British Geological Survey

NATURAL ENVIRONMENT RESEARCH COUNCIL

Arolwg Daearegol Prydain

CYNGOR YMCHWILYR AMGYLCHEDD NATURIOL

Normal background concentrations of contaminants in the soils of Wales. Exploratory data analysis and statistical methods

Science Facilities Directorate

Commissioned Report CR/12/107N

BRITISH GEOLOGICAL SURVEY

Science Facilities Directorate
COMMISSIONED REPORT CR/12/107N

Normal background concentrations of contaminants in the soils of Wales. Exploratory data analysis and statistical methods

E L Ander, M R Cave and C C Johnson

Keywords

soil, contaminated land, Statutory Guidance, contaminant.

Bibliographical reference

ANDER, EL, Cave, MR and Johnson CC. 2013. Normal background concentrations of contaminants in the soils of Wales. Exploratory data analysis and statistical methods. *British Geological Survey Commissioned Report*, CR/12/107. 144pp.

Keyworth, Nottingham British Geological Survey 2013

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 bgslondon@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 work is an extension of the Defra-funded Science and Research project SP1008 to establish normal background concentrations of contaminant in soils, expanding the initial study of English soils to those of Wales. The principal output for this work is a series of technical guidance sheets (TGSs) similar to those produced for English soil. This report documents the data sources, exploratory data analysis and application of the statistical methodology established for English soils and covers the same contaminants, namely, arsenic (As), cadmium (Cd), copper (Cu), mercury (Hg), nickel (Ni) and lead (Pb). A TGS for benzo[*a*]pyrene (BaP) has also been produced for Welsh soil, though as this is based on the same data set as explored for England utilising results from across Great Britain, it is not discussed further in this report.

Acknowledgements

The authors acknowledge that this work on Welsh soil data builds on the efforts of a broader project team listed in the acknowledgments of earlier project reports (see Appendix 1 for listing of project reports). In particular, we would like to thank David Devaney, Defra Soils Policy Team who initiated this extension to the original project SP1008.

The principal soil data sets used are: BGS G-BASE (© NERC) project (available for use under licence); and, the National Soil Inventory (NSI) samples from Wales (reanalysed by XRFS at BGS laboratories) and used under licence (ref. L0212/006570). We also thank Professor Steve McGrath at Rothamsted Research for use of the data from the XRFS analyses of the National Soil Inventory samples. Domain extents have been explored using the BGS Parent Material Model and Ove Arup the Metalliferous Mineralisation and Mining Database. Urban areas have been defined using the Ordnance Survey (OS) Strategi® data under OS OpenData™ terms and other OS topographic data is used under licence reference number 100021290.

Contents

FIGURES

TABLES

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 soils in support of the Part 2A Contaminated Land Statutory Guidance revision. The work was initially focused on England but has been expanded here to look at Welsh soil. This work explores contaminant spatial variability and population distributions in topsoils for six contaminants, namely, arsenic (As), cadmium (Cd), copper (Cu), mercury (Hg), nickel (Ni) and lead (Pb). A seventh contaminant, benzo[*a*]pyrene (BaP), was previously explored for British soil because of a paucity of available data and so is not further explored here.

Normal levels of contaminants are quantified as normal background concentrations (NBCs). These are attributed to different regions of Wales for each contaminant, based on factors that are seen to contribute to higher concentrations in some areas, referred to here 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/industrialisation.

Methods used for calculating Welsh soil NBCs essentially follow those described in earlier Project reports, though any differences in methodology from those used in the study of English soil are described here in more detail, *e.g.* the definition of urban areas. Technical guidance sheets (TGSs) have been produced for seven contaminants following the format of those produced for England. Supplementary information in support of the Welsh TGSs for each contaminant (data exploration, diagrams, tables and maps) are the core part of this report.

Using statistical measures developed in the initial Project (*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 when there is sufficient and representative data sets.

There are significantly more knowledge gaps concerning the distribution of contaminants in soils from Wales than from England, because, unlike England, there is a very limited number of high density and systematically sampled soils from the British Geological Survey's (BGS) Geochemical Baseline Survey of the Environment (G-BASE) project. The reduced amount of data means there is far greater uncertainty attached to the NBCs determined for Welsh soil. Furthermore, results for systematically sampled urban soils are dominated by the BGS Swansea soil data set. Swansea is one of the most contaminated urban areas in the UK and cannot be considered as being representative of Welsh urban or built-up areas.

1 Introduction

In April 2012 revised Part 2A contaminated land Statutory Guidance (SG) for Wales was issued by the Welsh Government (Welsh Government, 2012). Within the SG references are made to the 'normal' presence and levels of contaminants in soils and the British Geological Survey (BGS) has been commissioned by Defra to give guidance on what are normal concentrations of contaminants in English and now Welsh soils. Originally, 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). Asbestos in soil has been explored but has been deemed an example of a contaminant with insufficient information to calculate a NBC (Ander *et al*., 2011) and so is not explored any further here. Additionally, so sparse is the data for BaP in systematically sampled soils that all available British data was used to quantify the NBC for England, with the result that there is no further data analysis that could be undertaken for Wales.

This work contained within this report uses the same methodologies as those developed in Work Packages 1 – 3 of Defra Project SP1008 for the soils of England, summarised in Johnson *et al*. (2012) and other project reports (Ander *et al*. 2011; Cave *et al*. 2012; Ander *et al*. 2012). Any differences in the methodologies are highlighted and discussed in more detail in the following sections of this report. The project methodology and approach is also reported in Ander *et al*. (2013).

For England, the principal output was a series of short technical guidance sheets (TGSs - see Appendix 1) delivering Work Package 4 of Project SP1008. A similar series of TGSs for Wales has been prepared as a result of an extension of Project SP1008 to cover Welsh soils, referred to as Work Package 5. The information for each contaminant studied for Welsh soils (diagrams, tables, maps and results for the calculation of NBCs) is presented in subsequent sections of this report and the presentation of this work, to supplement the TGSs, is the main purpose of this report.

The principal difference between Work Package 4 (England) and Work Package 5 (Wales) is the way in which urban domains are defined. Firstly, the urban extent is defined by the Ordnance Survey (OS) Strategi® data, as the Generalised Land Use Database (GLUD), derived from the National Land Use Database (NLUD) (Office for National Statistics, 2006) used for England, is not available for Wales (see Section 2.2.2). Additionally, whereas in England there are many sizable cities with long-established industrial histories, in Wales industrialisation is dominated by the area of the South Wales Coalfield and a few associated urban centres. Urban areas are defined by the 'Large Urban' class of the OS Stategi™ data set and this is the Urban Domain defined for BaP and Hg. The area of the three valleys draining the western South Wales Coalfield (including Swansea) for many of the contaminants is defined as a separate domain – Urban 1 Domain. For Cu and Pb it is possible to distinguish several characteristically different urban areas, namely the Urban 1 Domain, as just described, and then all the urban areas outside the Urban 1 Domain, *i.e.* Urban 2 Domain.

2 General datasets/information

2.1 SOIL DATA SETS WITH CONTAMINANT RESULTS

Information on the sources of soil data used in this study are summarised below, but are described and discussed in more detail in Ander *et al*. (2011). Compared with England, there is very much less systematically collected soil data for Wales because the BGS G-BASE project has done no major regional soil sampling programme in Wales. A comparison of G-BASE topsoil sample coverage for England and Wales can be seen using the Google Earth G-BASE sample location index[1].

2.1.1 National Soil Inventory (NSI) samples reanalysed by XRFS

There are 798 NSI sample sites in Wales used in the data exploration here, so this data set represents the most comprehensive national coverage of soil data (McGrath and Loveland, 1992; Oliver *et al*., 2002; and Figure 1). This is a low density systematic soil survey with samples taken at a density of one sample every 25 km^2. Soils were collected with a screw auger and 25 subsamples were sampled from within a 20 m square, sampling the uppermost 15 cm of mineral soil, *i.e.* topsoil. These samples have recently been reanalysed by X-ray fluorescence spectrometry (XRFS) at the BGS laboratories (and hence referred to as NSI(XRFS)). These XRFS results and the sampling methodology make the soil results directly compatible with the BGS G-BASE and NSI(XRFS) results used to calculate the English domain NBCs for As, Cd, Cu, Ni and Pb. The NSI(XRFS) results for England and Wales are reported as an electronic atlas (Rawlins *et al*. 2012).

2.1.2 G-BASE urban

Whilst the G-BASE project has not collected any systematic regional soil samples, it has collected topsoil samples from two urban centres in Wales - Swansea and Cardiff (see Figure 1). The methods for collection of urban samples by G-BASE have been reviewed in Fordyce *et al*. (2005). For the G-BASE urban soil sampling, the methods used are identical to regional sampling, though in urban areas samples are collected at 4 sites per km^2 (regional is 1 site per 2 km^2). Soils are collected with a 1-m auger, each sample is made up of five subsamples from corners and centre of a 20 m square and are sampled from the uppermost 15 cm of mineral soil (nominally a depth of 5 – 20 cm allowing for vegetation cover and surface litter). There are 507 topsoils from Cardiff and 370 from Swansea. The results from Swansea and Cardiff are presented, respectively, in Morley and Ferguson (2001) and Brown (2001). More detailed analysis of the Swansea results are given in Waters *et al*. (2005) and Flight and Scheib (2012), and both of which demonstrate the exceptionally high levels of metal contaminants added to the Swansea environment as a result of the legacy of metallurgical industries. The Swansea soil urban data set cannot be seen as being typical for Welsh urban areas.

[1] http://www.bgs.ac.uk/gbase/sampleindexmaps/home.html

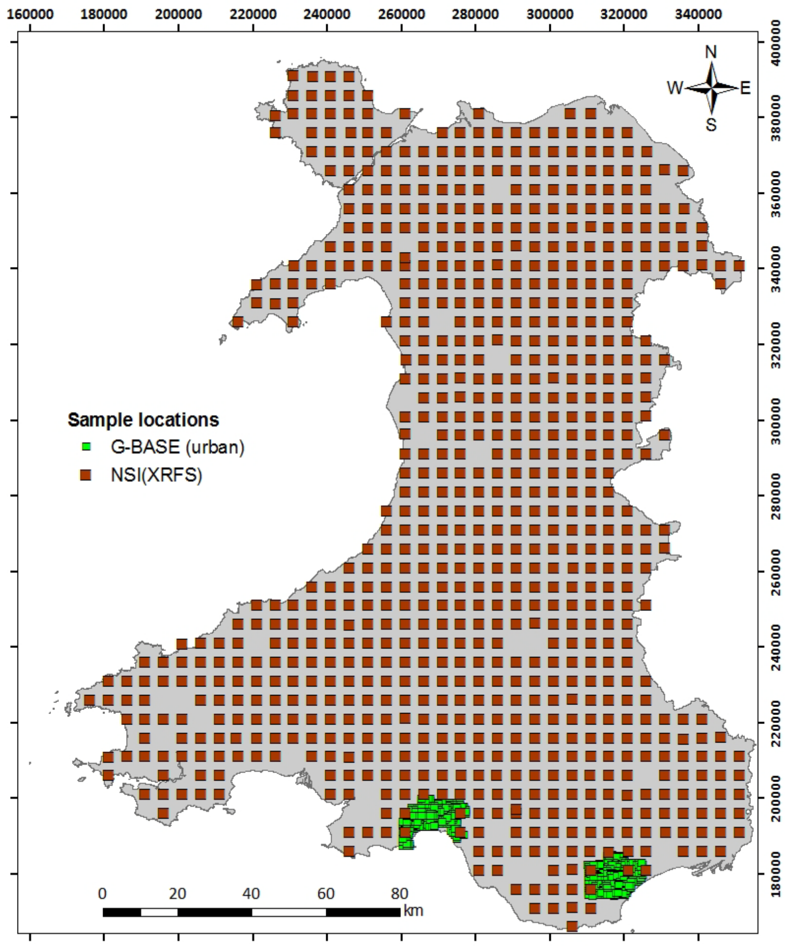


Figure 1: Soil site locations

2.1.3 Soil data sets for Hg

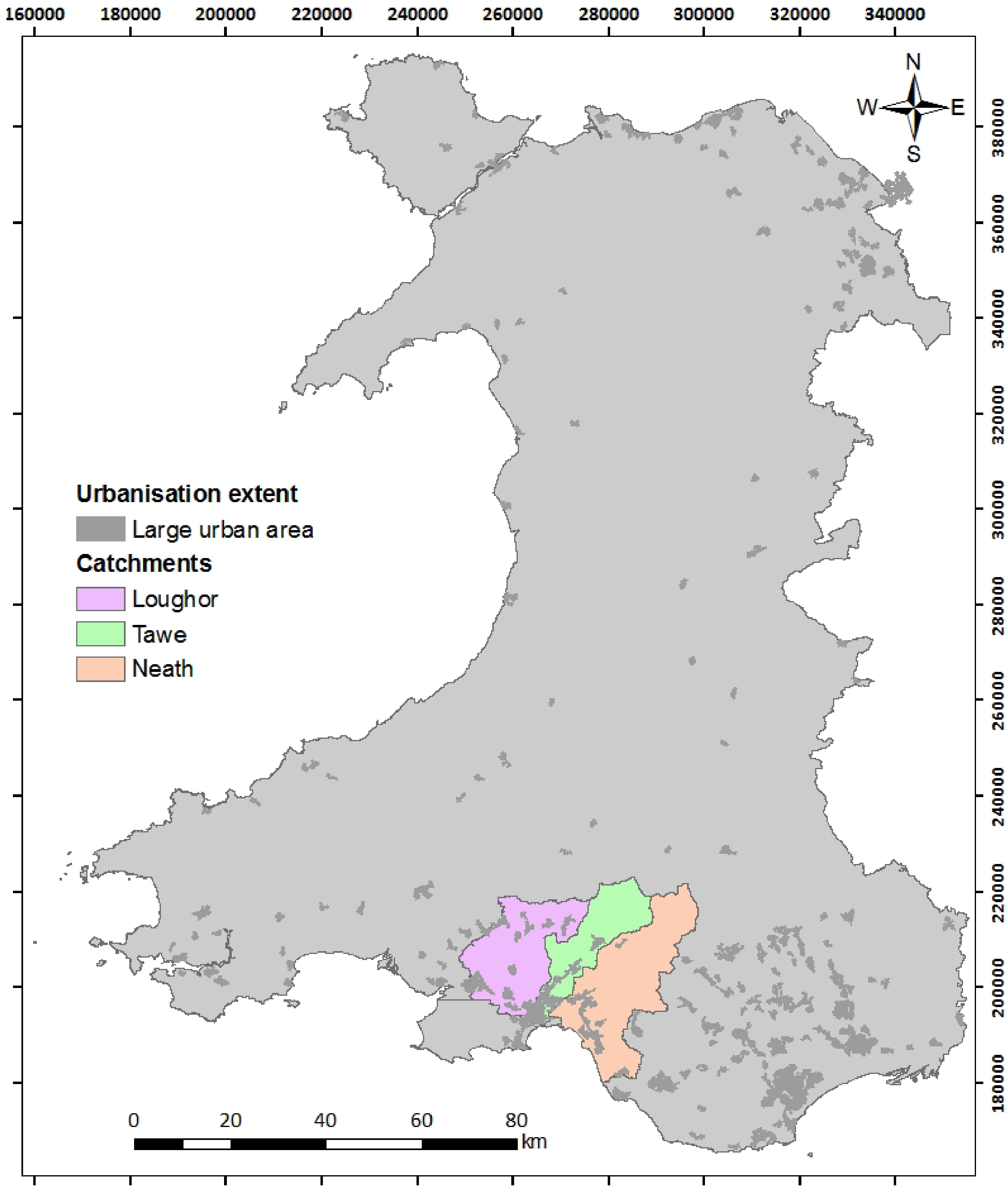
The XRFS analyses of G-BASE and NSI samples provides a large number of inorganic element analyses but not for Hg. Problems with assembling a data set for Hg in soils for England are described in Ander *et al.* (2012) and the problems for Wales are similar. Other smaller data sets have been used for Hg results including: Countryside Survey (CS - Emmett *et al*., 2010); UK Soil and Herbage Survey (SHS – Barraclough, 2007); Forum of Europeam Geological Survey geochemical atlas of Europe (FOREGS – Salminen *et al*., 2005); and EuroGeoSurveys Geochemical Mapping of Agricultural Soils (GEMAS – Reimann *et al*., 2012).

2.2 TOPOGRAPHICAL DATA SETS

All geographically referenced data have been explored using ESRI ArcGIS v9.3. Topographical data used is from the Ordnance Survey, for example, the polygon of Wales, used in many maps for this report, with a calculated land area of 21,224 km^2.

2.2.1 Catchment boundaries

Catchment boundaries have been used to explore contaminant distributions, in particular the catchments of the Loughor (Afon Llwchwr), Tawe (Cwm Tawe and more commonly referred to as the Swansea Valley), and Neath (Cwm Nedd) (Figure 2). In South Wales the 'Welsh valleys' draining the South Wales Coalfield have traditionally been associated with industry. A catchment-based approach to exploring contaminant distributions was not used in England.



Contains Ordnance Survey data © Crown copyright and database rights (2012)
Urban areas defined from OS Strategi® used under OS OpenData™ terms

Figure 2: Extent of 'Large urban areas' and Loughor, Tawe and Neath valleys

2.2.2 Urbanisation extent

The Generalised Land Use Database (GLUD) derived from the National Land Use Database (NLUD) (Office for National Statistics, 2006) was used to define urban areas in England (see

Ander *et al*. 2011), but is not available to define urban areas in Wales. Instead the Strategi® data set (Ordnance Survey, 2011). This data is free to view, download and use under OS OpenData™ terms. The 'Large Urban" classification of the Strategi® data is used to define the urbanisation extent and is shown in Figure 2.

