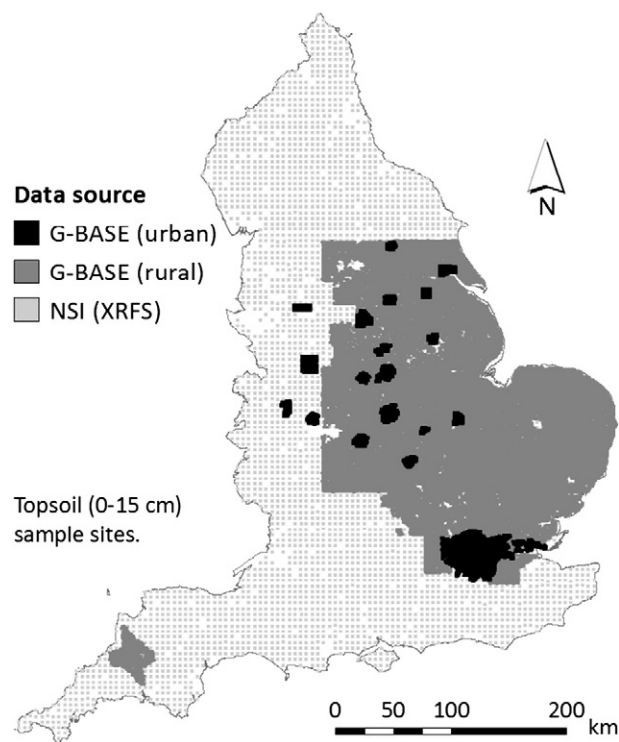


Fig. 3. Map showing the distribution of samples used in the NBC data exploration for As, Cd, Cu, Ni and Pb. NSI (XRFS) covers the whole of England at a sample density of 1:25 km<sup>2</sup>. G-BASE sampling densities for rural and urban are 1:2 km<sup>2</sup> and 4:1 km<sup>2</sup>, respectively.

areas not yet sampled by G-BASE can be supplemented by the NSI (XRFS) data set for which topsoil samples, collected and prepared in a similar way to the G-BASE project, have been reanalysed at the BGS XRFS laboratories to give total element concentrations. The earlier NSI data only contained a limited number of element results, following determination by ICP-AES using an *aqua regia* extraction. The NSI data has a sampling density that approximates to 1 sample per 25 km<sup>2</sup>.

#### 4.2. Supplementary soil chemical data for England

There are other important data sets that usefully augment the G-BASE and NSI data, although collected at much lower sampling densities. The EuroGeoSurveys' Geochemical Mapping of Agricultural Soil (GEMAS – Reimann et al., 2012) and European geochemical atlas (FOREGS (Forum of European Geological Surveys) – Salminen et al., 2005; De Vos and Tarvainen, 2006), the Environment Agency's UK Soil and Heritage Survey (UKSHS – Barraclough, 2007), and the Centre for Ecology & Hydrology (CEH) Countryside Survey (CS – Emmett et al., 2010) provide additional contaminant data for Hg and organic substances. However, there are issues regarding the use of these data sets. For example, some only determine elements following an *aqua regia* extraction and so do not unambiguously represent total contaminant concentrations. Additionally, they also tend to have a much reduced sampling density and thus will fail to capture local, and some regional, variability. The use of the UKSHS and Countryside Survey 2000 data for this purpose is also impeded by the fact that site coordinates are degraded to the nearest 10 km in order to satisfy land access agreements.

Some other significant data sets are inappropriate to use as they target a specific land use or land group and would, therefore, bias any NBCs towards that particular land use. The largest of these is the BGS Mineral Reconnaissance Programme (MRP) soil analyses. As these samples were collected in predominantly metalliferous mineralised areas, often associated with a long legacy of mining, sampling strategies were geared towards finding high results for metals. The MRP is also an example of a programme for which there was great variability in the sampling and analytical methodology used, so the data set cannot be interpreted as a single entity. Site investigations targeting contaminated land will similarly produce data that cannot be used to establish normal backgrounds. Results will predominantly be for contaminated soil, which is what is required when investigating such a site, but is not good for establishing local or regional backgrounds. Projects, specific to a particular land use, and targeting the humus layer rather than mineral soil, are also of limited value to the project. A big Europe-wide project – LUCAS (Land Use Coverage and Area frame Survey) (Joint Research Centre (JRC) laboratory of the European Commission) – is currently in progress (Montanarella et al., 2011) with some 1373 sample sites in the UK. Heavy metal analysis of top soil is proposed, but not yet completed. Land access agreements may also prevent site coordinates being readily available when this project delivers some data.

Finally, an important source of soil results, particularly for Hg and organic contaminants is contained in peer-reviewed publications, e.g., Tipping et al. (2011), Cousins et al. (1997), and Jones et al. (1989). Jones et al. is an example of a paper containing original data, with site coordinates, but for Wales rather than England, data which can be extrapolated to supplement sparse BaP information for English soil. However, many publications contain just summary tables without site locations, and care has to be taken to note how much of the data is original or compiled from other publications.

Although the estimation of NBCs is concerned with soil, other sample media are used to define the surface chemical environment and so can be used, particularly wherever there is no soil data, to define areas where contaminant concentrations are particularly high. The use of fine stream sediment, collected from small (low order) streams, is the media of preference for defining the regional geochemical baseline by geochemists (Johnson et al., 2008). Appleton et al.

(2008) have used the soil-stream sediment relationship to estimate national-scale potentially harmful element ambient background concentrations. For England, there is good high density stream sediment data available from the BGS G-BASE project and the Wolfson Geochemical Atlas (Appleton et al., 2008).

## 5. Data exploration and definition of domains

Prior to developing a methodology for determining NBCs, the spatial distribution of the selected contaminants was explored in order to understand the spatial variability and the most important domains (Ander et al., 2011, 2012). Of the initial eight contaminants selected for study, the absence of any systematic information on the background distribution of asbestos in soil meant that no data exploration was possible for this contaminant, and this is an example of a contaminant for which a NBC could not be determined. This is a reflection of the difficulty in applying a common methodology to the abundance of minerals in soils and the concentration of chemical substances. The data exploration contained two components, firstly exploration of the spatial distribution of contaminants in topsoil across England and, secondly, exploring various landscape data sets with which to define contaminant domains.

### 5.1. Spatial distribution

Where there is sufficient contaminant data, national maps can be plotted by interpolating between sample sites to give geochemical images, such as that shown in Fig. 4. These maps (for As, Cd, Cu, Ni and Pb) are published in the contaminant NBC Technical Guidance Sheets, supplementary information (e.g., Defra, 2012b). Further mapping, using other statistical techniques, such as k-means cluster analysis (see Ander et al., 2011), in addition to these maps, gives an excellent visualisation of a contaminant's variability in soils across England. The variability can be attributed to three main factors:

1. The single most important controlling factor is the underlying parent material (geology), which provides the geogenic component of 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 elements that are defined as contaminants and occur naturally.
2. A further geogenic component is when a contaminant is enriched in soil, 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.