2.3 GEOLOGICAL DATA SETS

2.3.1 Soil Parent Material Model

The BGS Soil-Parent Material Model[2] (SPMM) v6 (Lawley, 2011) has been developed using, as its basis, the mapped boundaries of the national 1:50,000 superficial and bedrock geological data (DigMapGB-50[3]). Soil 'Parent Material' is the first recognisably geological material found beneath a soil profile, and is the lithology on which that soil has developed (Avery, 1990). Soils thus inherit many properties, including chemical composition, from this material.

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

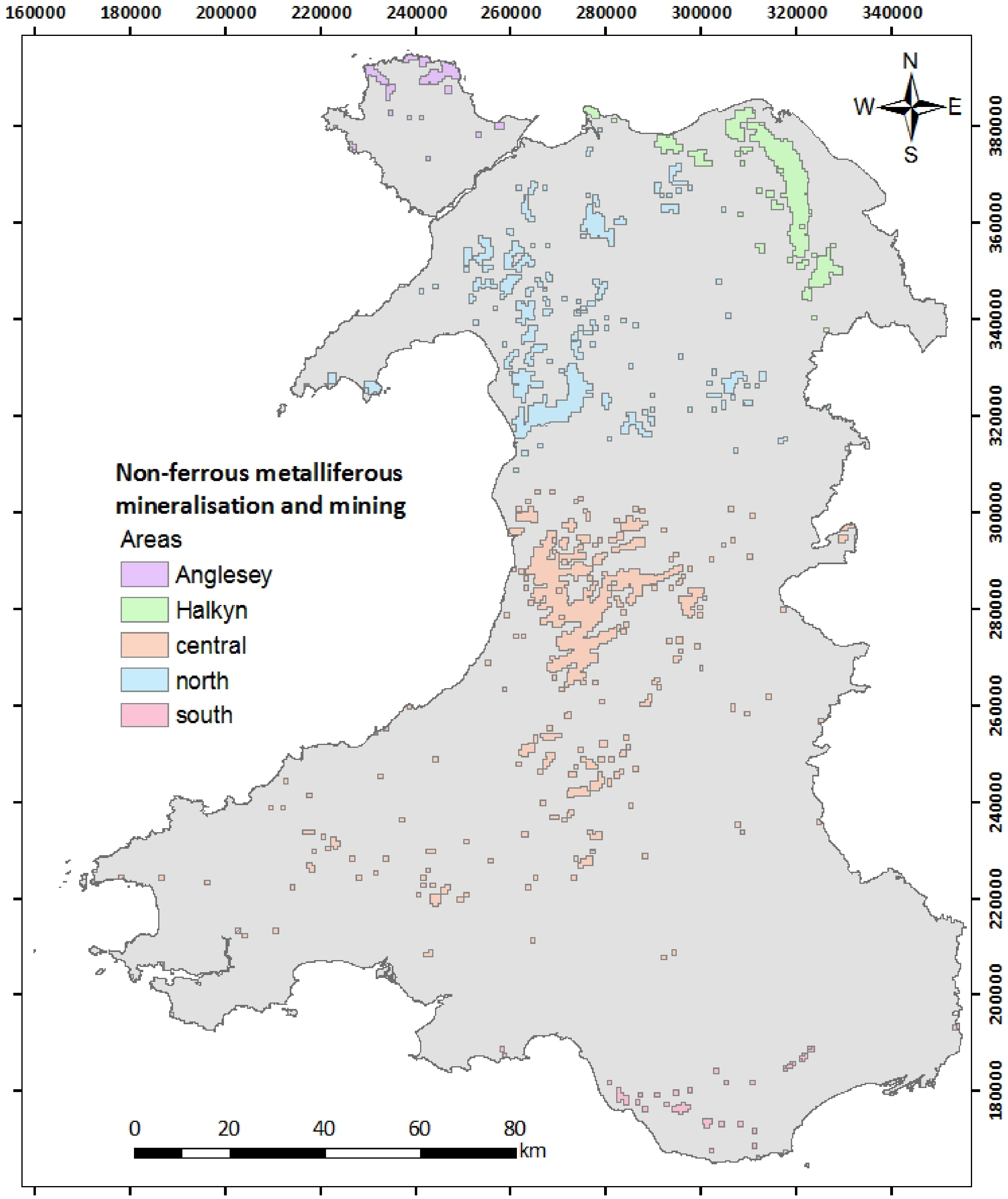
2.3.2 Non-ferrous metalliferous mineralisation and mining

The data set examined in this work is that of Metalliferous Mineralisation and Mining Database, originally produced in hard-copy by Ove Arup (1990) for Department of the Environment, but which has been 'cleaned' and turned into a polygon layer by BGS. The polygons have been further attributed in this study by giving names to the major Welsh ore-fields (Figure 3), to allow soil sample sites and geochemical data to be joined to the ore-fields and separately analysed for typical soil concentrations. The limitations and scale of use of this data set are discussed by Ander *et al*. (2011).

---

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

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



(after data from the Ove Arup database)

Figure 3: Non-ferrous metalliferous mineralisation and mining areas in Wales

3 Arsenic (As)

3.1 EXPLORATORY DATA ANALYSIS

Arsenic topsoil results are provided by the NSI(XRFS) and G-BASE urban data sets, which have been used as the primary data source in this work. The locations of data available from the NSI(XRFS) and G-BASE urban surveys are shown in Figure 1, and the summary statistics of these data in Table 1. Whilst it may appear that the urban data are significantly higher concentrations than the rural data (Figure 4), this is due entirely to the data from Swansea (Figure 5, Table 1), with Cardiff soil As concentrations falling within the range seen elsewhere in Wales. The data are mapped in Figure 6 showing the full range of concentrations. In order to focus on higher concentrations for NBC analysis, k-means cluster analysis thresholds are used in Figure 7, which highlight the areas of highest concentration. However, the systematically higher concentrations in the G-BASE Swansea data set, combined with the large number of samples this comprises, has the effect of spatially skewing the data – thus the NSI data only are plotted using k-means cluster analysis in Figure 8.

A baseline survey of northwest Pembrokeshire found a mean soil As concentration of 20 mg/kg (260 samples) (Bradley *et al.*, 1978), but also noted a tendancy for higher concentrations (unspecified value) over the Coal Measure shales. This is very close to the NSI(XRFS) median of 20 mg/kg. Archer and Hodgson (1987) reported a median concentration for England and Wales of 10 mg/kg.

(a) As (mg/kg)	**Number**	**Mean**	**Minimum**	**25th percentile**	**Median**	**75th percentile**	**Maximum**	**Skewness**
All data	1675	36.2	5.69	15.0	20.2	33.4	2062	17

(b) As (mg/kg) Area type	**Number**	**Mean**	**Minimum**	**25th percentile**	**Median**	**75th percentile**	**Maximum**	**Skewness**
NSI(XRFS)	798	28.5	5.69	14.8	19.7	26.4	826	11
Urban	877	43.1	6.06	15.1	21.2	45.3	2062	15

(c) As (mg/kg) Area name	**Number**	**Mean**	**Minimum**	**25th percentile**	**Median**	**75th percentile**	**Maximum**	**Skewness**
Cardiff	507	18.3	6.06	13.1	16.1	20.2	151	6
Swansea	370	77.2	8.07	32.3	52.4	81.6	2062	11

Note that here, and throughout this report, concentration units are given as mg/kg . This represents the number of mg of the element being discussed present in 1 kg of dry sieved soil.

Table 1: Statistical summary of topsoil As in the main data sets (a) whole data set (b) by sampling density (c) by urban centres sampled by G-BASE

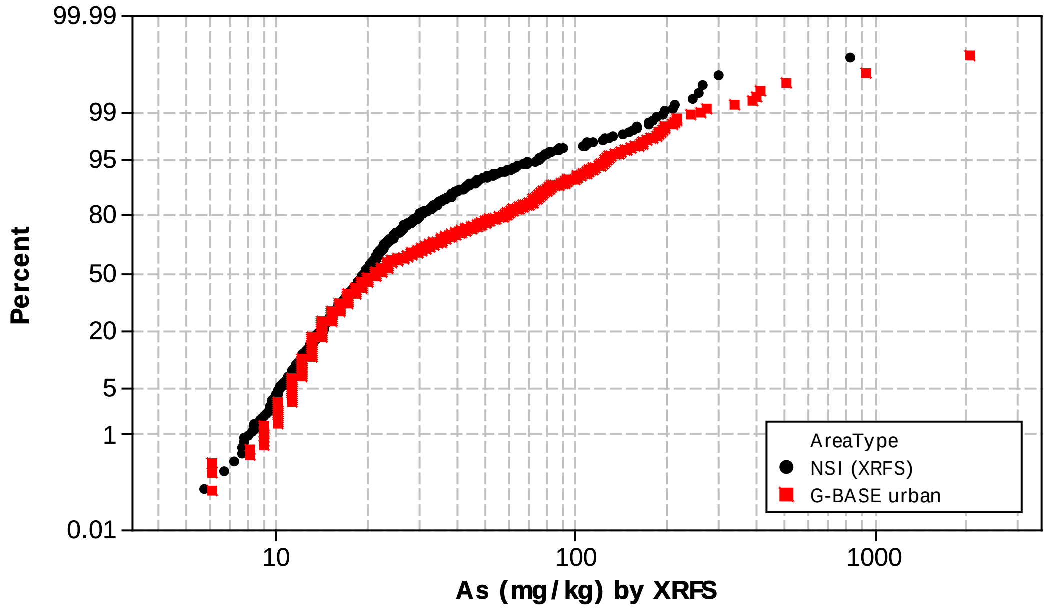
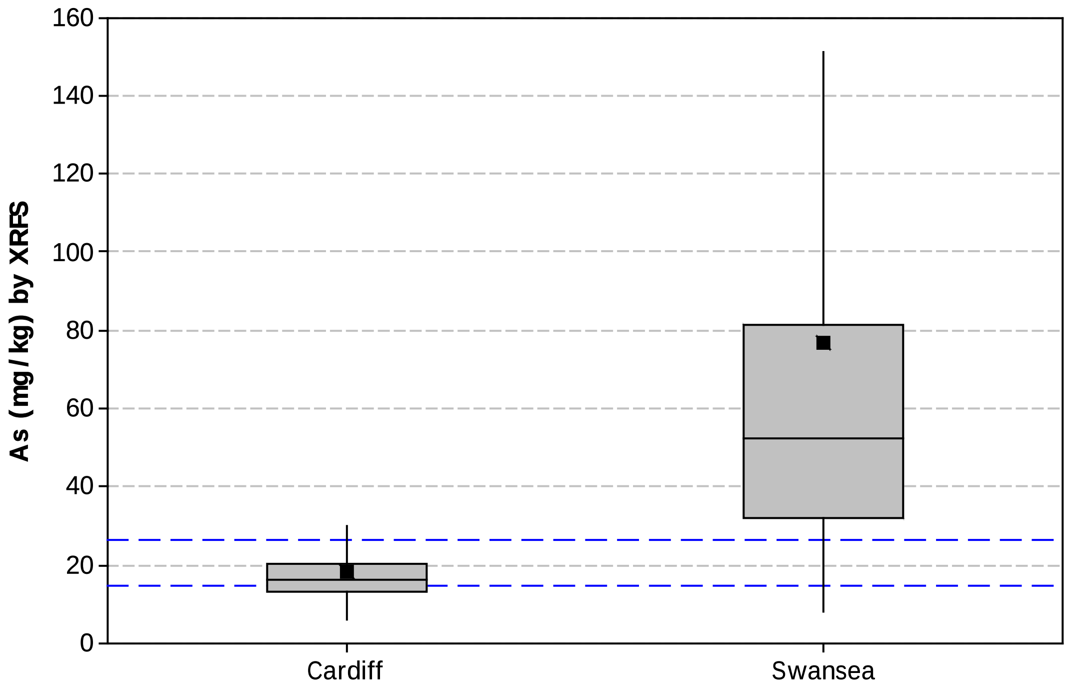
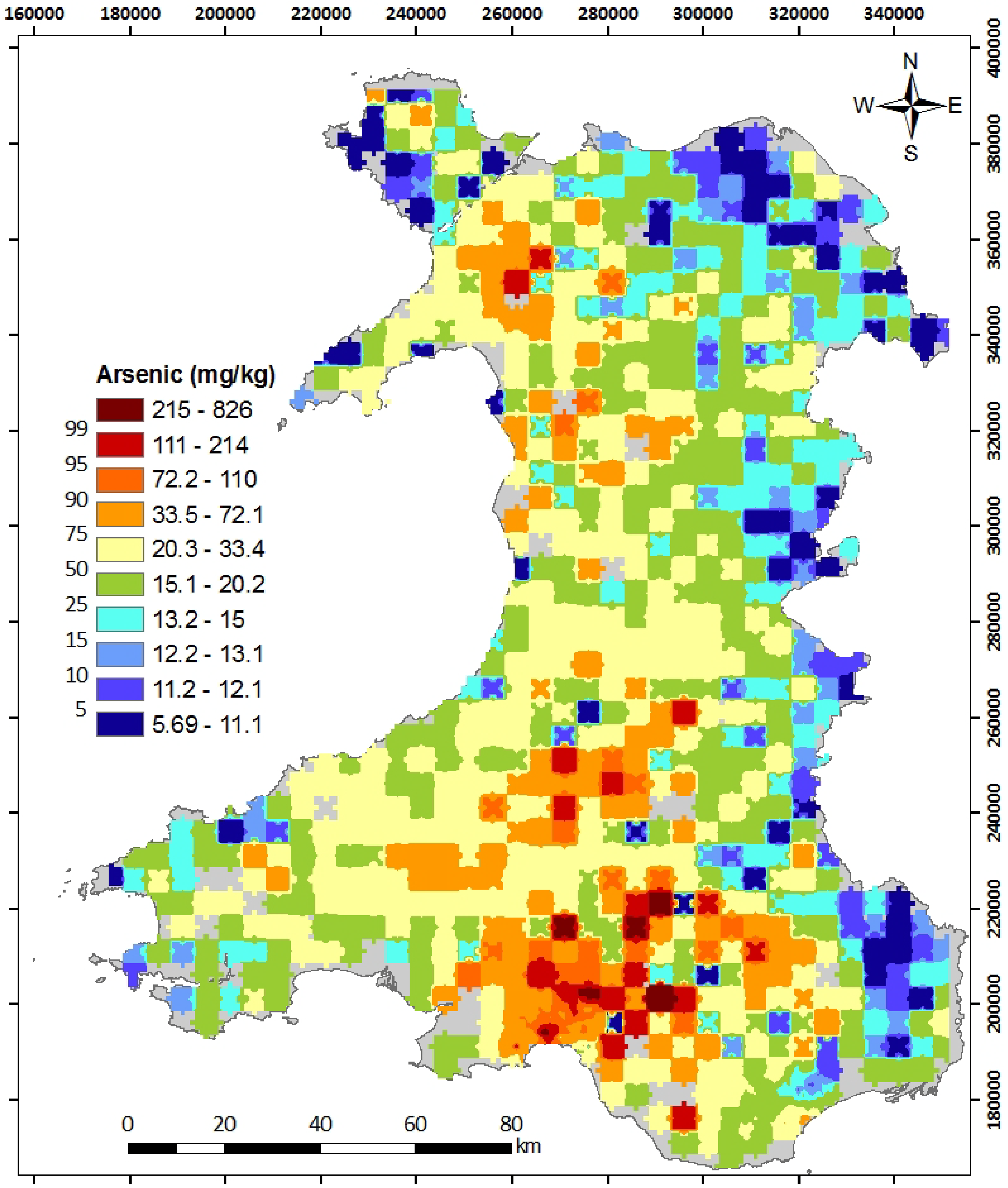


Figure 4: Cumulative probability plot of topsoil As concentrations for G-BASE urban and NSI(XRFS) results



Blue dashed lines indicate the interquartile range of the national NSI(XRFS) data. The boxes show the interquartile range with the median as a line within the box, black square represent the mean.

Figure 5: Boxplot of urban topsoil As results from the two G-BASE urban areas



Colour thresholds are designed for highly skewed data

Figure 6: Interpolated map of all topsoil As data

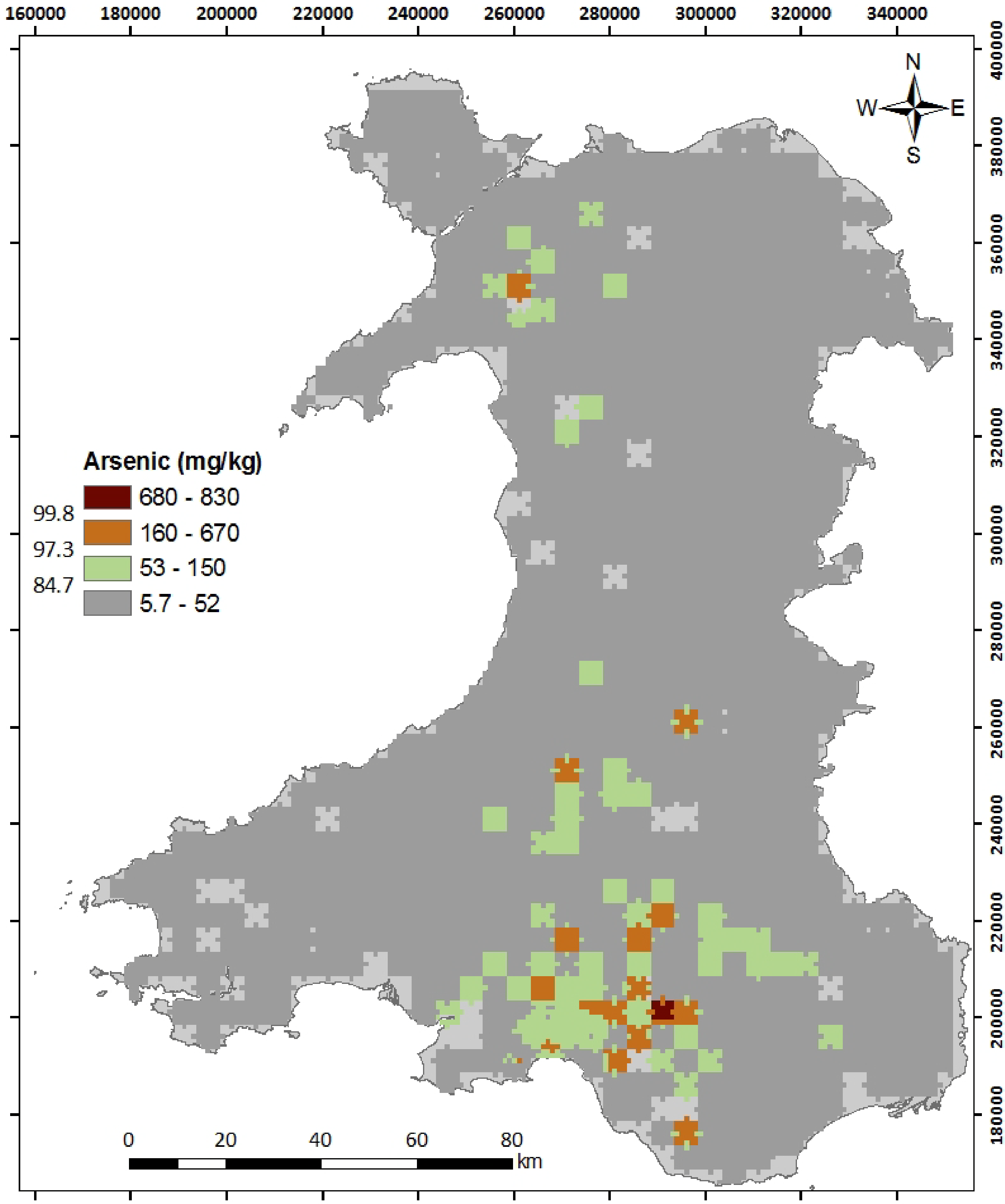


Figure 7: Interpolated map of topsoil As concentrations. Thresholds determined by using k-means cluster analysis

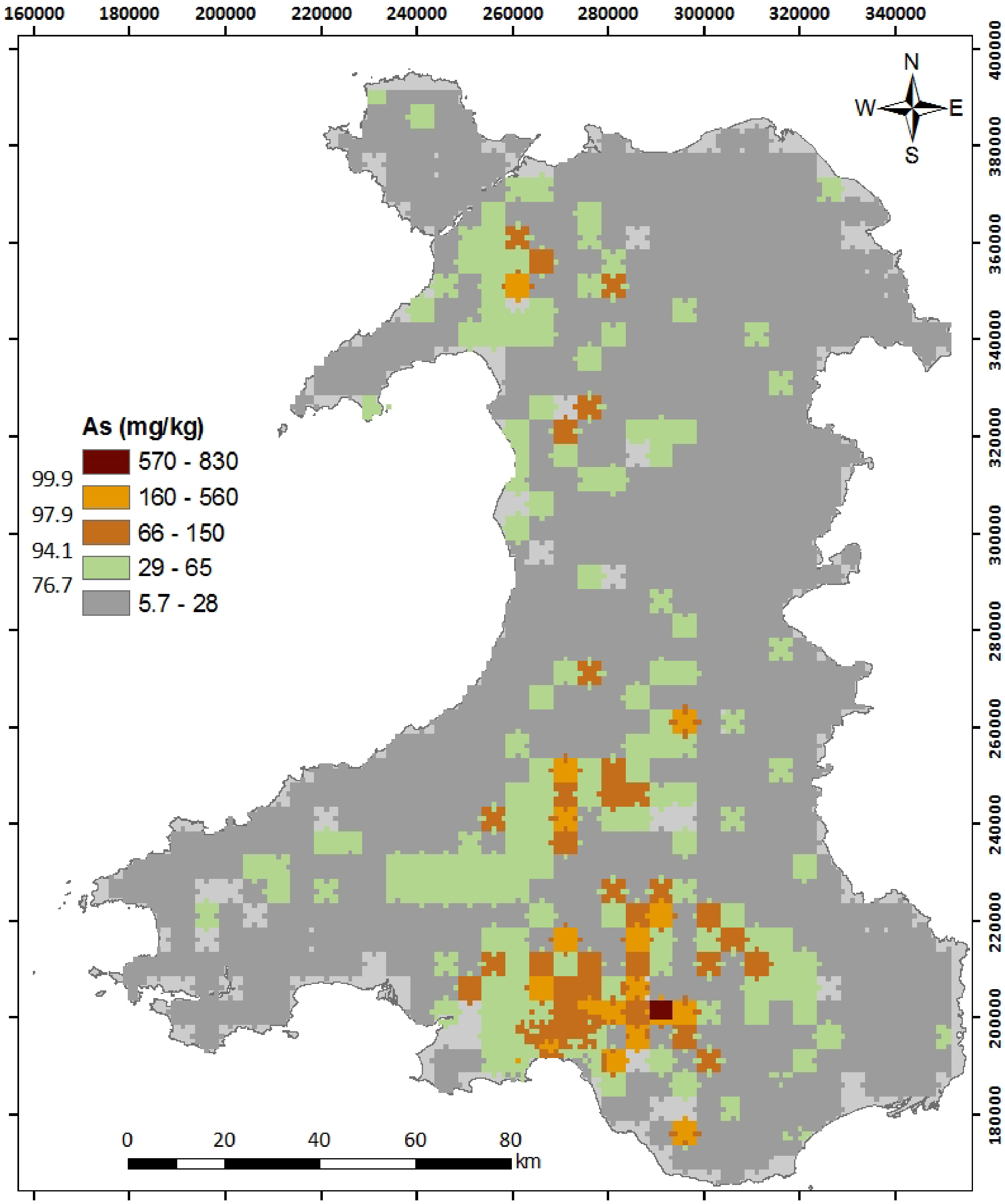
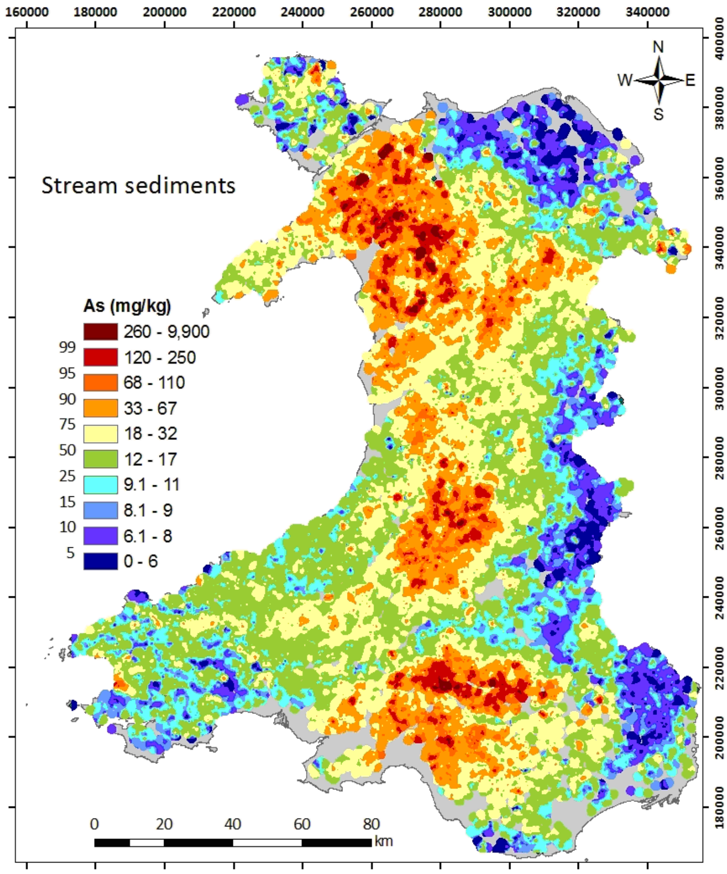
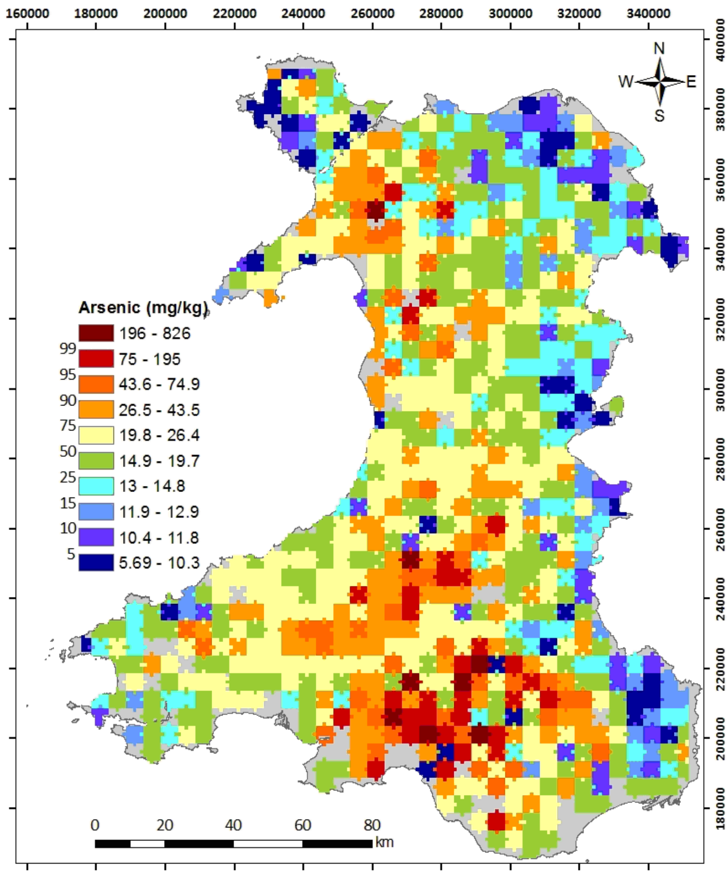


Figure 8: Interpolated map of topsoil As concentrations from the NSI(XRFS) data set only. Thresholds determined by using k-means cluster analysis



(a)



(b)

Figure 9: Arsenic in: (a) G-BASE stream sediments and (b) NSI(XRFS) soils

3.2 DOMAIN SELECTION

In the previous sections, data are used to establish areas of typically higher concentration of topsoil As. However, due to the low density of soil data and highly variable geological environments (including non-ferrous metalliferous mineralisation) found across Wales, G-BASE stream sediment results (British Geological Survey, 2000) have also been explored to allow comparison with the NSI(XRFS) soil data (Figure 9). The low availability of systematically collected soil data for Wales means that a greater reliance has to be given to other sample media in order to define environments high in contaminant concentrations. The stream sediment results for As (average sample density of 1 sample every 2 km^2) readily establishes higher environmental background concentrations in mineralised areas of Wales, as well as in the South Wales Coalfield (British Geological Survey, 2000), which are less well defined by the NSI(XRFS) soil sampling density. The stream sediment concentration information has therefore been used to help define key domains.

Domains investigated are:

- urban/industrial activity
- non-ferrous metalliferous mineralisation and mining

3.2.1 Urban areas

As was seen in the English As data exploration (Ander *et al.*, 2011), there is no indication that urbanisation, *per se*, systematically increases As concentrations (Table 1). However, there is evidence that Swansea has higher concentrations, which may be expected to reflect an even wider area of the Swansea valley, where there was smelting of Cu (and later Zn) (Hughes and Reynolds, 1992). This expectation of higher concentrations is also supported by regionally elevated stream sediment concentrations over the South Wales Coalfield (Figure 9).

The lack of generally elevated topsoil As concentrations in urban areas is also confirmed when the data are compared with the intensity of urbanisation. Figure 10 shows that topsoils classified as being from Swansea is a more important association for the high As soils than the intensity of urbanisation classification. However, when the sample sites are within three selected valleys in the west of the South Wales Coalfield, it can be seen that these have higher concentrations than the rest of the available data (Figure 11 and Table 2).

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Loughor	40	68.5	15.2	34.7	48.4	87.7	262	2
Tawe	138	98.5	8.1	39.2	64.5	101	2062	8
Neath	164	72.9	5.7	29.5	54.4	89.1	410	2
All other	1333	24.2	6.1	14.1	18.1	24.2	826	14

Table 2: Summary statistics for topsoil As (mg/kg) classified by selected South Wales valleys

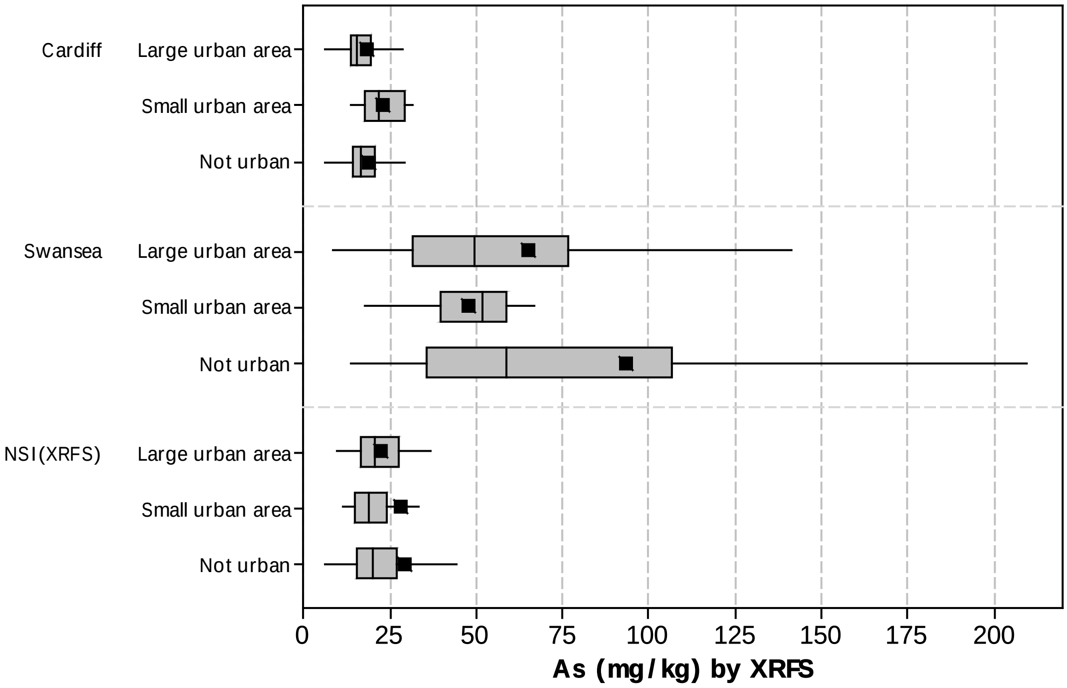


Figure 10: Topsoil As boxplot by urbanisation extent (OS Strategi® classification) and by project source

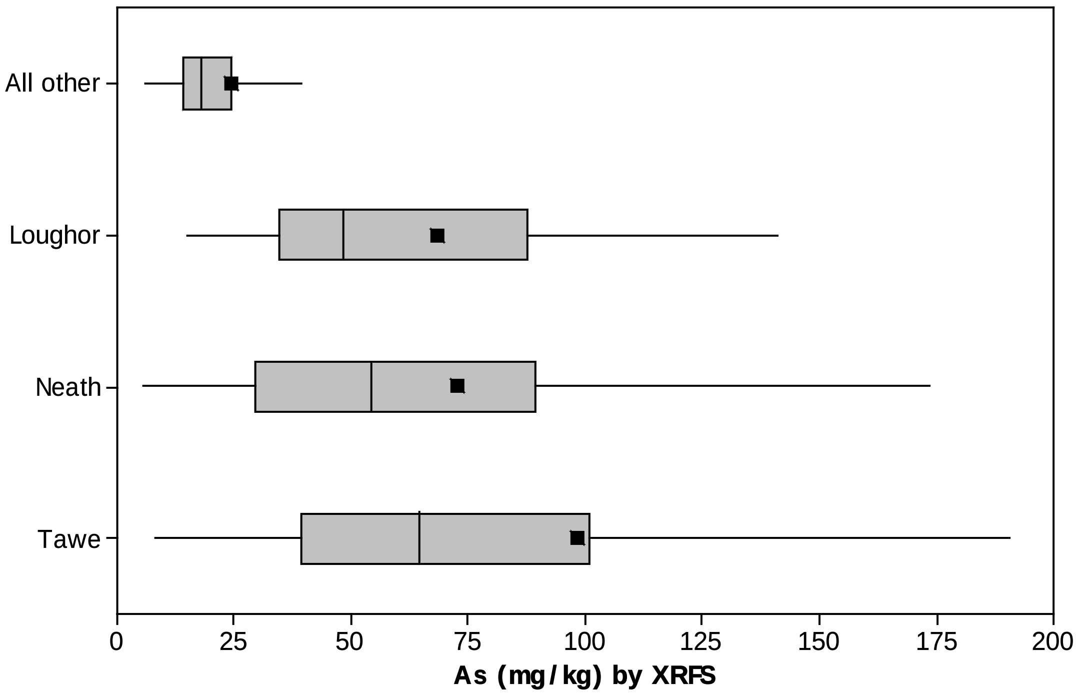


Figure 11: Boxplot of topsoil As concentrations in selected South Wales valleys

3.2.2 Non-ferrous metalliferous mineralisation and mining

Arsenic has previously been recognised as occurring at higher natural background concentrations over mineralised areas of Wales (British Geological Survey, 2000). However, the