These form the basis for the landscape data used to define contaminant domains. In a GIS environment (Arc GIS v9.2), this is integrated with the spatial distribution maps so as to recognise domains associated with elevated contaminant concentrations in soil. Domains are defined as polygons and the contaminant data captured within the domain polygons. 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. Domains identified for the contaminants investigated are shown in Figs. 5 and 6. These domains along with summary contaminant information are listed in Table 2. Our objective has been to identify the most significant domains on a national and regional scale. At a local scale, where higher density sampling provides more detailed information, further domains could undoubtedly be identified, say for example, within an

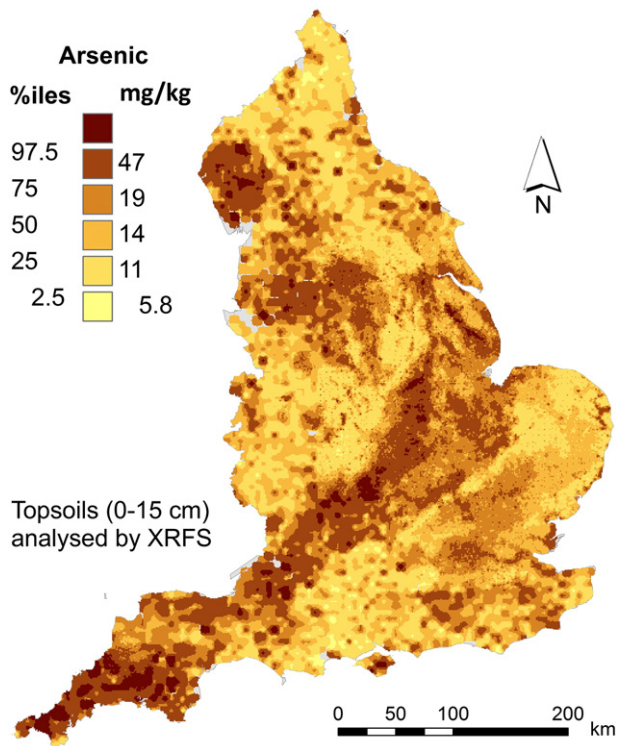


Fig. 4. Example of an interpolated geochemical image used to demonstrate contaminant variability in topsoil across England. (Inverse distance squared weighting option, cell size 1000 m and search radius 5000 m. %iles = percentiles.)

urban domain residential and industrial areas could be attributed with NBCs if the land use is significantly affected by widespread diffuse pollution.

Mercury and BaP have fewer data points to use in the data exploration, thus sites are widely spaced and so interpolated maps cannot be validly made for these elements. Instead of interpolated images, classified point maps are plotted to illustrate the spatial variability in these contaminants (Defra, 2012c,d). Furthermore, for Hg and BaP, because far fewer data are available, data sets used lack the methodology consistency, and so there is greater uncertainty associated with the data exploration (Ander et al., 2011, 2012). In order to have sufficient data points for BaP it was necessary to include all British data, which maintained all other data selection criteria shown above (Section 4) and are still taken from within the UK (Ander et al., 2011). Sampling and analytical methods used for Hg and BaP are summarised in the supplementary technical guidance sheets written for these contaminants (Defra, 2012c,d).

## 5.2. Landscape data

### 5.2.1. Soil parent material

The Soil-Parent Material Model (SPMM) for England, Wales and Scotland (Lawley, 2011) has been developed by BGS, based on the mapped boundaries of the national 1:50,000 superficial and bedrock geological data. The results of this project were used 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. Soil thus inherits 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 useful 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.

In the absence of a soil parent material map, a digital geology map (solid and drift editions) can be used to identify where elevated contaminant concentrations in soil are related to the underlying geology. Soil maps are also a landscape resource that can be used in this context, though these are often based on the underlying geology. Furthermore, for this work in England, geological and parent material maps are available at appropriate scales (i.e., 1:50,000) to do the required interpretation.

### 5.2.2. Urban areas

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 generated (Fig. 7), and the delineated urban areas were 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).

Other data sets are available in the UK for defining urban domains. At a regional to national scale light pollution maps (e.g., CPRE, 2003) are a good way of defining the built environment. At a more local scale, detailed land use information is available from the Ordnance Survey Strategi® maps (Ordnance Survey, 2011).

### 5.2.3. Metalliferous mineralisation and mining

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 mineralisation or coal mining, which are also noted to have significant controls on contaminant levels in soil. 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 in the context of underlying parent material/geology. In the GIS environment, we have used the British 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.

## 5.3. Data distributions

The data distribution of concentrations of contaminants in soil, used for characterising domains, have been explored, as this is a fundamental part of the methodology to determine NBCs (see next section). Cumulative distribution plots and boxplots are useful tools to explore such distributions (Figs. 8 and 9). These plots are included in the technical guidance sheets for each contaminant (e.g., Defra, 2012b).

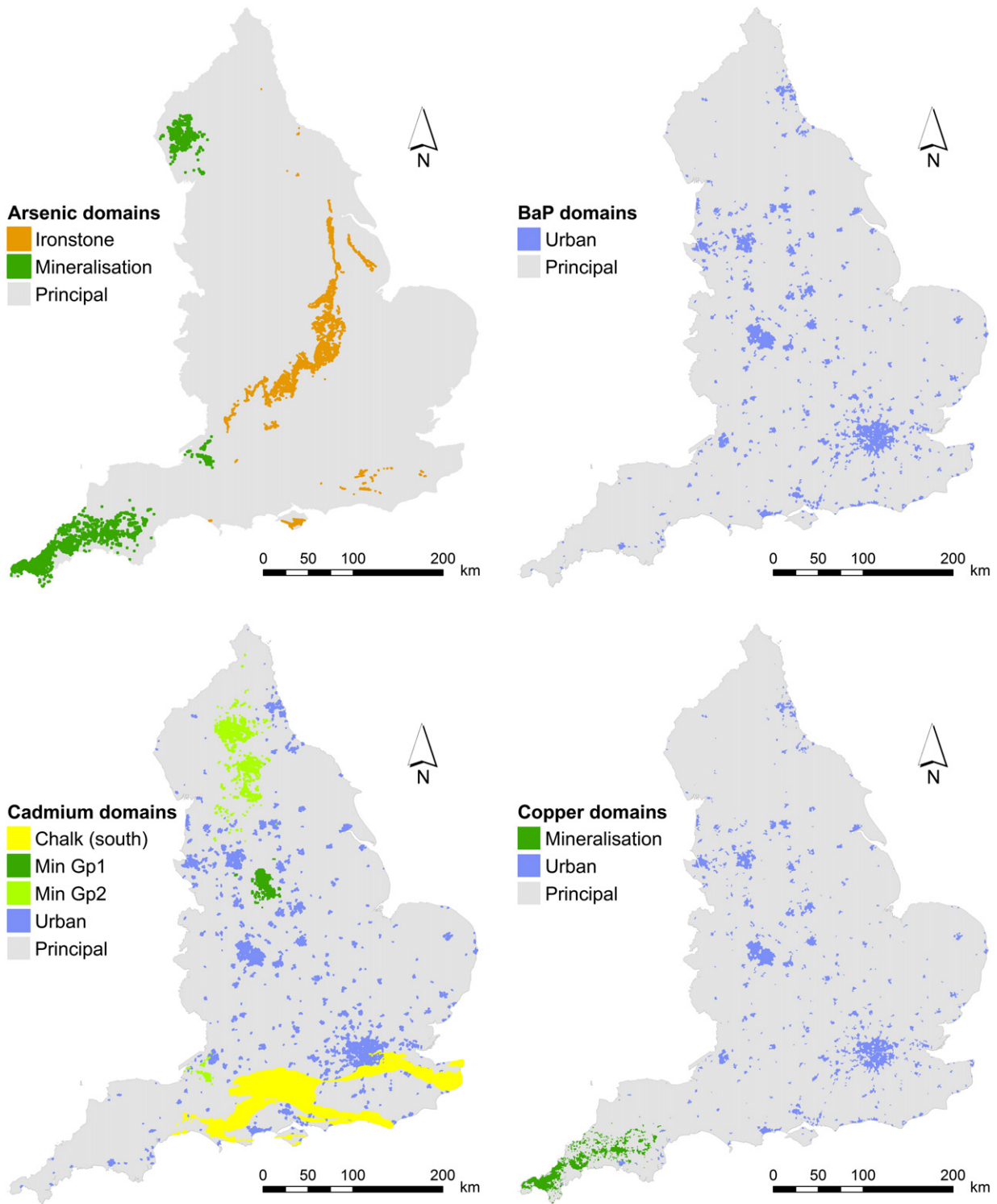


Fig. 5. Maps showing the contaminant domains for As, BaP, Cd and Cu.