NSI(XRFS) data does not show such regional distributions of concentration (Figure 9) and when the data are co-located with the records of metalliferous mineralisation and mining there are very few sample sites (Table 3). The majority of these have As concentrations which are close to that of the remainder of the data set (see also Figure 12). Because of the low sample numbers, the data are combined into all other than Halkyn (where the mineralisation does not include accessory As (British Geological Survey, 2000)) to create one proposed domain 'mineralisation'.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Anglesey	3	21.6	8.72	8.72	17.6	38.4	38.4	1.1
north	22	42.7	10.8	18.7	30.9	52.9	210	3.3
Halkyn	15	13.8	9.4	11.4	12.4	16.3	20.8	0.8
central	31	26.4	14.2	17.9	21.7	27.5	85.8	2.8
south	7	40.7	14.1	18.1	23.2	28.2	159	2.6
none	*1597*	*36.5*	*5.69*	*14.8*	*20.2*	*34.2*	*2060*	*17*

Table 3: Summary statistics for topsoil As (mg/kg) by ore-field areas

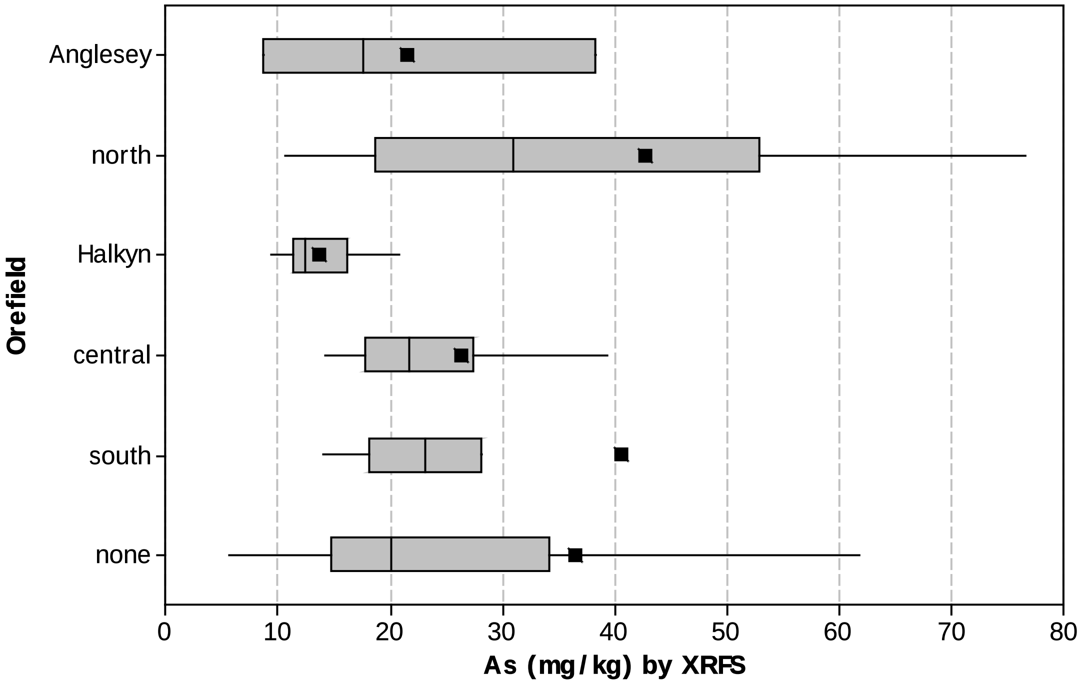


Figure 12: Boxplot of mineralised areas topsoil As concentrations

3.3 DOMAIN DATA SUMMARY

Data have been assigned to the Urban 1 (the three South Wales valleys) and Mineralisation Domains, for which the summary statistics (Table 4) and plots (Figure 13 and Figure 14) show the data distributions. All other samples not in these two domains are assigned to the Principal Domain. This shows that the Urban 1 Domain has the highest typical concentrations from the available data, whilst the Mineralisation Domain contains slightly higher soil As concentrations compared to the rest of the data from the Principal domain. The concentrations associated with the mineralised areas may turn out to be higher when further higher density systematic sampling is undertaken, as suggested by the BGS high density stream sediment survey (British Geological Survey, 2000). A map of the domains is shown in Figure 15.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Mineralisation	63	33.4	8.72	18.0	23.2	33.4	210	4
Urban 1	342	82.7	5.69	34.5	57.4	93.2	2062	11
Principal	1270	23.8	6.06	14.1	18.1	23.6	826	15

Table 4: Summary statistics for the As (mg/kg) domain data

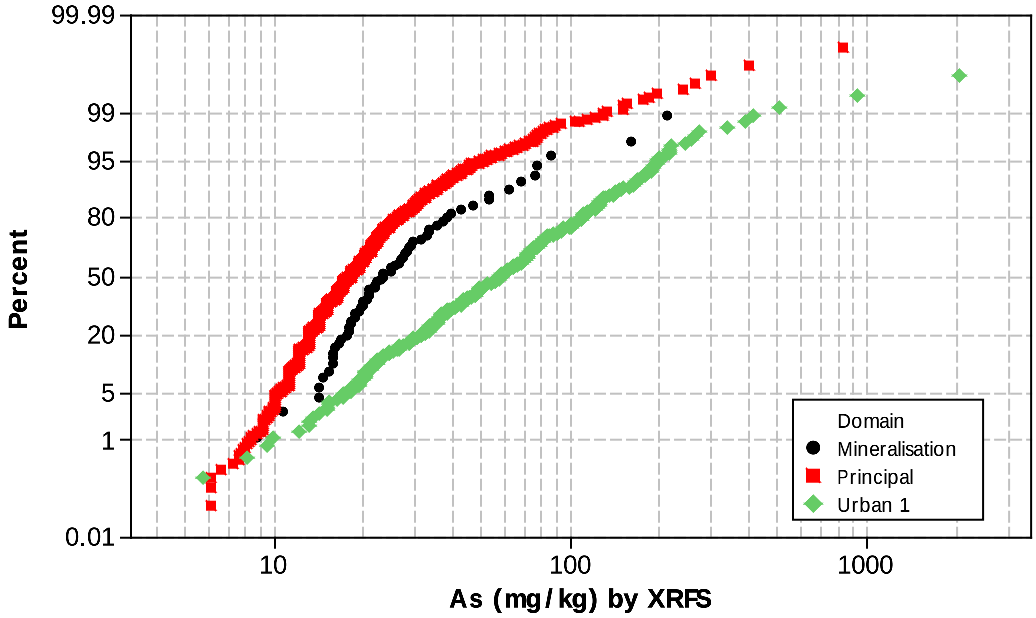


Figure 13: Probability plot of As topsoil data classified by domains

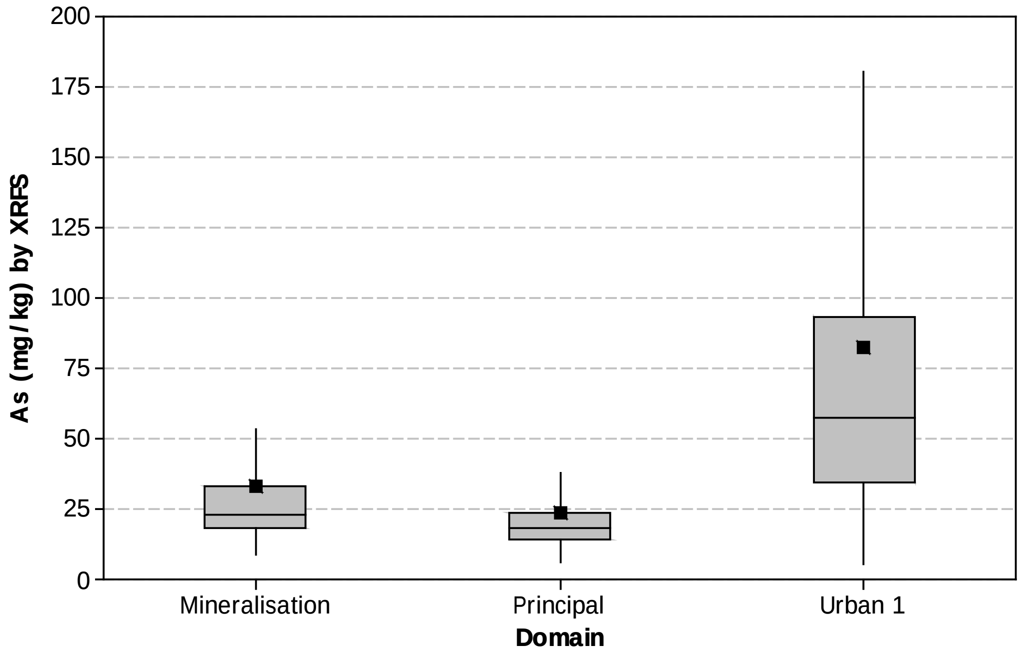
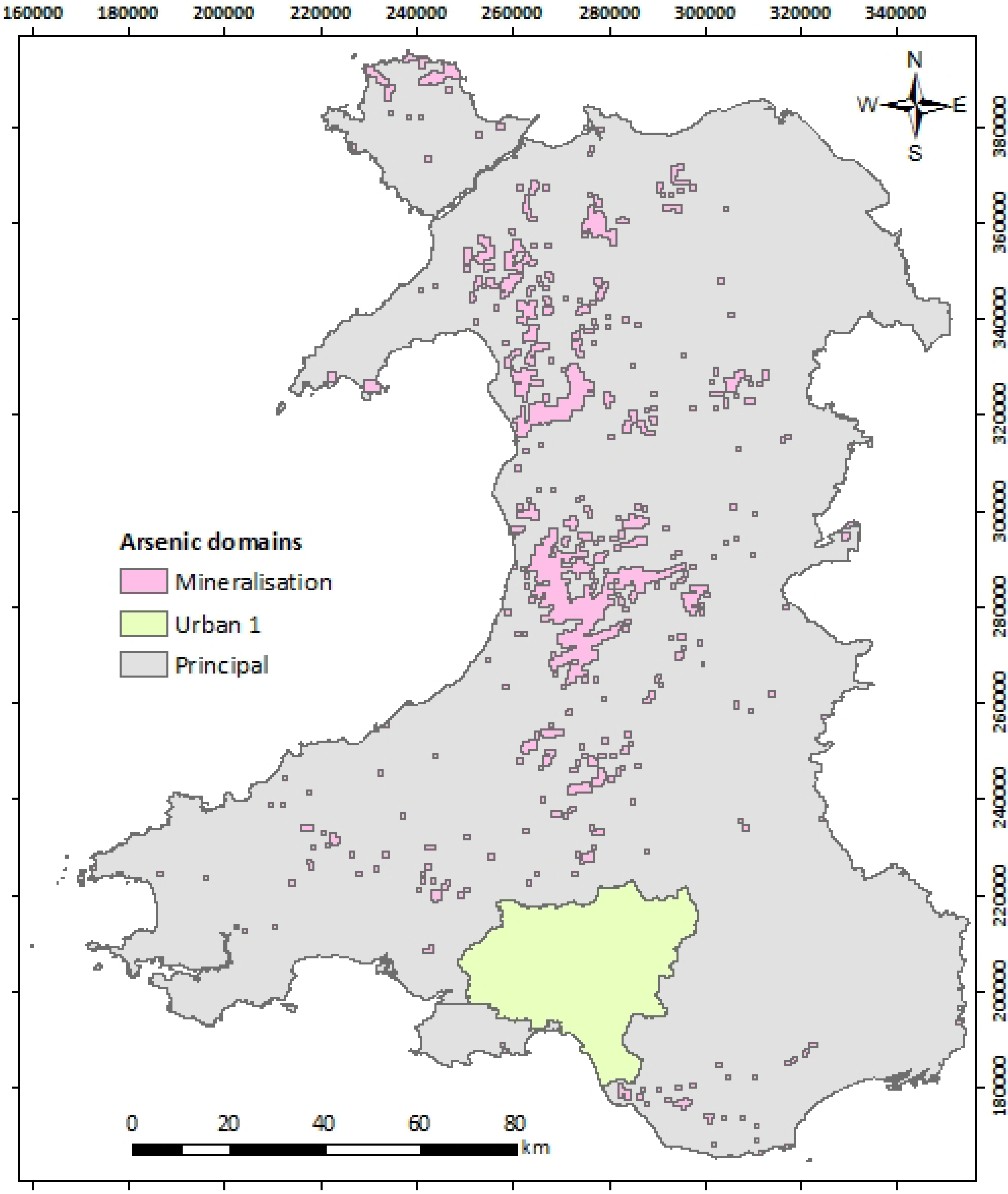


Figure 14: Boxplot of As topsoil data classified by domains



Contains Ordnance Survey data © Crown copyright and database rights (2012)
Urban areas defined from OS Strategi® used under OS OpenData™ terms
Mineralisation after the Ove Arup data

Figure 15: Map showing the location of As domains

3.4 As NBC CALCULATION

The NBC calculations for As have been carried out according to the method described in Cave *et al.* (2012).

3.4.1 Principal Domain

Untransformed skewness and octile skew coefficients fail the test criteria for underlying Gaussian distribution. The log transformation (log_e) produces a skewness coefficient (SC) that is >1, but octile skew (OS) is <0.2 suggesting a log distribution with outliers (Figure 16 and Figure 17). Therefore, the robust statistics method of percentiles calculation is used (Table 9) to give a Principal Domain NBC in Figure 18.

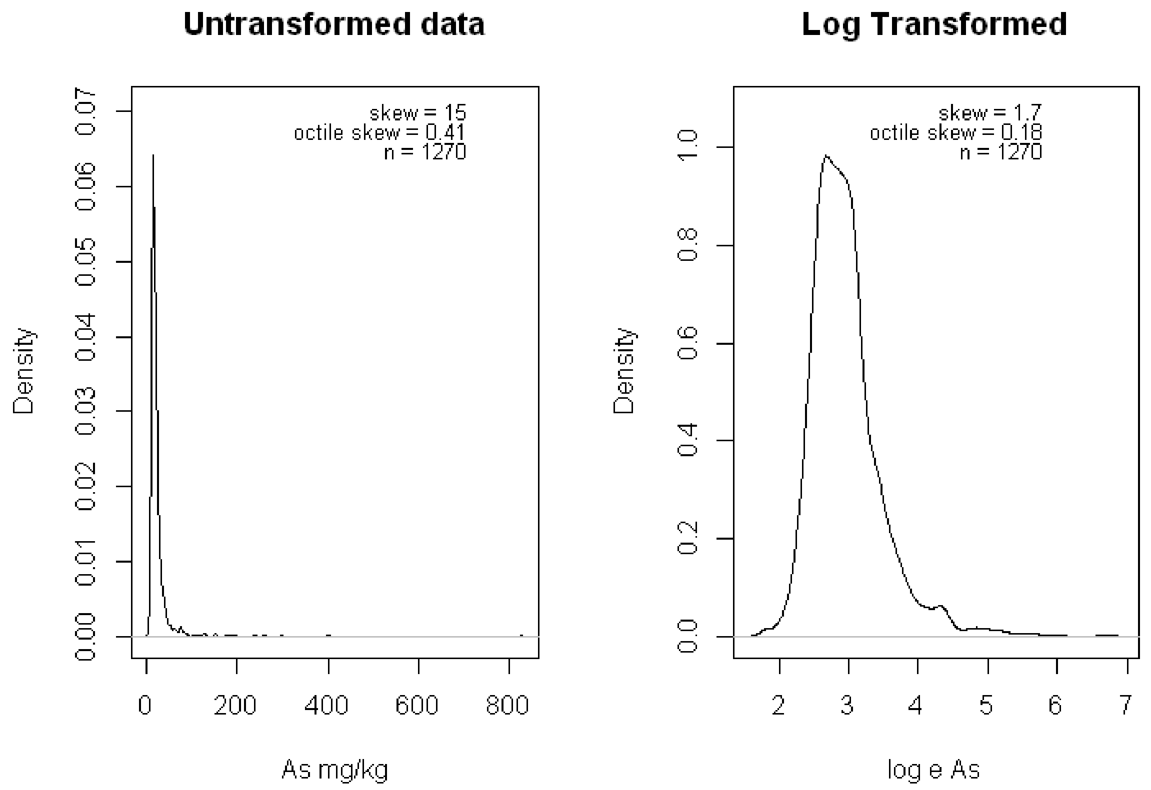


Figure 16: Density distributions for the raw data and the log_e transformed data for As in the Principal Domain (n = number of samples)

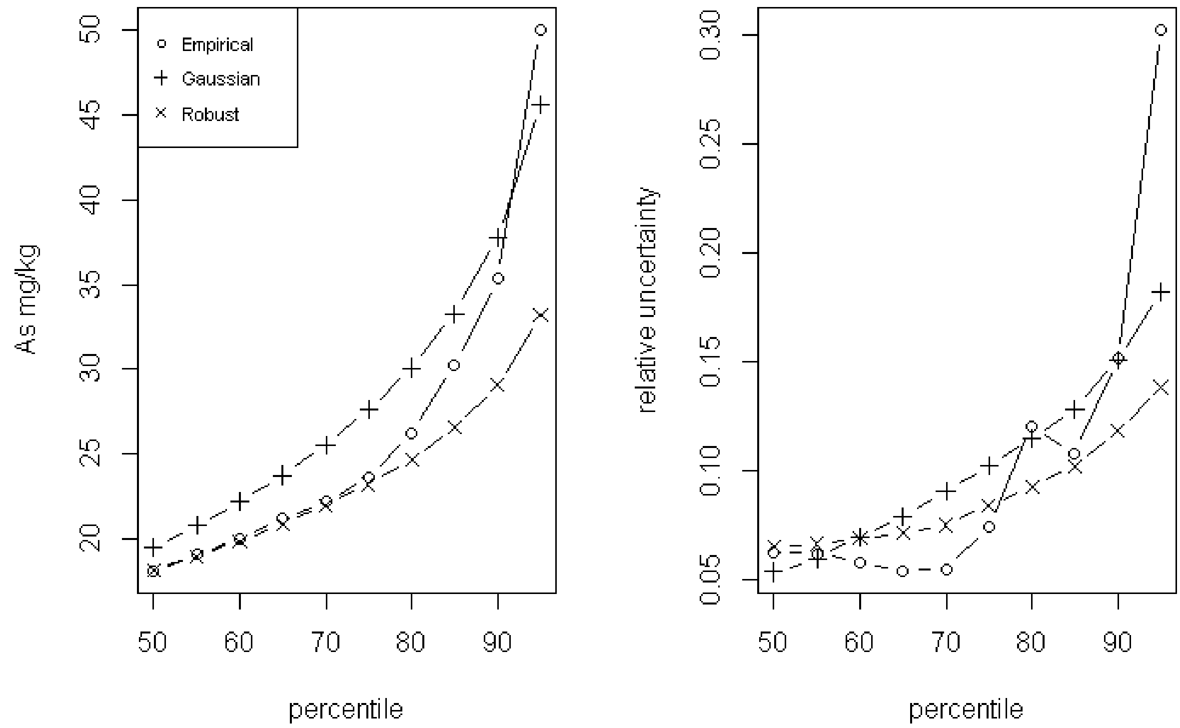


Figure 17: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for As in the Principal Domain

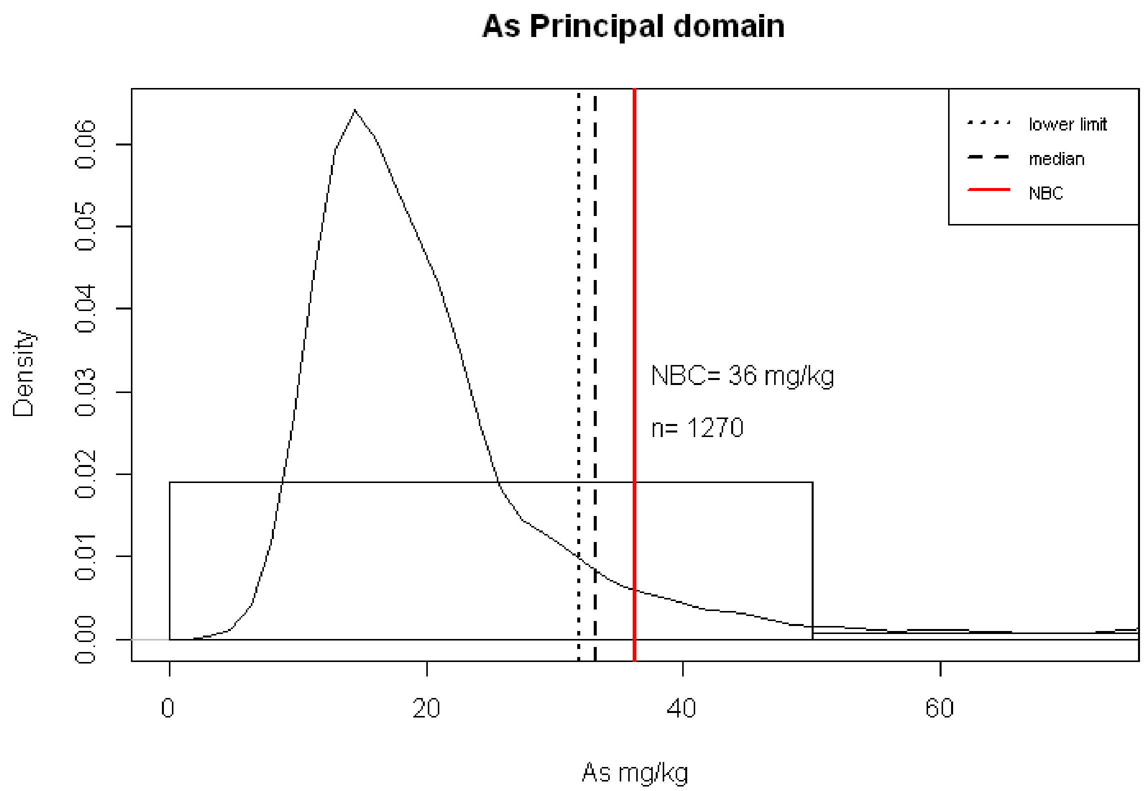


Figure 18: Summary density plot and histogram of the distribution for As in the Principal Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	17	18	18	19	19	20	**17**	**18**	**18**
55	18	19	19	20	21	21	**18**	**19**	**19**
60	19	20	20	21	22	23	**19**	**20**	**20**
65	20	21	21	23	24	25	**20**	**21**	**21**
70	21	22	23	24	25	27	**21**	**22**	**23**
75	23	24	25	26	28	29	**22**	**23**	**24**
80	25	26	28	28	30	32	**24**	**25**	**26**
85	28	30	32	31	33	35	**25**	**27**	**28**
90	33	35	38	35	38	41	**28**	**29**	**31**
95	44	50	58	42	46	50	**32**	**33**	**36**

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 5: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for As (mg/kg) in the Principal Domain

3.4.2 Mineralisation Domain

Untransformed and $\log_e$ transformed skewness and octile skew coefficients fail the test criteria for underlying Gaussian distribution. Therefore, a Box-Cox transformation was applied (Figure 19 and Figure 20). However, for the Box-Cox transform the uncertainties are high and unstable particularly for the robust data (Figure 21 and Table 6). This is a consequence of the low number of samples and two outliers. Using $\log_e$ transform gives more stable results (*cf*. Figure 21 and Figure 22) which gives a more plausible result with less weight to the outliers.

Percentiles using Box-Cox (Table 6) and using $\log_e$ transform (Table 7) are given to permit comparison of the relative effects of the data transformations. The suggested NBC has been derived from the $\log_e$ transformation (Figure 23). There is a clear need for more data for this domain.

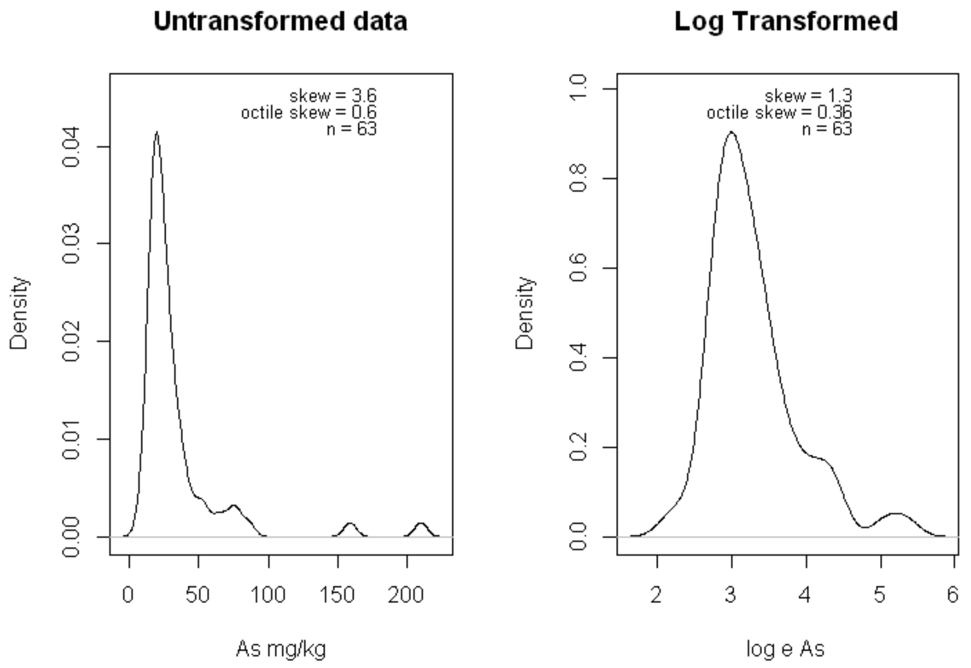


Figure 19: Density distributions for the raw data and the $\log_e$ transformed data for As in the Mineralisation Domain (n = number of samples)

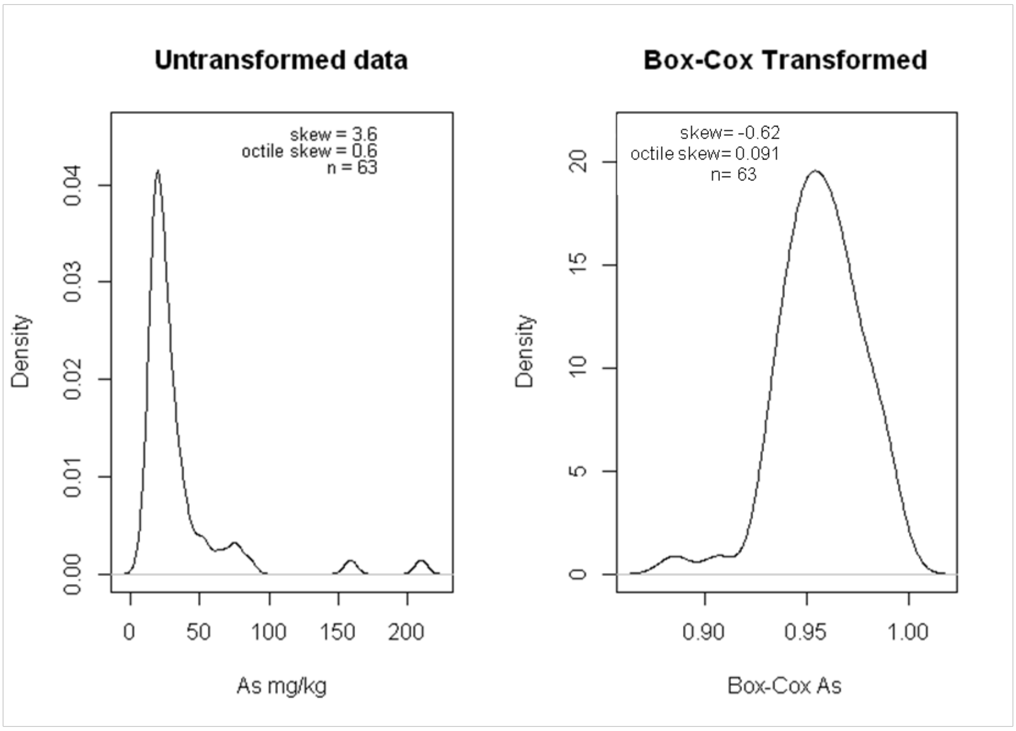


Figure 20: Density distributions for the raw data and the Box-Cox transformed data for As in the Mineralisation Domain (n = number of samples)

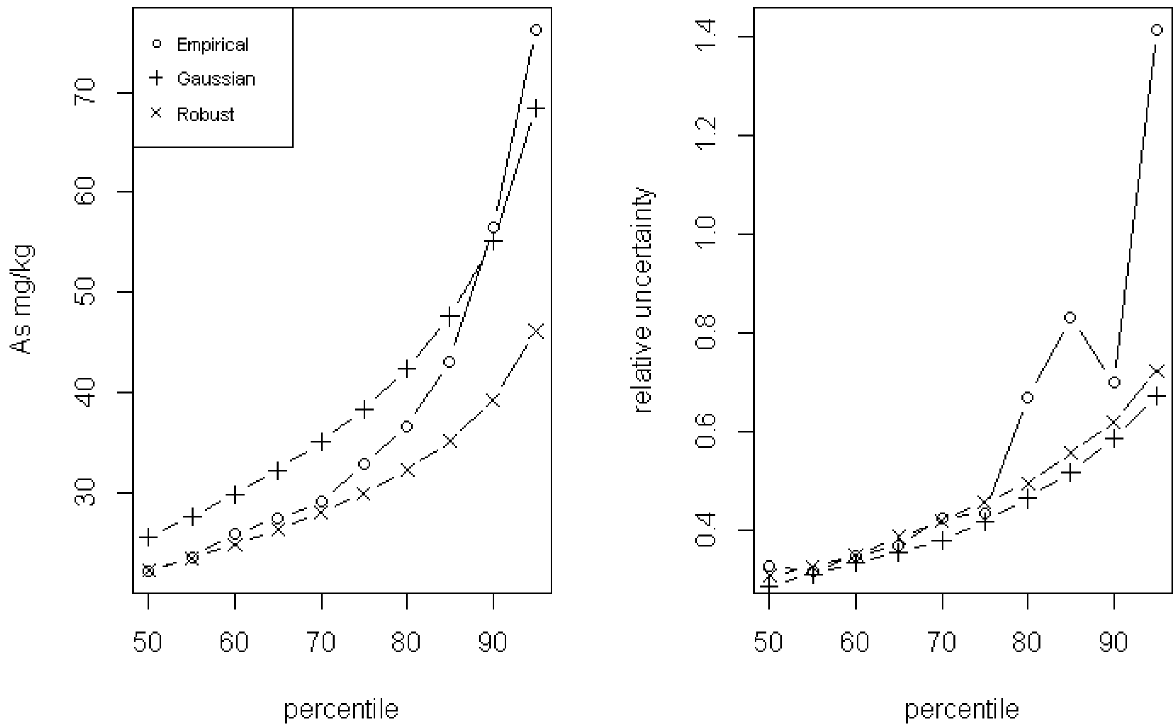


Figure 21: Comparison of $\log_e$ empirical, Gaussian and Robust percentiles and relative uncertainty for As in the Mineralisation Domain

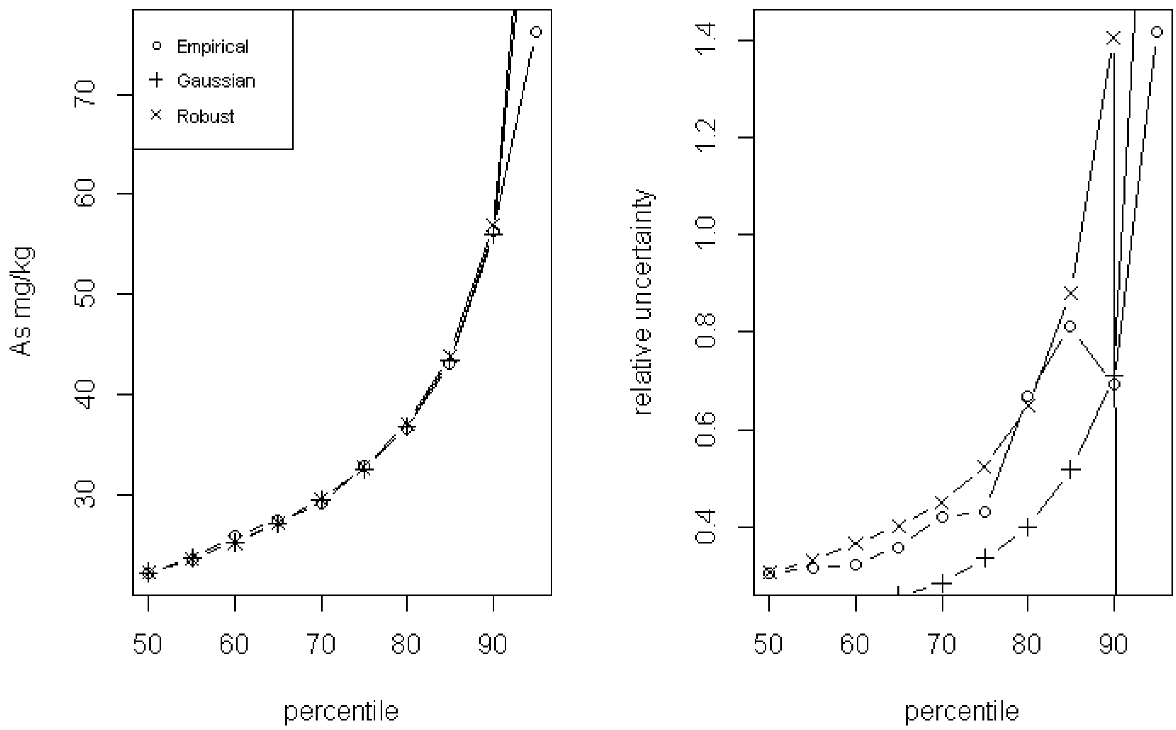


Figure 22: Comparison of Box-Cox empirical, Gaussian and Robust percentiles and relative uncertainty for As in the Mineralisation Domain