An important property of the distribution of data on a variable is its degree of symmetry, and this was also investigated in the data exploration. The Gaussian distribution is symmetrical, and is the distribution that we would expect for a variable that arises from the additive combination of random effects (Allègre and Lewin, 1995). In geochemistry, it is common to find that concentrations have an asymmetric distribution, typically with a long upper tail of larger values (Reimann and Filzmoser, 2000). This can arise from non-linear combination of random effects,

and such data can be transformed to a symmetrical distribution, for example by calculating the logarithm of their values. The symmetry of the distribution of a variable can be measured by the skewness coefficient, SC:

$$SC = \frac{\sum_{i=1}^n (x_i - m)^3}{ns^3},$$



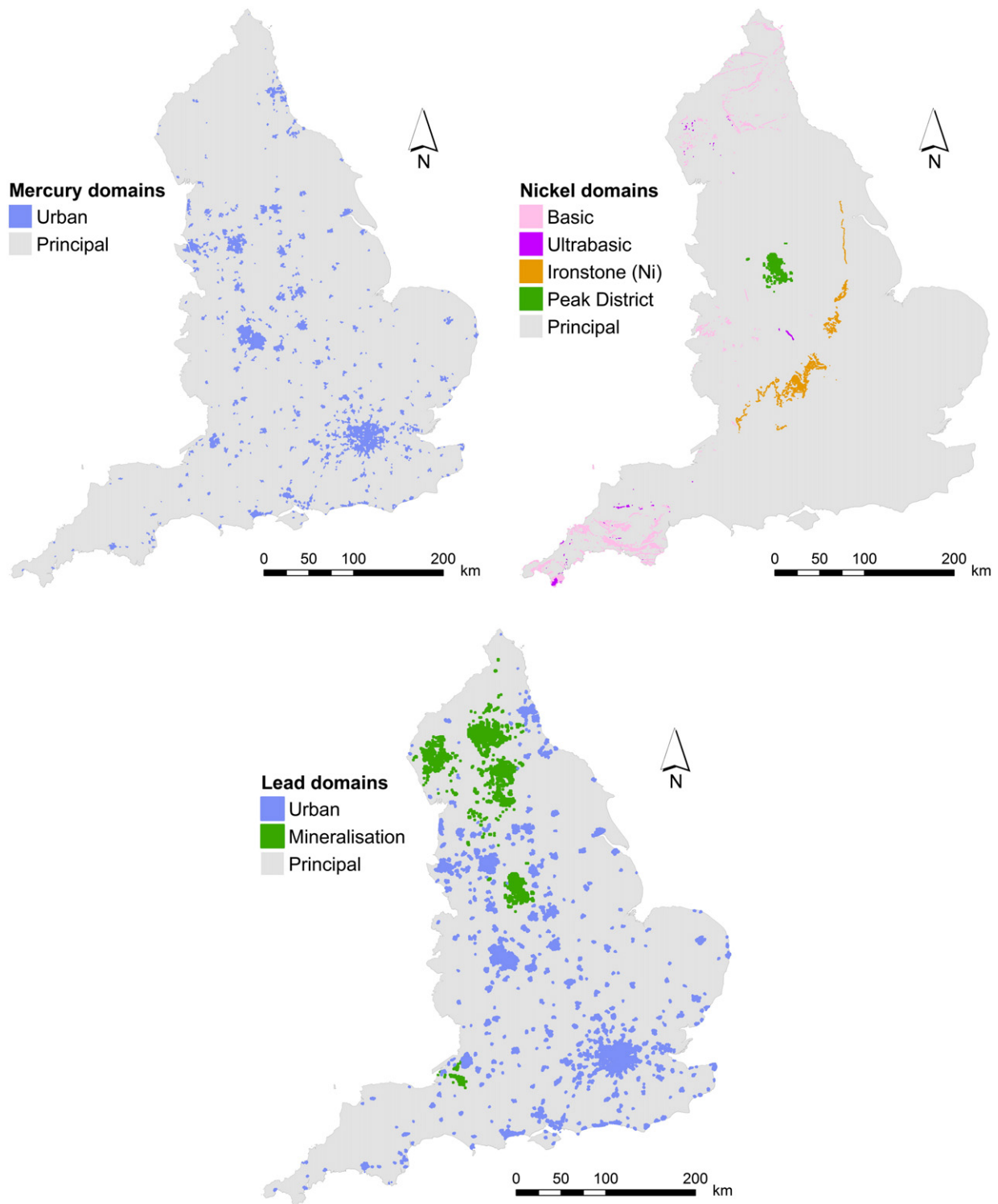


Fig. 6. Maps showing the contaminant domains for Hg, Ni and Pb.

where  $x_i$  is the  $i$ th of  $n$  observations,  $m$  is the sample average and  $s$  is the sample standard deviation. Data from a Gaussian distribution have a skewness coefficient of zero, because it is symmetrical. It is a rule of thumb that we transform data if SC lies outside the interval  $[-1, 1]$  (Webster and Oliver, 2007).

If data from a domain can be regarded as variables with a Gaussian distribution, perhaps after a transformation, then this distribution

allows us to characterise the NBCs of the domain, reflecting the multiple processes that generate variation within a domain. However, the distribution may be affected unduly by a relatively small number of extreme observations resulting from point contamination. Plots may help to identify when such processes are occurring. Such outlying data will also affect the value of SC. This means that a large positive SC may be due to an underlying asymmetry in the distribution of

**Table 2**

List of contaminant domains with summary statistical information for associated English topsoil (all concentrations in mg/kg).

	Domain	% area of England	Samples	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness coefficient
As	Principal	97	41,509	16	<0.5	10.6	14.1	18.6	1008	18
	Ironstone	1	437	73	4.1	27.8	45	83.4	555	3
	Mineralisation	2	187	181	6.9	27.8	45.6	105.5	15,110	13
BaP	Principal	96	165	0.14	<DL	0.02	0.06	0.15	1.44	4
	Urban	4	13	0.67	0.07	0.30	0.52	1.00	1.56	1
Cd	Principal	89	4418	0.5	0.3	0.3	0.3	0.5	20	17
	Urban	4	9308	0.9	<0.5	<0.5	<0.5	0.8	165	33
	Chalk South	5	265	1	0.3	0.5	0.9	1.4	5.6	2
	Min_Gp1	1	224	4.5	<0.5	2	3	5	48	4
	Min_Gp2	<1	95	0.9	0.3	0.3	0.5	0.9	13	6
	Principal	95	34,504	27	<1	14.3	19.8	27.7	5326	41
Cu	Urban	4	7475	74	1.2	30.3	47.5	79.7	4577	14
	Mineralisation	1	153	92	3	30.4	47.4	77.6	2766	10
	Principal	96	1134	0.34	<0.07	0.07	0.12	0.23	31	15
Hg	Urban	4	512	0.55	<0.07	0.18	0.33	0.65	9.6	5
	Principal	99	41,768	25.3	1	16.6	23.4	31.7	506	5
Ni	Basic	<1	23	60.9	20.5	33.4	62.5	80.9	107	0
	Ultrabasic	<1	4	213	25.4	48.2	199	393	430	0
	Ironstone (Ni)	<1	117	78.9	4.1	42.3	69.4	112	182	1
	Peak district	<1	221	44.5	5.9	22.4	34.2	51.2	384	4
	Principal	94	34,257	72	3	31	41	66	10,000	22
Pb	Urban	4	7529	276	2	89	166	322	10,000	8
	Mineralised	2	347	665	35	151	290	638	10,196	5

Using G-BASE and NSI topsoil results determined by laboratory based XRFs except BaP and Hg (see text). Cadmium data has variable detection limits between the Urban and Min\_Gp1 domains (0.5 mg/kg) and other domains (0.25 mg/kg) as fully described in Ander et al. (2012).