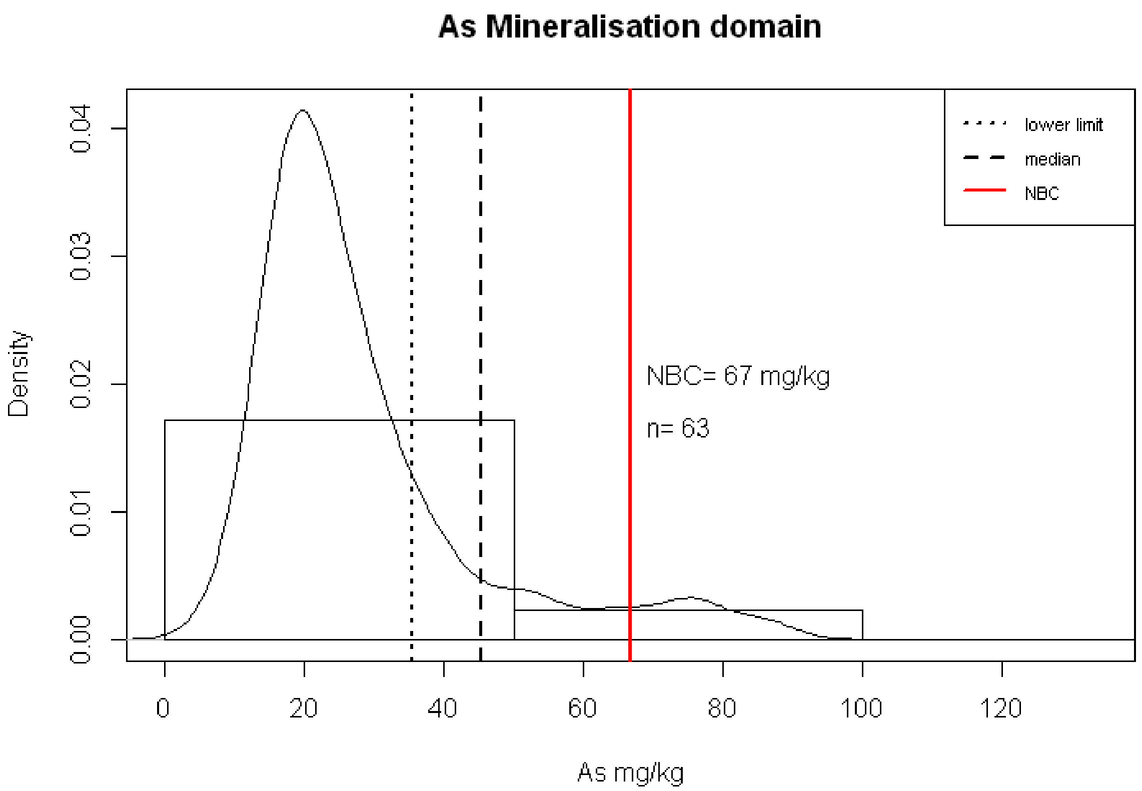


Figure 23: Summary density plot and histogram of the distribution for As in the Mineralisation Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	19.8	22.2	26.6	20.0	22.3	25.0	19.8	22.2	26.6
55	20.7	23.6	28.1	21.1	23.7	26.7	20.8	23.6	28.5
60	21.7	25.9	30.2	22.6	25.3	28.6	21.9	25.3	30.7
65	23.0	27.4	33.0	24.2	27.2	31.1	23.2	27.2	33.6
70	24.6	29.1	37.3	25.9	29.5	34.1	24.7	29.6	37.5
75	26.8	32.9	41.0	27.9	32.6	38.5	26.1	32.7	42.8
80	28.2	36.7	52.7	30.9	36.8	45.0	28.3	37.1	52.3
85	32.7	43.1	67.8	34.7	43.4	55.7	31.4	43.9	70.9
90	37.0	56.4	79.2	40.8	56.0	83.0	36.1	57.0	126
95	51.2	76.2	159	55.2	98.0	298	44.7	102	-1565

Low (L) and High (H) values represent the 95% confidence intervals around the median

Table 6: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for As (mg/kg) in the Mineralisation Domain using a Box-Cox transformation

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	21	23	27	23	27	31	**21**	**23**	**27**
55	21	25	28	25	29	34	**22**	**24**	**29**
60	22	27	33	26	31	37	**23**	**26**	**31**
65	24	28	34	28	34	40	**24**	**27**	**33**
70	26	30	38	30	36	44	**25**	**29**	**36**
75	27	33	45	32	40	49	**26**	**31**	**39**
80	29	38	53	34	44	55	**27**	**33**	**43**
85	33	46	68	37	49	64	**29**	**35**	**48**
90	38	60	82	42	57	76	**32**	**39**	**55**
95	52	76	159	49	71	98	**36**	**45**	**67**

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 7: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for As (mg/kg) in the Mineralisation Domain using a $\log_e$ transformation

3.4.3 Urban 1 domain

Untransformed skewness and octile skew coefficients fail the test criteria for underlying Gaussian distribution. When a $\log_e$ transformation was used, the skewness coefficient was <1 and octile skew <0.2, indicating an underlying Gaussian distribution, for which parametric percentiles were used (Figure 24 and Figure 25). Percentiles are given in Table 8 and an example NBC is shown in Figure 26.

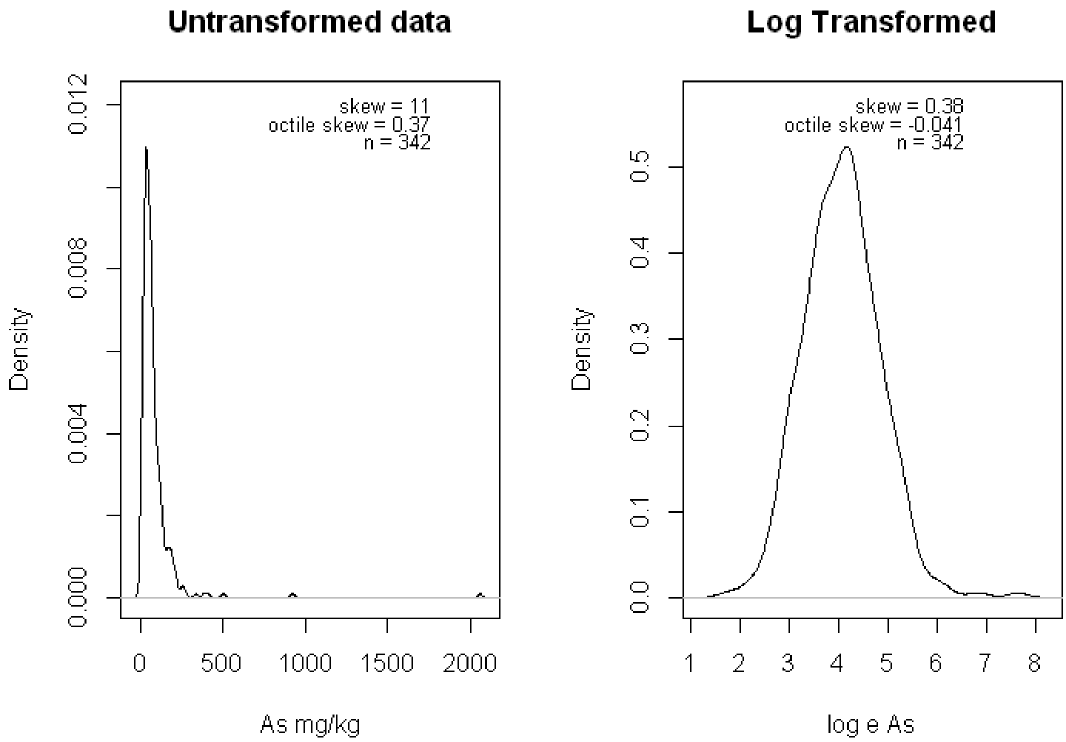


Figure 24: Density distributions for the raw data and the $\log_e$ transformed data for As in the Urban 1 Domain (n = number of samples)

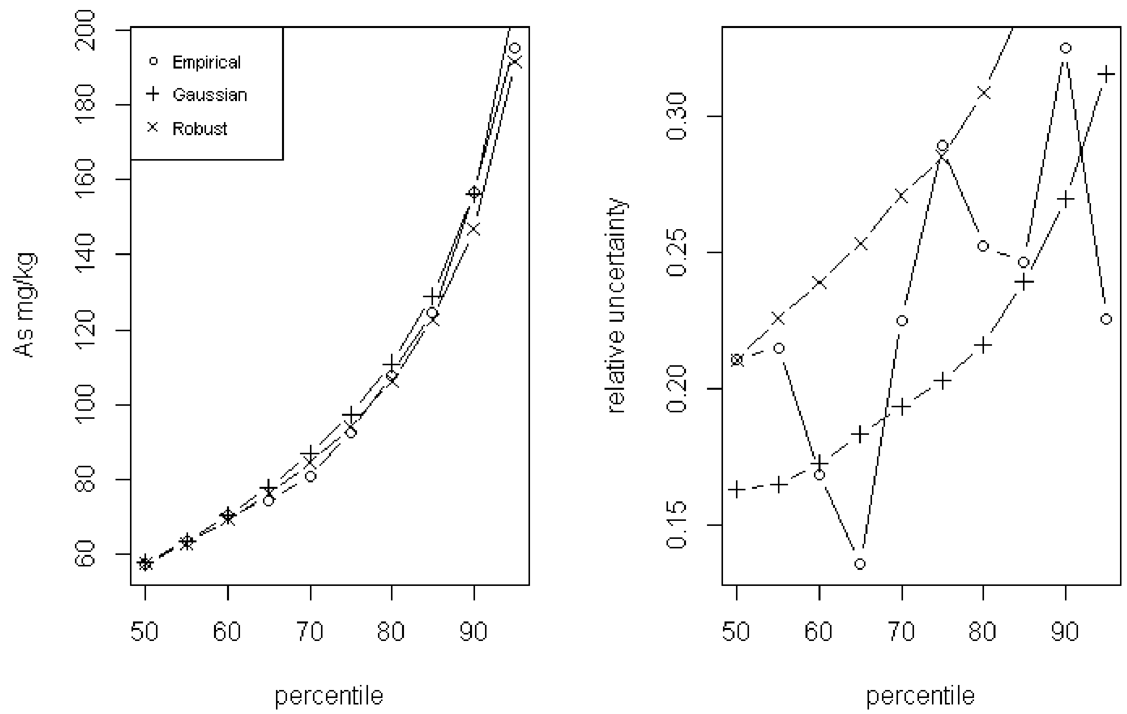


Figure 25: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for As in the Urban 1 Domain

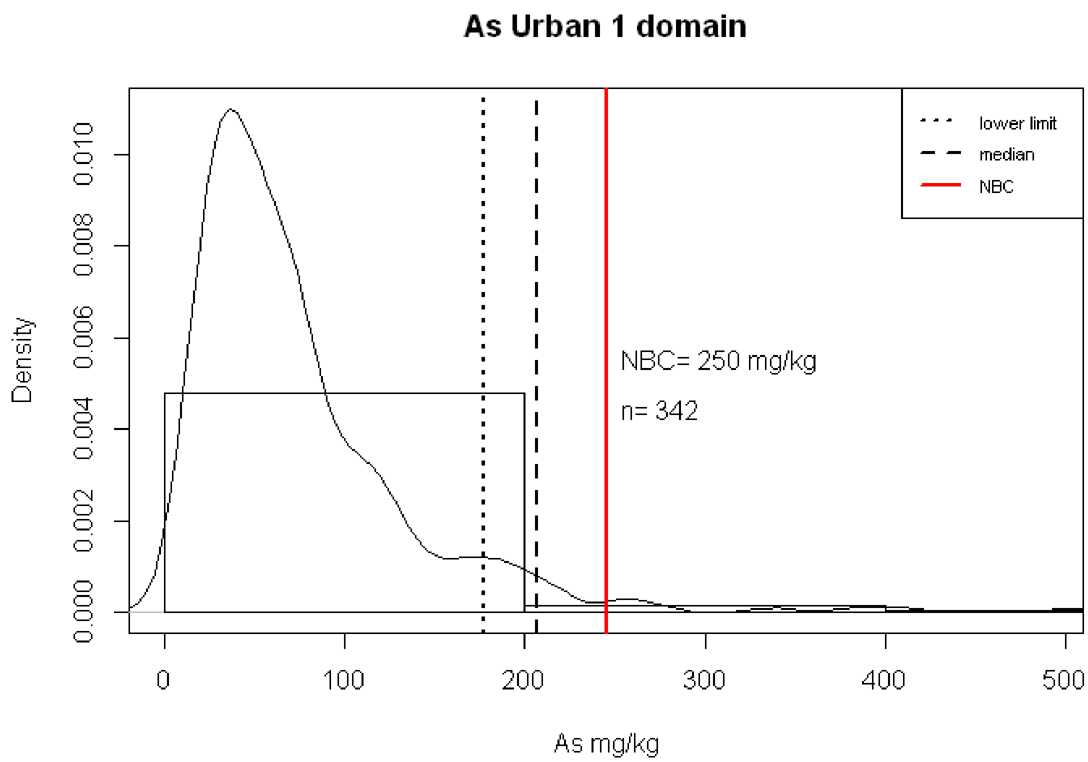


Figure 26: Summary density plot and histogram of the distribution for As in the Urban 1 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	**P L**	**Parametric**	**P H**	R L	Robust	R H
50	51.4	57.4	63.5	**53.5**	**57.7**	**62.7**	51.4	57.4	63.5
55	57.4	63.5	71.5	**58.8**	**63.6**	**69.4**	55.9	63.0	70.1
60	62.7	70.5	74.6	**64.7**	**70.2**	**77.0**	60.9	69.1	77.5
65	70.2	74.2	80.6	**71.3**	**77.8**	**85.7**	66.6	76.2	86.0
70	73.5	80.6	92.7	**79.0**	**86.7**	**95.9**	73.2	84.3	95.9
75	80.6	92.7	108	**88.2**	**97.4**	**109**	80.9	94.1	108
80	94.7	108	122	**100**	**111**	**125**	90.4	106	124
85	110	125	141	**115**	**129**	**147**	103	123	145
90	126	157	179	**137**	**156**	**180**	121	147	178
95	173	195	217	**177**	**207**	**245**	155	192	240

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 8: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for As (mg/kg) in the Urban 1 Domain

3.4.4 Arsenic NBC summary

A summary of all the NBC percentiles for each domain is shown in Table 9, with details of the number of samples used and the transformation applied as well as the area of Wales each covers.

Percentile	**Principal Domain n = 1270 Area = 18,900 km^2 Transformation: $\log_e$**			**Mineralisation Domain n = 63 Area = 1100 km^2 Transformation: $\log_e$**			**Urban 1 Domain n = 342 Area = 1200 km^2 Transformation: $\log_e$**		
	lower	**middle**	**upper**	**lower**	**middle**	**upper**	**lower**	**middle**	**upper**
50	17	18	18	21	23	27	53	58	63
55	18	19	19	22	24	29	59	64	69
60	19	20	20	23	26	31	65	70	77
65	20	21	21	24	27	33	71	78	86
70	21	22	23	25	29	36	79	87	96
75	22	23	24	26	31	39	88	97	110
80	24	25	26	27	33	43	100	110	130
85	25	27	28	29	35	48	110	130	150
90	28	29	31	32	39	55	140	160	180
95	32	33	**36**	36	45	**67**	180	210	**250**

n = number of samples used in calculation

Table 9: Arsenic NBC summary table

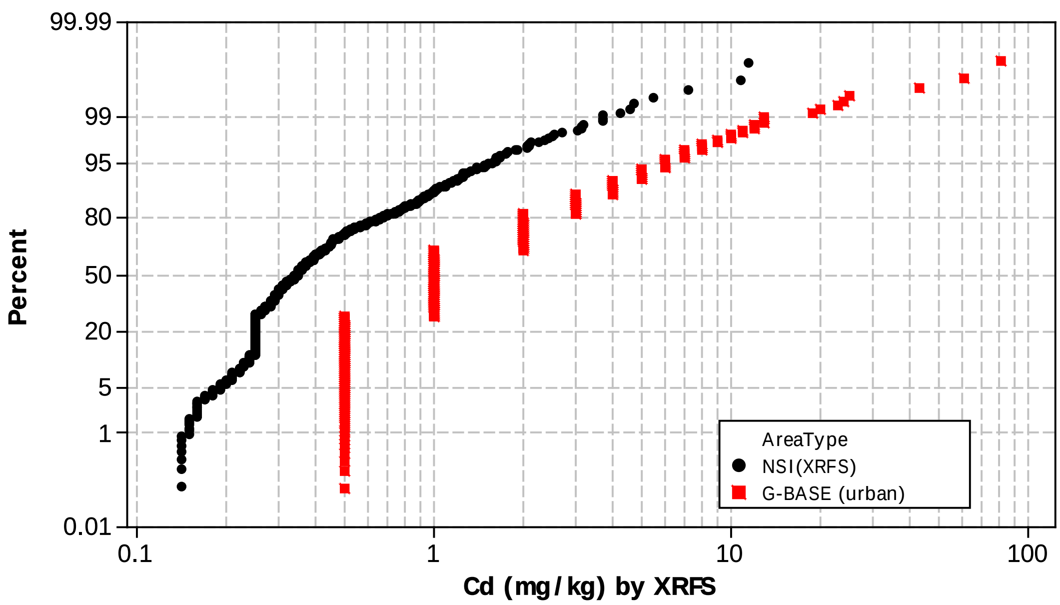
4 Cadmium (Cd)

4.1 EXPLORATORY DATA ANALYSIS

Cadmium data from the G-BASE and NSI(XRFS) projects have greatly varying detection limits, with the lowest (0.25 mg/kg) from the NSI(XRFS) project, whilst that from the G-BASE urban data is 1 mg/kg, which is high in relation to the natural abundance of Cd and results in many data being below the detection limit (Table 10). It is also the case that data just above the detection limit was reported to only one significant figure – leading to multiple identical concentrations being reported from these surveys. The distribution of the data is shown in Figure 27 and Figure 28.

The data are mapped in Figure 29, using only the NSI(XRFS) data due to the difference in detection limits between the two surveys. The data are also presented with thresholds derived from k-means cluster analysis in Figure 30, which show the areas of highest concentrations, thus guiding the definition of domains associated with background concentrations higher than the Principal Domain. To augment these data the stream sediment concentrations (collected at 1 per 2 km^2) are also shown (Figure 31), and although many data are below the lower limit of detection, higher concentrations can be seen around parts of south Wales, Snowdonia and Halkyn areas.