NBCs in the domain, the superimposed effects of point contamination, or both. A useful tool to discriminate between these circumstances is the octile skewness coefficient, OC:

$$OC = ((P_{87.5} - P_{50}) - (P_{50} - P_{12.5})) / (P_{87.5} - P_{12.5}),$$

where  $P_y$  is the  $y$ th percentile of the data, here we consider the first and seventh octile and the median. The OC is zero for a symmetrical variable, but because it is based on order statistics it is not very sensitive

to a few extreme observations. A comparable rule of thumb for the interpretation of OC is that the data require transformation if OC is outside the range  $[-0.2, 0.2]$  (Rawlins et al., 2005).

## 6. Methodology for determining normal background concentrations (NBCs)

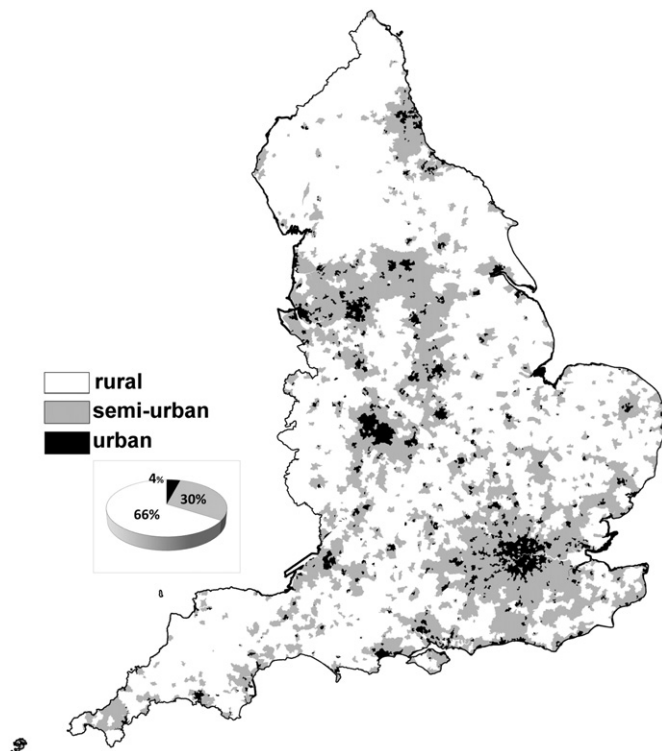
The methodology is based on the assumption that the contaminant data conform to a random variable:

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

The fixed effects are sources of variation in the observed concentrations that are attributable to geogenic sources or diffuse anthropogenic activities that are represented by the domain mean. 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 shown in Fig. 2. 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.

A robust statistical methodology was used 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 statistical analysis has been done using open source R code (R Development Core Team, 2011) employing 'fBasics', 'Hmisc', 'boot' and 'car' packages.

Previous sections have described the gathering and division of soil contaminant data into domains. The steps in this initial process are summarised in Fig. 10. Having selected the contaminant for investigation the first question to be asked is "can the contaminant be considered for a NBC?" Asbestos and manufactured organic contaminants with no natural origin, for example, fail this question as there is a lack of data on their natural distribution and concentration in soil. The contaminant data set is then divided into domain data sets. A minimum of 30 results are considered necessary to determine a NBC (BSI, 2011). Once the data



**Fig. 7.** 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).

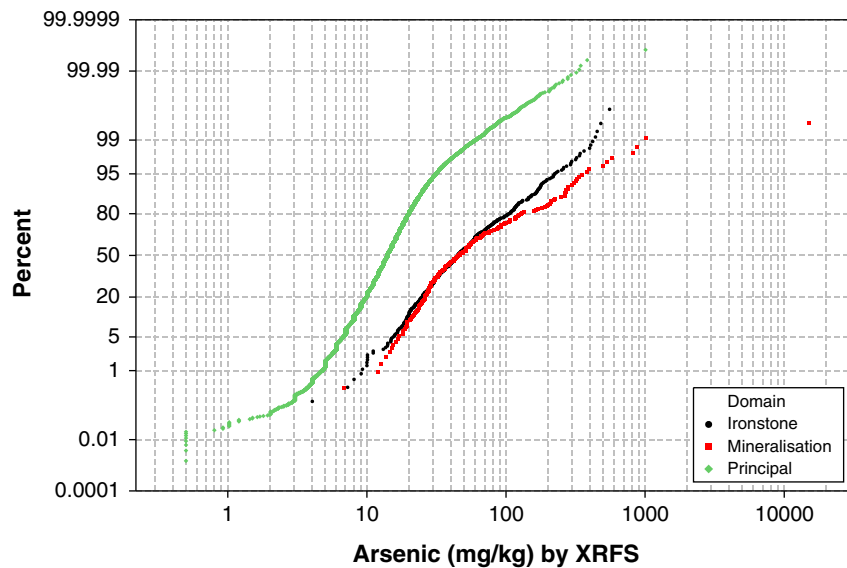


Fig. 8. Example of a frequency distribution plot used in the data exploration – cumulative probability plot of topsoil As results categorised by domains.

has been allocated into domains, then exploratory statistics indicating the skewness coefficient and inspection of frequency distribution plots can be done to select the appropriate data transform and method of calculating percentiles. The steps applied are:

- i) If the data are symmetrically distributed, SC is  $<1$  and the OS is  $<0.2$ , 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.
- ii) If the data show a mostly symmetrical distribution with potential outliers in the distribution tail, SC  $>1$  but OS  $<0.2$ , then 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 (Reimann and Filzmoser, 2000).
- iii) If the data distribution is skewed, SC is  $>1$  and the OS is  $>0.2$ , then the data are not suitable for fitting to a Gaussian model, and the data need to be transformed to using either a logarithmic

( $\log_e$ ) or the more general family of transformations known as the Box–Cox transform (Box and Cox, 1964):

$$\left( U^{\lambda} - 1 \right) / \lambda.$$

Where  $U$  is vector of soil concentrations for a given contaminant and  $\lambda$  is the power coefficient which is optimised to give the transformed data best fit to the normal distribution. This has been shown to work well for geochemical applications (Reimann et al., 2008). After transform the distribution is re-examined and steps i to iii are repeated. After calculation of the percentiles the data are back transformed to their original units;

- iv) Finally, if the data cannot be made to be consistent with a Gaussian distribution (even after transform) the empirical percentiles for the data set are calculated.