A national data set reported by Davies (1985) had a median of 0.29 mg/kg (722 samples), but it is not clear if these samples and data are independent of the NSI data reported by McGrath and Loveland (1992). Archer and Hodgson (1987) report a median Cd concentration of 0.5 mg/kg for England and Wales, slightly higher than the median of the NSI(XRFS) data (0.35 mg/kg).



Vertical lines of data points are results reported as being below the detection limit, here represented by a value 0.5 × detection limit and truncated to one significant figure

Figure 27: Cumulative probability plot of topsoil Cd concentrations for G-BASE urban and NSI(XRFS) results

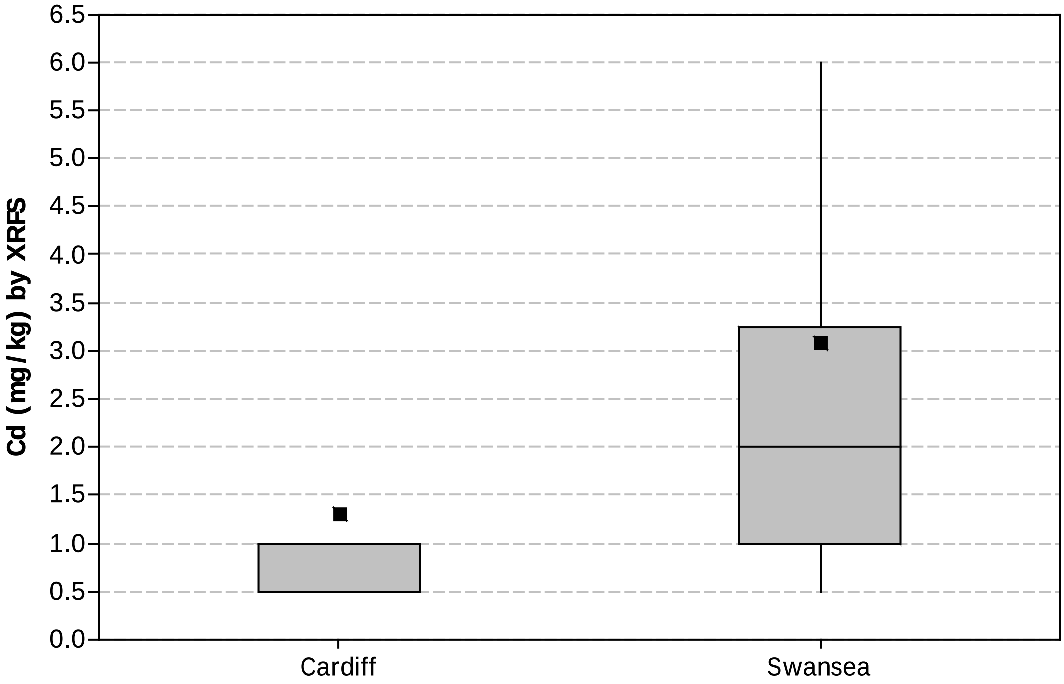


Figure 28: Boxplot of urban topsoil Cd data collected by the G-BASE project

(a) Cd (mg/kg)	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
All data	1675	1.35	<1	<1	<1	1.00	82	15

(b) Cd (mg/kg) Area type	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
NSI(XRFS)	798	0.57	<0.25	0.25	0.35	0.53	11.5	8
Urban	877	2.1	<1	<1	<1	2.0	82.0	12

(c) Cd (mg/kg) Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Cardiff	507	1.3	<1	<1	1.0	1.0	82	17
Swansea	370	3.1	<1	1.0	2.0	3.3	61	7

Table 10: Statistical summary of topsoil Cd in the main data sets (a) whole data set (b) by sampling density (c) by urban centres sampled by G-BASE

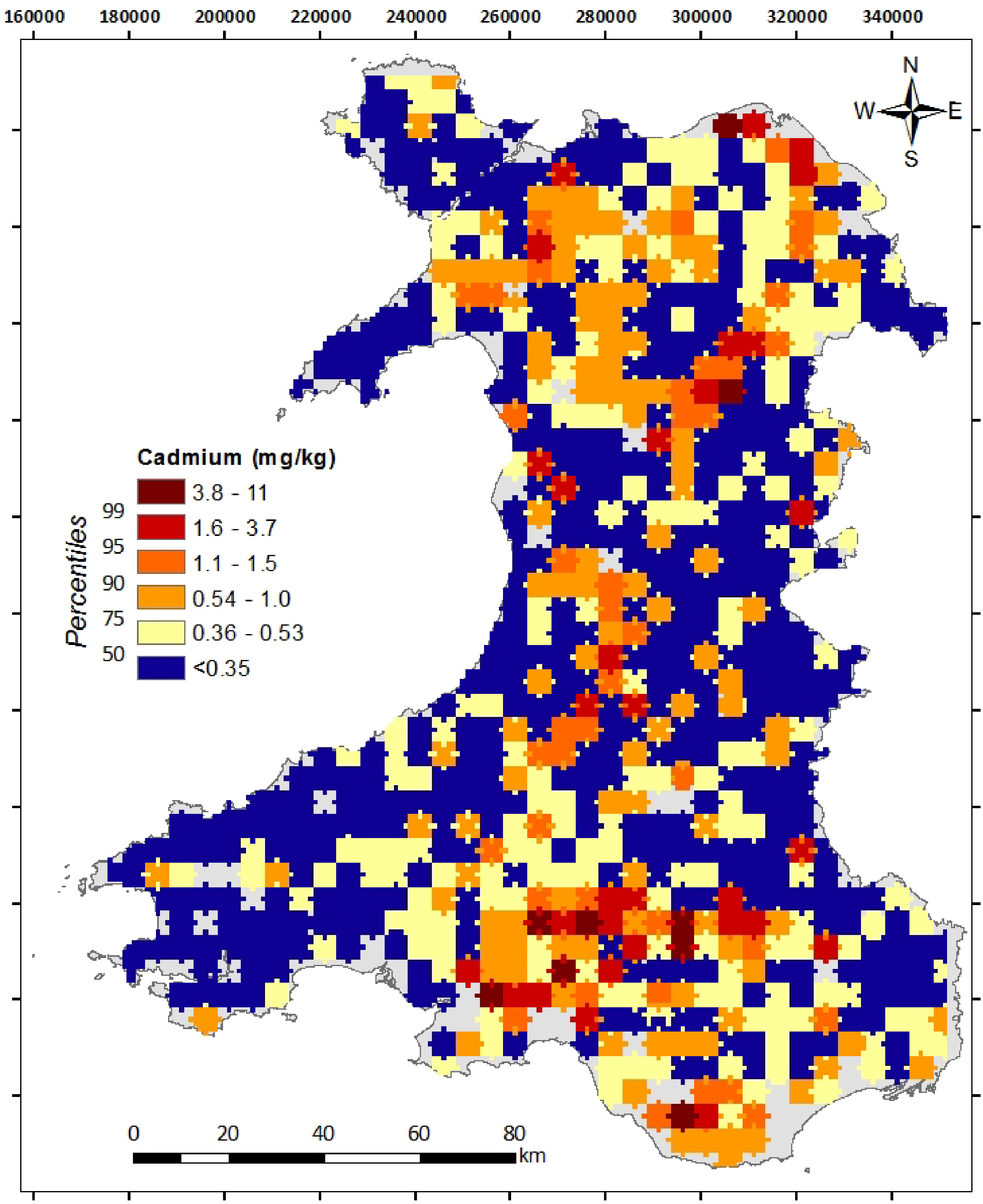


Figure 29: Interpolated map of topsoil Cd, using only NSI(XRFS) data. Colour thresholds are designed for highly skewed data

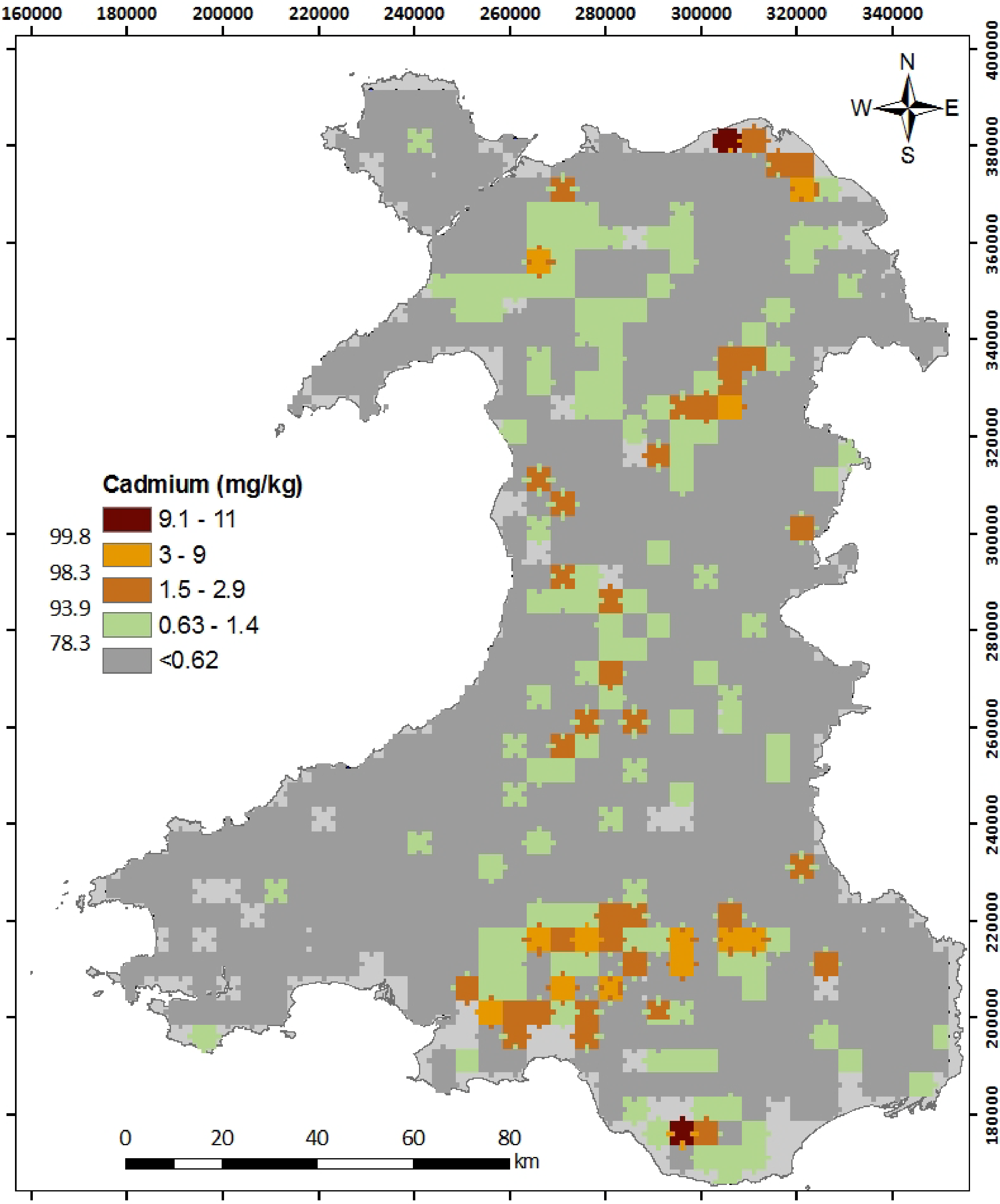
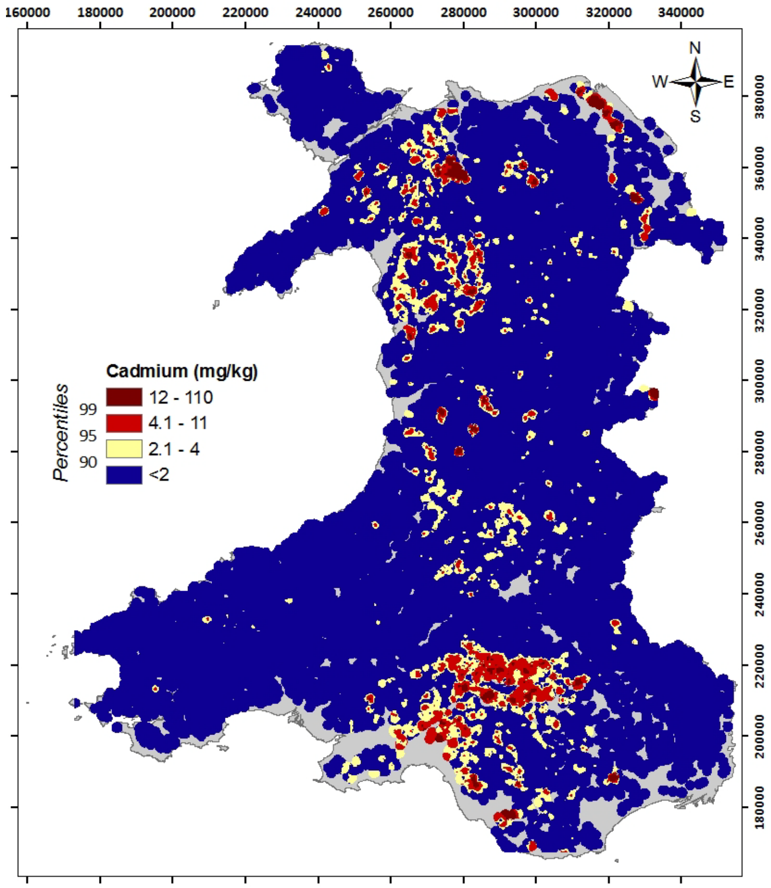
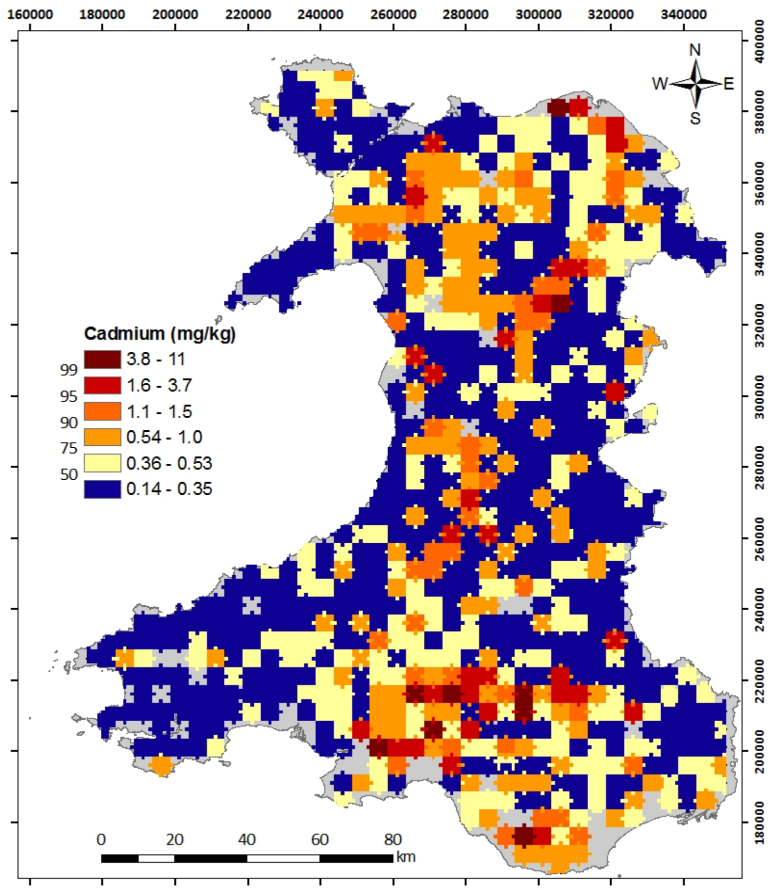


Figure 30: Interpolated map of topsoil NSI(XRFS) Cd data only. Thresholds determined by using k-means cluster analysis



(a)



(b)

Figure 31: Cadmium in (a) G-BASE stream sediments and (b) NSI(XRFS) soils

4.2 DOMAIN SELECTION

The following information was used to compile the evidence for domains in:

- urbanisation extent
- mineralisation areas

4.2.1 Urban areas

The NSI sampling strategy did not include urban centres, so the majority of urban data is derived from the G-BASE Swansea and Cardiff data sets. Only 41 NSI samples fall within areas defined as 'Large urban area' and 'Small urban area' (Table 11) from the OS Strategi® data. It can be seen that the variation in these results depends on how they are classified, whether they are within the G-BASE Swansea data, rather than by the extent of urbanisation (Figure 32 and Figure 33). This is also consistent with typically higher G-BASE stream sediment concentrations of Cd in the Swansea area (Figure 31). A study of an industrial area, including smelting, in the Swansea valley had a median soil Cd concentration of 3.3 mg/kg by partial extraction (Davies, 1997).

When the NSI(XRFS) data are examined for 3 selected valleys in South Wales, including the Tawe, it can be seen that these do reflect higher concentrations outside Swansea, but within the local area (Figure 34) and probably reflect industrial activities associated with the wider urban environment of the Swansea valley and surrounds (Waters *et al*. 2005).