In practice the empirical, parametric and robust percentiles have been reported for each domain to check for consistency between methods (Ander et al., 2012; Cave et al., 2012). The methodology assumes that data for a given domain come predominantly from a single population, and that the data are either normally distributed or have a positive SC. For the contaminants and domains considered these assumptions hold true. Examples of distribution plots on untransformed and transformed data, along with the skewness calculations are displayed in Fig. 11. An example table (for As) of percentiles for different contaminant domains is shown in Table 3.

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. We define arbitrarily the upper limit of normal background concentrations as the 95th percentile. The percentile values are subject to uncertainty based on the number of data points and the shape of the distribution. Where sample sizes are smaller, this results in wider confidence intervals for the domain percentiles (Ander et al., 2012; Cave et al., 2012). An assessment of the uncertainty on the 95th percentile was estimated by empirical, parametric and robust parametric methods, and has been included in the statistical estimations, using a bootstrap resampling routine implemented by employing the ‘boot’ package within the R programming language. The bootstrap routine used 1000 resamples of the original or transformed data

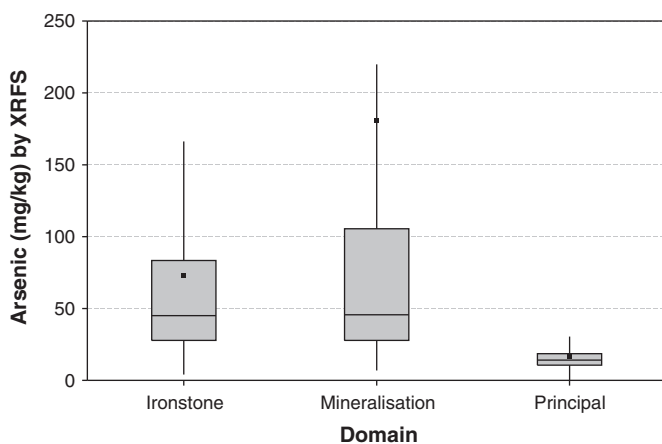


Fig. 9. Example of a boxplot used to display the range of domain concentrations for As. The box represents the interquartile range (Q1, Q3), with the median (Q2) as a line within the box. The point symbol shows the mean value. The upper whisker =  $Q3 + 1.5(Q3 - Q1)$ ; lower whisker =  $Q1 - 1.5(Q3 - Q1)$ .

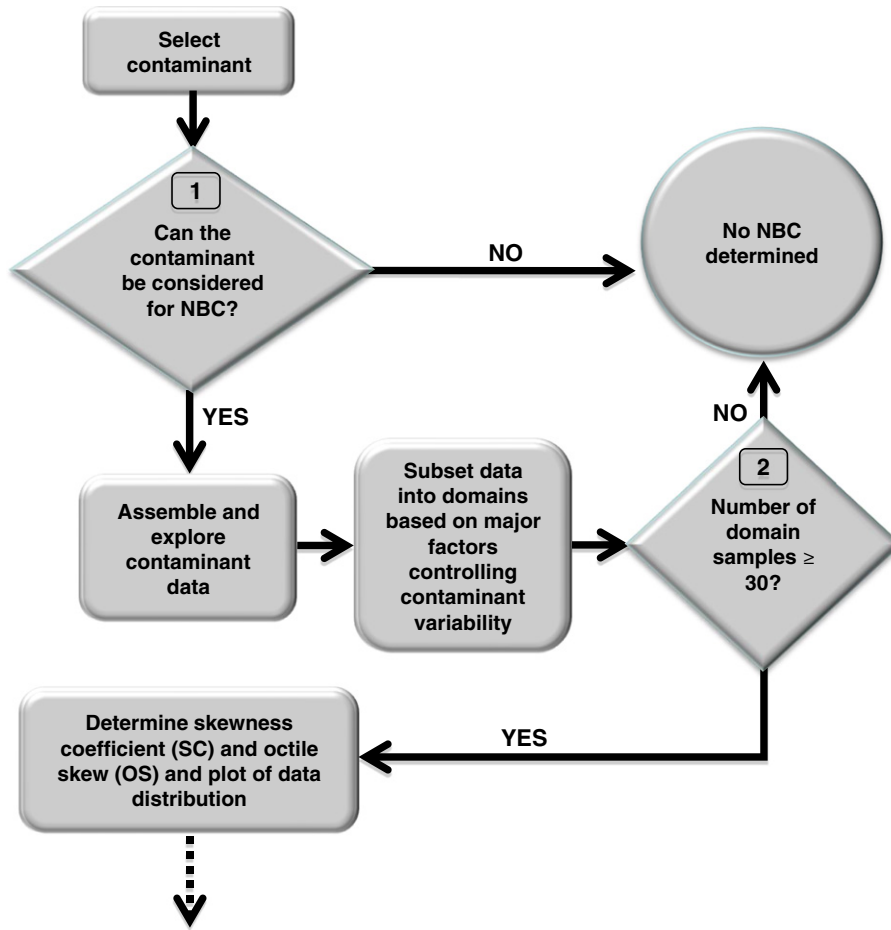


Fig. 10. Flow chart for the estimation of the NBC for a given contaminant – data gathering and exploration phase. For discussion of questions 1 & 2 see text.

providing a 95% percentile confidence interval on the estimated percentile.

Estimation of the uncertainty using bootstrapping gives the upper 95% confidence limit to the estimated 95th percentile. Comparison of the uncertainty on the estimations confirms that the parametric

uncertainties from the Gaussian fit are much less erratic than the empirical values, and thus the parametric limits have been chosen as being a better representation of the uncertainty on the percentiles.

This definition of a NBC considers what is typical and widespread (words used in the SG to describe ‘normal’ levels). The median (50th

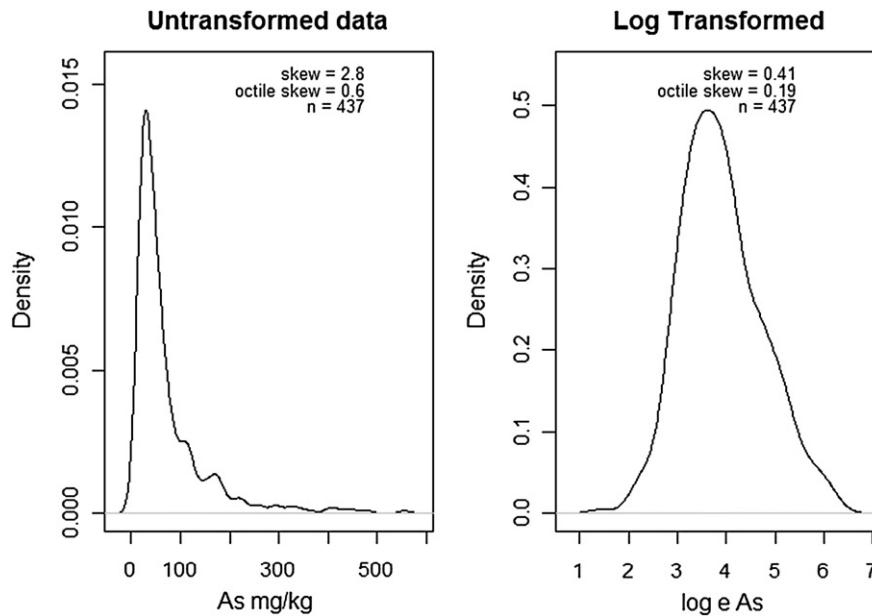


Fig. 11. Example density distributions for the raw data and the log<sub>e</sub> transformed data for As in the Ironstone Domain (n = number of samples).

**Table 3**  
Empirical (Emp), Parametric Gaussian (P) and Robust Gaussian (R) percentile values for As in the ironstone domain (concentrations in mg/kg). L and H values represent low and high confidence intervals around the median. Shaded values indicate percentile method used to estimate NBC – this is determined by exploration of the data distribution.

Percentile	Empirical	Emp L	Emp H	Parametric (P)	P L	P H	Robust (R)	R L	R H
50	45.0	41.7	49.4	50.0	46.2	53.7	45.0	41.3	49.4
55	49.7	45.3	55.8	55.5	51.1	59.7	49.8	45.5	55.1
60	56.0	50.2	60.0	61.7	56.7	66.6	55.2	49.8	61.6
65	60.9	56.7	69.6	68.8	63.2	74.5	61.4	54.8	68.7
70	71.4	61.7	81.0	77.3	70.6	84.1	68.6	60.7	77.5
75	82.7	72.6	97.7	87.5	79.6	95.9	77.4	67.7	88.7
80	104.8	86.4	116.6	100.6	91.1	111.0	88.5	76.4	102.0
85	122.7	108.9	147.1	118.3	106.4	131.7	103.5	88.1	120.7
90	165.6	140.7	179.3	145.1	129.1	163.5	126.1	105.2	149.3
95	221.7	179.7	291.4	196.2	171.6	223.9	168.8	136.6	205.4

percentile) is a measure of the central tendency, so half the observations are expected to exceed it under normal background variation. The NBC should be represented by a larger percentile. Selecting the 95th percentile is somewhat arbitrary, but will encompass a large proportion of the normal background variation whilst excluding extremes. It has also been used in other approaches to define background concentrations for environmental purposes (e.g., in Italy, APAT-ISS, 2006). As it has been argued that the NBC should represent the highest concentration of contaminant in this domain that is likely to come from normal background, then the NBC should be the upper end of the confidence interval. The NBC is defined, therefore, as the upper 95% confidence limit of the 95th percentile (taking into account data transformations).

Any uncertainty in the 95th percentile value is fully captured by taking the upper limit of the 95th percentile confidence interval. In the determination of geochemical background, of great economic value in the search for ore bodies, the mean ( $\bar{x}$ ) plus  $2\sigma$  ( $\sigma$  = standard deviation) is commonly used (Matschullat et al., 2000), which represents c. 97% of the data for normalised distributions.

## 7. NBCs for As, BaP, Cd, Cu, Hg, Ni and Pb

A summary of domain NBCs for the investigated contaminants, using the methodology described above, is given in Table 4. 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). Table 5 is an example of the summary information (in this case As) presented in the contaminant technical guidance sheet.

## 8. Discussion and concluding remarks

1. There are large amounts of high quality and systematically collected soil data covering England containing inorganic element contaminant results that have enabled us to estimate NBCs for contaminant domains. 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. We use total element concentrations for inorganic contaminants (except Hg), determined by XRFs. The total amount of an element present in a sample 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 are desirable (Darnley et al., 1995). During the data exploration, data sets for which determinands in soil had been determined by XRFs and other methods (e.g., GEMAS (Reimann et al., 2012); TELLUS (Smyth, 2007); FOREGS (Salminen et al., 2005)) have been compared, and for all the contaminants there was no

significant differences over normal ranges of concentrations for the methods concerned, which could be modelled by linear regression (e.g., see Defra, 2012b).

3. 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. 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.

**Table 4**

Summary of domain normal background concentrations (NBCs) for the studied contaminants. All concentrations in mg/kg. Number of samples used in each domain NBC estimation is shown in brackets. The Ni Basic and Ultrabasic Domains are shown here, but as they are each defined by less than 30 samples, the NBCs are not estimated. Note that the number of samples in the Hg Principal Domain is eight fewer than cited in Table 2 because these eight samples had problems regarding the spatial resolution of the site locations.

As	DOMAIN				
	Principal	Mineralisation	Ironstone		
	32 (41,509)	290 (187)	220 (437)		
BaP	DOMAIN (Great Britain)				
	Principal	Urban			
	0.5 (371)	3.6 (32)			
Cd	DOMAIN				
	Principal	Min. Grp. 1	Min. Grp. 2	Urban	Chalk (south)
	1.0 (4,418)	17 (224)	2.9 (95)	2.1 (9,308)	2.5 (265)
Cu	DOMAIN				
	Principal	Mineralisation	Urban		
	62 (34,504)	340 (153)	190 (7,475)		
Hg	DOMAIN				
	Principal	Urban			
	0.5 (1,126)	1.9 (512)			
Ni	DOMAIN				
	Principal	Ironstone (Ni)	Peak District	Basic	Ultrabasic
	42 (41,768)	230 (117)	120 (221)	* (23)	* (4)
Pb	DOMAIN				
	Principal	Mineralisation	Urban		
	180 (34,257)	2,400 (347)	820 (7,529)		

**Table 5**

NBCs for the arsenic domains (cited to 2 significant figures, n is number of samples used in the estimation). Arsenic is determined by laboratory-based X-ray fluorescence spectrometry (XRFS), i.e., total As in soil sampled from a depth 0–15 cm. The NBC is the upper confidence interval on the 95th percentile of the domain data.

Domain	Area (km <sup>2</sup> )	Area (%)	NBC (mg/kg)	n
Ironstone	1300	1	220	437
Mineralisation	2300	2	290	187
Principal	129,300	97	32	41,509

- Where soil data are sparse, sample media other than soil, 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 and other high density stream sediment sampling of Great Britain and Northern Ireland (Webb et al., 1978; Johnson et al., 2005) can provide supplementary information to define contaminant domains for the areas not covered by the high density G-BASE soil sampling.
- Contaminant information for soil 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.
- The statistical methodology used in this work can be used by others either to determine NBCs not investigated here or estimated 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 re-estimated to greater levels of confidence, particularly in those areas where there are knowledge gaps.
- 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 (Defra, 2012a). They are not a planning or risk assessment tool and must be used in the context of the Guidance.
- We define the NBC as the upper confidence limit of the 95th percentile. Other percentiles and their confidence limits have been estimated and are available, should others wish to consider 'normal' levels in the context of other statistical information.

## Acknowledgements

This research work was commissioned by the UK Department for Environment, Food and Rural Affairs as a Science and Research Project (SP1008) in support of the Part 2a contaminated land Statutory Guidance revision published in April 2012. We also thank Professor Steve McGrath at Rothamsted Research for use of the data from the XRFS analyses of the National Soil Inventory samples. This work is published with the permission of the Executive Director of the British Geological Survey (NERC).

## References

- Allègre CJ, Lewin E. Scaling laws and geochemical distributions. *Earth Planet Sci Lett* 1995;132:1–13.
- Ander EL, Cave MR, Johnson CC, Palumbo-Roe B. Normal background concentrations of contaminants in the soils of England. Available data and data exploration Commissioned report CR/11/145. Keyworth, Nottinghamshire: British Geological Survey/Defra; 2011 [124 pp. <http://nora.nerc.ac.uk/19958/>].
- Ander EL, Cave MR, Johnson CC, Palumbo-Roe B. Normal background concentrations of contaminants in the soils of England. Results of the data exploration for Cu, Ni, Cd and Hg. Commissioned report CR/12/041. Keyworth, Nottinghamshire: British Geological Survey/Defra; 2012 [88 pp. <http://nora.nerc.ac.uk/19952/>].
- APAT-ISS. Protocollo Operativo per la determinazione dei valori di fondo di metalli/metalloidi nei suoli di interesse nazionale. Revisione 0. Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici and Istituto Superiore di Sanita; 2006 [in Italian].

- Appleton JD. Potentially harmful elements from natural sources and mining areas: characteristics, extent and relevance to planning and development in Great Britain. British Geological Survey/Technical report WP/95/3. UK: Commission report for the Department of Environment (DoE); 1995 [63 pp.].
- Appleton JD, Rawlins BG, Thornton I. National-scale estimation of potentially harmful element ambient background concentrations in topsoil using parent material classified soil:stream-sediment relationships. *Appl Geochem* 2008;23(9): 2596–611.
- Archer FC, Hodgson IH. Total and extractable trace element contents of soils in England and Wales. *J Soil Sci* 1987;38:421–31.
- Barracough D. UK soil and herbage pollutant survey. Introduction and summary. UK: Environment Agency (EA); 2007.
- Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc Ser B* 1964;26:211–46.
- BSI. Soil quality: guidance on the determination of background values. BS EN ISO 19258:2011. International Organisation for Standardisation; 2011 [revision of BS ISO 19258:2005; 2011].
- Cave MR, Johnson CC, Ander EL, Palumbo-Roe B. Methodology for the determination of normal background contaminant concentrations in English soils. British Geological Survey commissioned report. CR/12/003; 2012 [41 pp. <http://nora.nerc.ac.uk/19959/>].
- Cole S, Jeffries J. Using soil guideline values. Environment agency, UK SC050021/SGV introduction; 2009 [32 pp.].
- Communities and Local Government. Generalised land use database statistics for England 2005. Product Code 06CSR04342. Department for Communities and Local Government; 2007 [February; [http://data.gov.uk/dataset/land\\_use\\_statistics\\_generalised\\_land\\_use\\_database](http://data.gov.uk/dataset/land_use_statistics_generalised_land_use_database)].
- Cousins IT, Kreibich H, Hudson LE, Lead WA, Jones KC. PAHs in soils: contemporary UK data and evidence for potential contamination problems caused by exposure of samples to laboratory air. *Sci Total Environ* 1997;203(2):141–56.
- CPRE. Night blight in England and UK. Light pollution images. Campaign to protect rural England (CPRE); 2003 [April; <http://www.cpre.org.uk/resources/country-side/dark-skies/item/1977>].
- Darnley AG, Bjorklund A, Bolvikken B, Gustavsson N, Koval PV, Plant JA, et al. A global geochemical database for environmental and resource management. UNESCO publishing 19; 1995 [122 pp.].
- De Vos W, Tarvainen T. Geochemical Atlas of Europe. Part 2. Interpretation of Geochemical Maps, Additional Tables, Figures, Maps, and Related Publications. Espoo, Finland: Geological Survey of Finland; 2006. [692 pp.].
- Defra. The Contaminated Land (England) Regulation. Statutory instruments 2006 No. 1380. The Stationery Office (UK) Ltd.; 2006 [4th August; <http://www.legislation.gov.uk/ukSI/2006/1380/contents/made>].
- Defra. Guidance on the Legal Definition of Contaminated Land. UK: Department for Environment Food and Rural Affairs (Defra); 2008 [July; 16 pp.].
- Defra. Simplification of the contaminated land regime. Impact assessment (IA) No Defra 113. UK: Department for Environment, Food and Rural Affairs (Defra); 2011 [October; 41 pp. <http://archive.defra.gov.uk/environment/quality/land/contaminated/documents/contaminated-land-ia.pdf>].
- Defra. Environmental Protection Act 1990: Part 2A Contaminated Land Statutory Guidance. UK: Department for Environment, Food and Rural Affairs (Defra); 2012a [April; 69 pp. <http://www.defra.gov.uk/publications/2012/04/10/pb13735contaminated-land/>].
- Defra. Technical Guidance Sheet on normal levels of contaminants in English soils: arsenic – supplementary information. In: Johnson CC, Ander EL, Cave MR, Palumbo-Roe B, editors. Normal background concentrations in Soil Technical Guidance sheet. London, UK: Department for Environment Food and Rural Affairs (Defra), Soils R&D Project SP1008; 2012b. [TGS01s: pp21. <http://nora.nerc.ac.uk/19964/>].
- Defra. Technical Guidance Sheet on normal levels of contaminants in English soils: benzo[a]pyrene – supplementary information. In: Johnson CC, Ander EL, Cave MR, Palumbo-Roe B, editors. Normal background concentrations in Soil Technical Guidance sheet. London, UK: Department for Environment Food and Rural Affairs (Defra), Soils R&D Project SP1008; 2012c. [TGS04s: pp16. <http://nora.nerc.ac.uk/19966/>].
- Defra. Technical Guidance Sheet on normal levels of contaminants in English soils: mercury – supplementary information. In: Johnson CC, Ander EL, Cave MR, Palumbo-Roe B, editors. Normal background concentrations in Soil Technical Guidance sheet. London, UK: Department for Environment Food and Rural Affairs (Defra), Soils R&D Project SP1008; 2012d. [TGS07s: pp16. <http://nora.nerc.ac.uk/19972/>].
- Defra, EA. Potential contaminants for the assessment of land. UK: Department for Environment, Food and Rural Affairs and Environment Agency; 2002 [34 pp.].
- DETR. Environmental Protection Act 1990: Part IIA Contaminated Land. DETR Circular 02/2000, Department of the Environment, Transport & the Regions (DETR). UK. London: HMSO; 2000 [20th March; 162 pp.].
- Emmett BA, Reynolds B, Chamberlain PM, Rowe E, Spurgeon DJ, Brittain SA, et al. Countryside Survey. Soils Report from 2007. Centre for Ecology & Hydrology CS Technical report no. 9/07; 2010 [194 pp.].
- Finnish Government. Government decree on the assessment of soil contamination and remediation needs. 1st March 2007. Decree 214/2007 Ministry of the Environment; 2007 [in Finnish – Unofficial translation].
- Flight DMA, Scheib A. Soil geochemical baselines in UK urban centres: the G-BASE project. In: Johnson CC, Demetriades A, Locutura J, Ottesen RT, editors. Mapping the Chemical Environment of Urban Areas. Chichester, UK: John Wiley & Sons Ltd; 2011. p. 186–206.
- Fordey FM, Brown S, Ander EL, Rawlins BG, O'Donnell KE, Lister TR, et al. GSUE: urban geochemical mapping in Great Britain. *Geochem Explor Environ Anal* 2005;5(4): 325–36.
- Grunsky EC. The interpretation of geochemical survey data. *Geochem Explor Environ Anal* 2010;10(1):27–74.

- Hamon RE, McLaughlin MJ, Gilkes RJ, Rate AW, Zarcinas B, Robertson A, et al. Geochemical indices allow estimation of heavy metal background concentrations in soils. *Global Biogeochem Cycles* 2004;18:1–6.
- Hawkes HE, Bloom H. Heavy metals in stream sediment used as exploration guides. *Min Eng* 1955;8:1121–6.
- Johnson CC, Ander EL. Urban geochemical mapping studies: how and why we do them. *Environ Geochem Health* 2008;30:511–30.
- Johnson CC, Breward N, Ander EL, Ault L. G-BASE: baseline geochemical mapping of Great Britain and Northern Ireland. *Geochem Explor Environ Anal* 2005;5(4):347–57.
- Johnson CC, Flight DMA, Ander EL, Lister TR, Breward N, Fordyce FM, et al. The collection of drainage samples for environmental analyses from active stream channels. In: de Vivo B, Belkin HE, Lima A, editors. *Environmental Geochemistry: site characterisation, data analysis and case histories*. Oxford: Elsevier; 2008. p. 59–92.
- Johnson CC, Demetriades A, Locutura J, Ottesen RT. Mapping the chemical environment of urban areas. Chichester, UK: John Wiley & Sons Ltd; 2011. [664 pp.].
- Johnson CC, Ander EL, Cave MR, Palumbo-Roe B. Normal background concentrations (NBCs) of contaminants in English soils: final project report. British Geological Survey commissioned report, CR/12/035; 2012 [40 pp. <http://nora.nerc.ac.uk/19946/>].
- Jones KC, Stratford JA, Waterhouse KS, Vogt NB. Organic contaminants in Welsh soils – polynuclear aromatic hydrocarbons. *Environ Sci Technol* 1989;23(5):540–50.
- Lawley R. The soil-parent material database: a user guide. British Geological Survey Open Report, OR/08/034; 2011 [53 pp. <http://nora.nerc.ac.uk/8048/>].
- Lepeltier C. A simplified statistical treatment of geochemical data by graphical representation. *Econ Geol* 1969;64(5):538–50.
- Lovering TS, Huff LC, Almond H. Dispersion of copper from the San Manuel copper deposit, Pinal County, Arizona. *Econ Geol* 1950;45:493–514.
- Martin I, Cowie C. Compilation of data for priority organic pollutants for derivation of soil guideline values. UK: Environment Agency; 2008 [SC050021/SR7; 171 pp.].
- Matschullat J, Ottenstein R, Reimann C. Geochemical background – can we calculate it? *Environ Geol* 2000;39:990–1000.
- McGrath SP, Loveland PJ. The soil geochemical atlas of England and Wales. Glasgow: Blackie Academic and Professional; 1992.
- Montanarella L, Tóth G, Jones DG. Soil component in the 2009 LUCAS survey. In: Tóth G, Németh T, editors. *The international conference on 'Land Quality and Land Use Information in the European Union' 26–27 May 2011 in Keszthely, Hungary*. Keszthely, Hungary: European Commission, Joint Research Centre, Institute for Environment and Sustainability; 2011. p. 209–18.
- Oliver MA, Loveland PJ, Frogbrook ZL, Webster R, McGrath SP. Statistical and geostatistical analysis of the National Soil Inventory of England and Wales. Defra (formerly MAFF) Soil Programme Technical Report Project SP0124; 2002.
- Ordnance Survey. Strategi@ user Guide and technical specification. Ordnance Survey; 2011 [V8.3–12/2011; 102 pp. <http://www.ordnancesurvey.co.uk/oswebsite/products/strategi/index.html>].
- Ottesen RT, Alexander J, Langedal M, Høygaard E. Soil pollution in day-care centres and playgrounds in Norway: national action plan for mapping and remediation. *Environ Geochem Health* 2008;30:623–37.
- Ove Arup. Mining instability in Britain. Department of Environment contract. Hard-copy maps digitised and adapted by the British Geological Survey; 1990 [<http://www.bgs.ac.uk/gbase/NBCDefraProject.html>].
- Paterson E, Towers W, Bacon JR, Jones M. Background levels of contaminants in Scottish soils. Final contract report to SEPA. The Macaulay Institute; 2003 [60 pp.].
- R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2011 [<http://www.r-project.org/>].
- Rawlins BG, Webster R, Lister TR. The influence of parent material on top soil geochemistry in eastern England. *Earth Surf Proc Land* 2003;28:1389–409.
- Rawlins BG, Lark RM, O'Donnell KE, Tye AM, Lister TR. The assessment of point and diffuse metal pollution of soils from an urban geochemical survey of Sheffield, England. *Soil Use Manag* 2005;21(4):353–62.
- Rawlins BG, McGrath SP, Scheib AJ, Cave MR, Breward N, Lister TR, et al. The advanced soil geochemical atlas of England and Wales. Keyworth, Nottingham: British Geological Survey; 2012 [Published as an electronic book. <http://www.bgs.ac.uk/gbase/advsoilatlasEW.html>].
- Reimann C, Filzmoser P. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. *Environ Geol* 2000;39(9):1001–14.
- Reimann C, Garrett RG. Geochemical background – concept and reality. *Sci Total Environ* 2005;350:12–27.
- Reimann C, Garrett RG, Filzmoser P. Background and threshold – critical comparison of methods of determination. *Sci Total Environ* 2005;346:1–16.
- Reimann C, Filzmoser P, Garrett RG, Dutter R. *Statistical data analysis explained – applied environmental statistics with R*. John Wiley and Sons, Ltd.; 2008 [343 pp.].
- Reimann C, de Caritat P, GEMAS Project Team, NGS Project Team. New soil composition data for Europe and Australia: demonstrating comparability, identifying continental-scale processes and learning lessons for global geochemical mapping. *Sci Total Environ* 2012;416:239–52.
- Salminen R, Batista MJ, Bidovec M, Demetriades A, de Vivo B, de Vos W, et al. Geochemical atlas of Europe. Part 1, background information, methodology and maps. Espoo, Finland: Geological Survey of Finland; 2005 [526 pp.].
- Sinclair AJ. Applications of probability graphs. Nepean, Ontario: Association of Exploration Geochemists; 1976 [95 pp.].
- Sinclair AJ. Univariate analysis. In: Govett GJS, editor. *Handbook of exploration geochemistry*. In: Howarth RJ, editor. *Statistics and data analysis in geochemical prospecting*, Amsterdam: Elsevier; 1983. p. 59–81. (Chapter 3).
- Sinclair AJ. Statistical interpretation of soil geochemical data. In: Fletcher WK, Hoffman SJ, Mehrtens MB, Sinclair AJ, Thompson I, editors. *Exploration geochemistry: Design and interpretation of soil surveys*. In: Robertson JM, editor. *Reviews in Economic Geology* Chelsea, MI, USA: Society of Economic Geologists; 1986. p. 97–115.
- Smyth D. Methods used in the Tellus geochemical mapping of Northern Ireland. British Geological Survey open report OR/07/022; 2007 [<http://nora.nerc.ac.uk/14008/>].
- Tarvainen T, Jarva J. Using geochemical baselines in the assessment of soil contamination in Finland. In: Johnson CC, Demetriades A, Locutura J, Ottesen RT, editors. *Mapping the Chemical Environment of Urban Areas*. Chichester, UK: John Wiley & Sons Ltd; 2011. p. 223–31.
- Tennant CB, White ML. Study of the distribution of some geochemical data. *Econ Geol* 1959;54:1281–90.
- Tidball RR, Erdman JA, Ebens RJ. Geochemical baselines for sagebrush and soil, Powder River Basin, Montana Wyoming. US Geological Survey open-file report, 74–250; 19746–13.
- Tipping E, Poskitt JM, Lawlor AJ, Wadsworth RA, Norris DA, Hall JR. Mercury in United Kingdom topsoils; concentrations, pools, and critical limit exceedances. *Environ Pollut* 2011;159:3721–9.
- Webb J, Thornton I, Thompson M, Howarth R, Lowenstein P. The Wolfson geochemical atlas of England and Wales. Oxford: Clarendon Press; 1978.
- Webster R, Oliver MA. *Geostatistics for environmental scientists*. 2nd ed. Chichester: John Wiley & Sons; 2007.
- Zhao FJ, McGrath SP, Merrington G. Estimates of ambient background concentrations of trace metals in soils for risk assessment. *Environ Pollut* 2007;148(1):221–9.