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Large urban areas								
Cardiff	239	1.1	<1	<1	1.0	1.0	5	3
Swansea	199	3.2	<1	1.0	2.0	3.0	61	8
NSI(XRFS)	14	0.42	<0.25	0.25	0.31	0.46	1.63	3
Small urban areas								
Cardiff	6	1.1	<1	<1	1.0	1.3	2.0	1.4
Swansea	7	3.4	<1	2.0	4.0	5.0	5.0	-0.9
NSI(XRFS)	27	0.38	<0.25	0.25	0.27	0.39	1.67	3
Not urban								
Cardiff	262	1.5	<1	<1	1.0	1.0	82	13
Swansea	164	2.9	<1	1.0	2.0	3.8	25	3
NSI(XRFS)	757	0.57	<0.25	0.25	0.35	0.55	11.5	8

Table 11: Cadmium (mg/kg) summary statistics for G-BASE urban and NSI(XRFS) data categorised by urbanisation extent

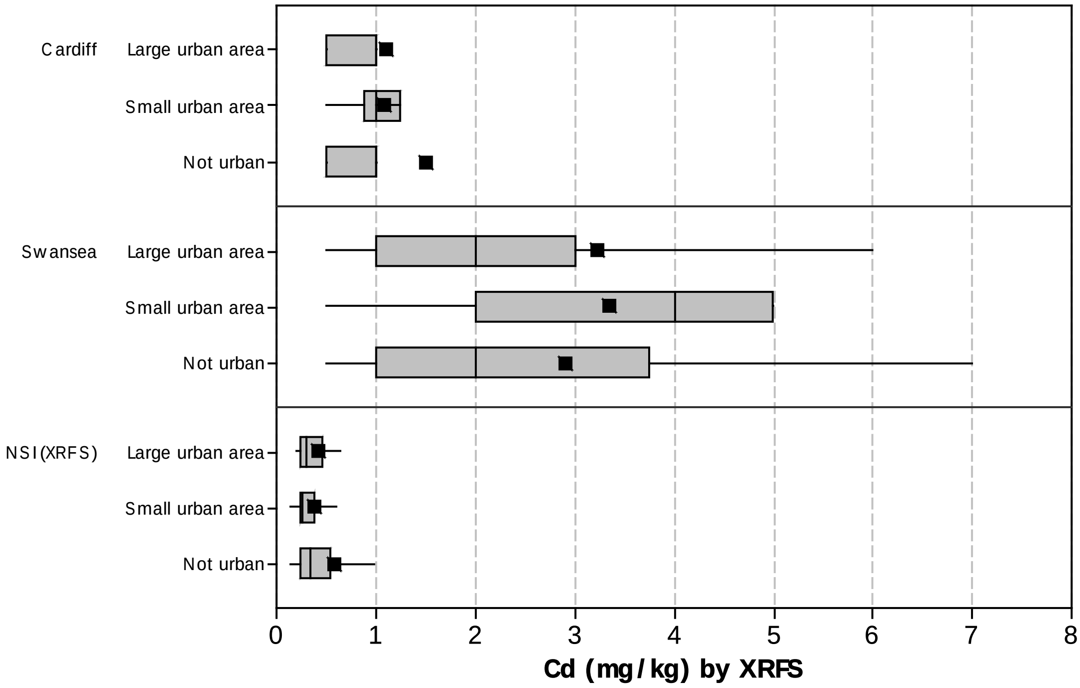


Figure 32: Boxplot of topsoil Cd data classified by extent of urbanisation (OS Strategi® data)

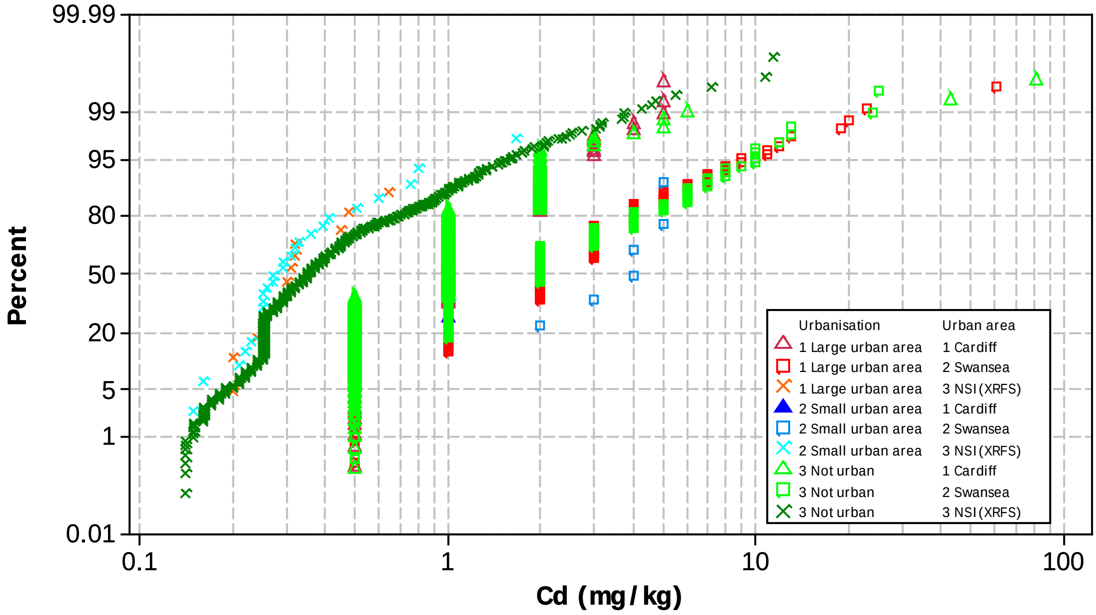
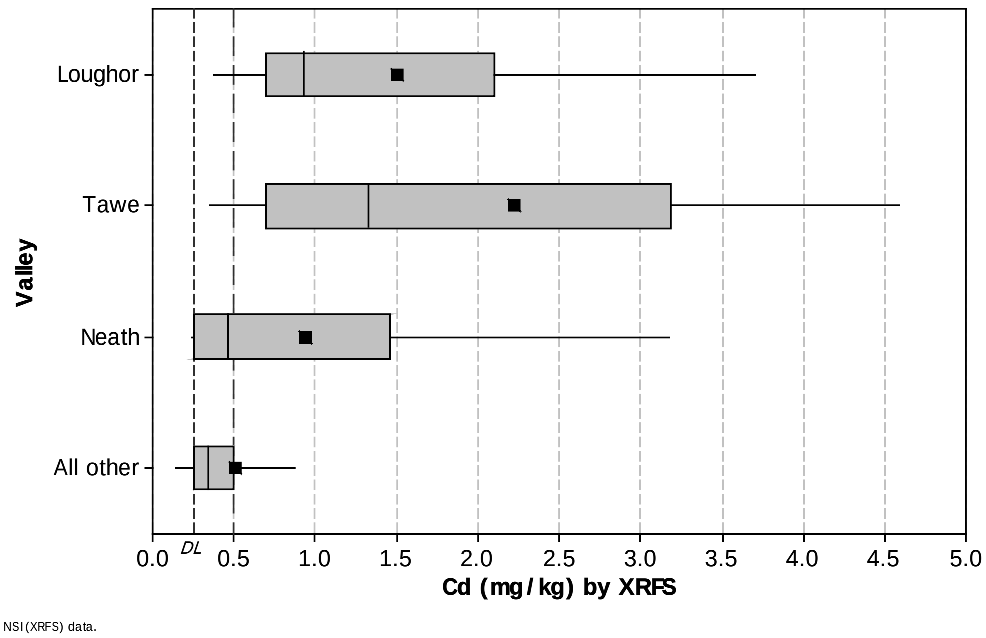


Figure 33: Probability plot of topsoil Cd concentrations in the G-BASE urban and NSI(XRFS) non-urban) data classified by urbanisation extent



The detection limit (DL) and 75th percentile of the 'other data' are shown as dashed black vertical lines

Figure 34: Boxplot of NSI(XRFS) topsoil Cd in Loughor, Tawe and Neath valleys in south Wales

4.2.2 Non-ferrous metalliferous mineralisation and mining

A study of base metal mining areas in Wales (Alloway and Davies, 1971) used the Aeron catchment to calculate background concentrations in comparison to mineralised catchments of Ceredigion. This defined a 'Normal range' of Cd 1.6 - 2.2 mg/kg, compared to alluvium in the Rheidol (Cd range 1.2 - 4.0 mg/kg) which is affected by mineralisation and mining wastes. Fuge *et al.* (1989) found mean concentrations of 2.6 mg/kg and 1.7 mg/kg in the Ystwyth and Rheidol valleys, respectively, but up to 980 mg/kg in soil adjacent to spoil heaps. Soils at Halkyn Mountain were found to have a mean Cd concentration of 23 mg/kg and a maximum of 543 mg/kg (from 60 samples) (Davies and Roberts, 1975), although the authors noted concerns about the analytical method then used.

When the NSI(XRFS) data are compared to the mineralisation extent, a small increase in concentrations can be seen. However, when the data are examined by areas it can be seen that the NW ore-field of Halkyn is responsible for the offset in concentration of the data sets (Figure 35), although there are low sample numbers associated with each of these (Table 12). Evidence of the presence of Cd in mineralised areas of Wales is discussed in the 'Thematic and multi-element studies' section of the BGS stream sediment atlas for Wales (British Geological Survey, 2000).

Thus, there is evidence in these data that the Halkyn area should form one domain for Cd in topsoil. In this area cadmium-tinted smithsonite derived from the weathering of sphalerite (zinc sulphide) is recorded[4]. However, it is surprising, given the evidence from the stream sediment survey, that some of the other ore-fields do not have higher Cd concentrations in relation to the 'other' data, which forms the majority of the Principal Domain. These have therefore been included as a separate mineralised area, notwithstanding the similarity in concentration of topsoil Cd in to the remainder of the data.

[4] http://www.museumwales.ac.uk/cy/800/?mineral=84

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Anglesey	3	0.38	<0.25	<0.25	0.27	0.61	0.61	1.7
north	22	0.70	<0.25	<0.25	0.42	0.78	5.52	4.3
Halkyn	15	1.62	<0.25	0.39	0.55	1.46	11.5	3.4
central	31	0.52	<0.25	<0.25	0.40	0.77	1.50	1.4
south	1	10.8	10.8	*	10.8	*	10.8	*
none	726	0.53	<0.25	<0.25	0.34	0.51	7.21	4.9

Table 12: Summary statistics for topsoil Cd (mg/kg) by ore area

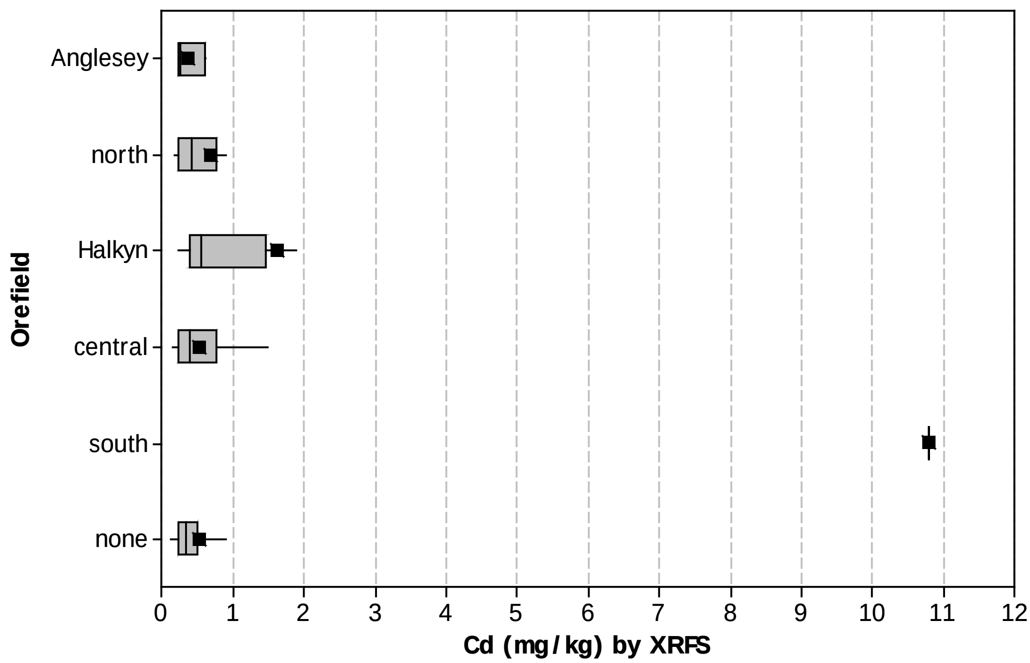


Figure 35: Boxplot of topsoil Cd by ore-field area

4.3 DOMAIN DATA SUMMARY

The data are categorised into the two mineralisation and mining grouped areas (Mineralisation 1 (Halkyn) and Mineralisation 2 (other) Domains) and the urban/industrial area of the Loughor, Tawe and Neath valleys (Urban 1 Domain). The remainder of the data is attributed to the Principal Domain (Table 13, Figure 36 and Figure 37). The extent of these domains in mapped in Figure 38. Only the NSI(XRFS) data has been used, due to the significant discrepancy in detection limits between these data and those from G-BASE urban data sets for Swansea and Cardiff.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Mineralisation 1	15	1.62	<0.25	0.39	0.55	1.46	11. 5	3.4
Mineralisation 2	57	0.76	<0.25	0.25	0.40	0.77	10.8	5.7
Urban 1	45	1.42	<0.25	0.41	0.76	1.93	7.21	2.0
Principal	681	0.47	<0.25	0.25	0.33	0.48	4.71	4.2

Table 13: Summary statistics of topsoil Cd (mg/kg) in the domains

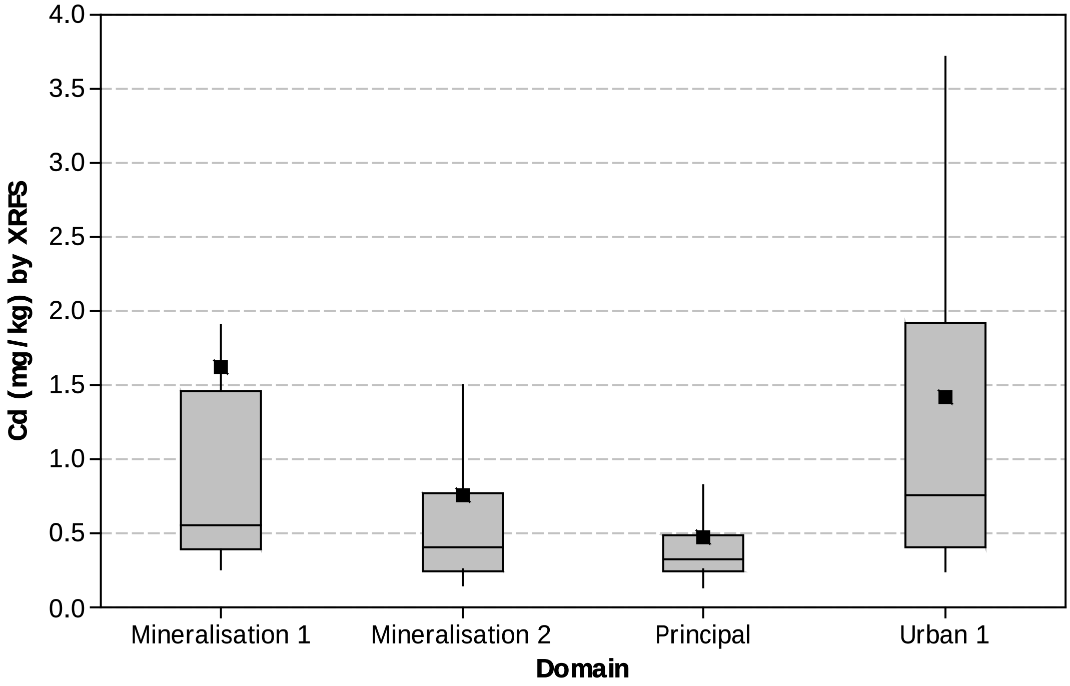


Figure 36: Boxplot of topsoil Cd concentrations by domain

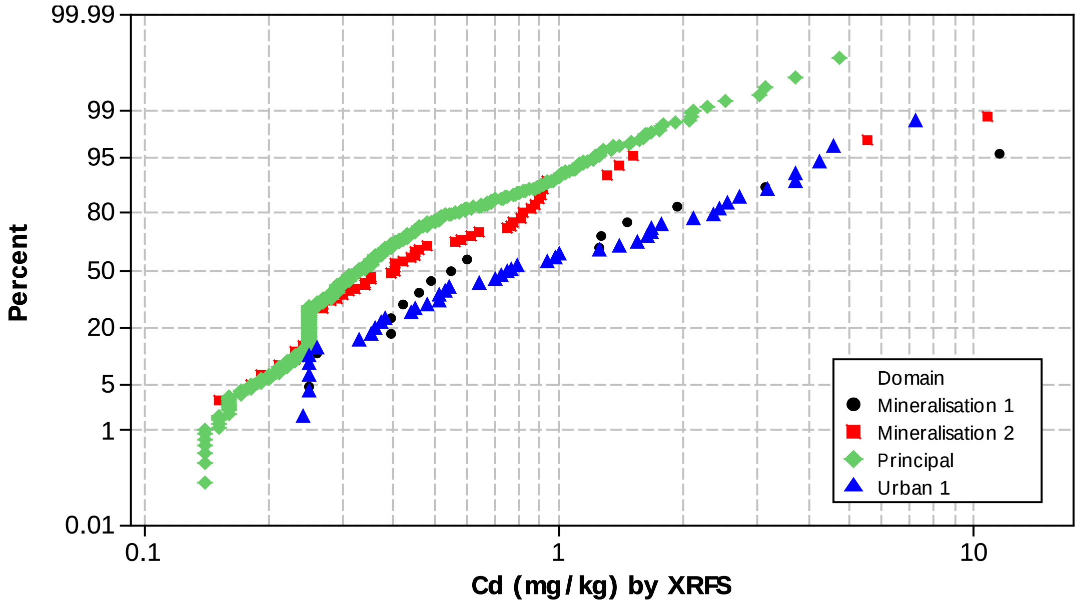
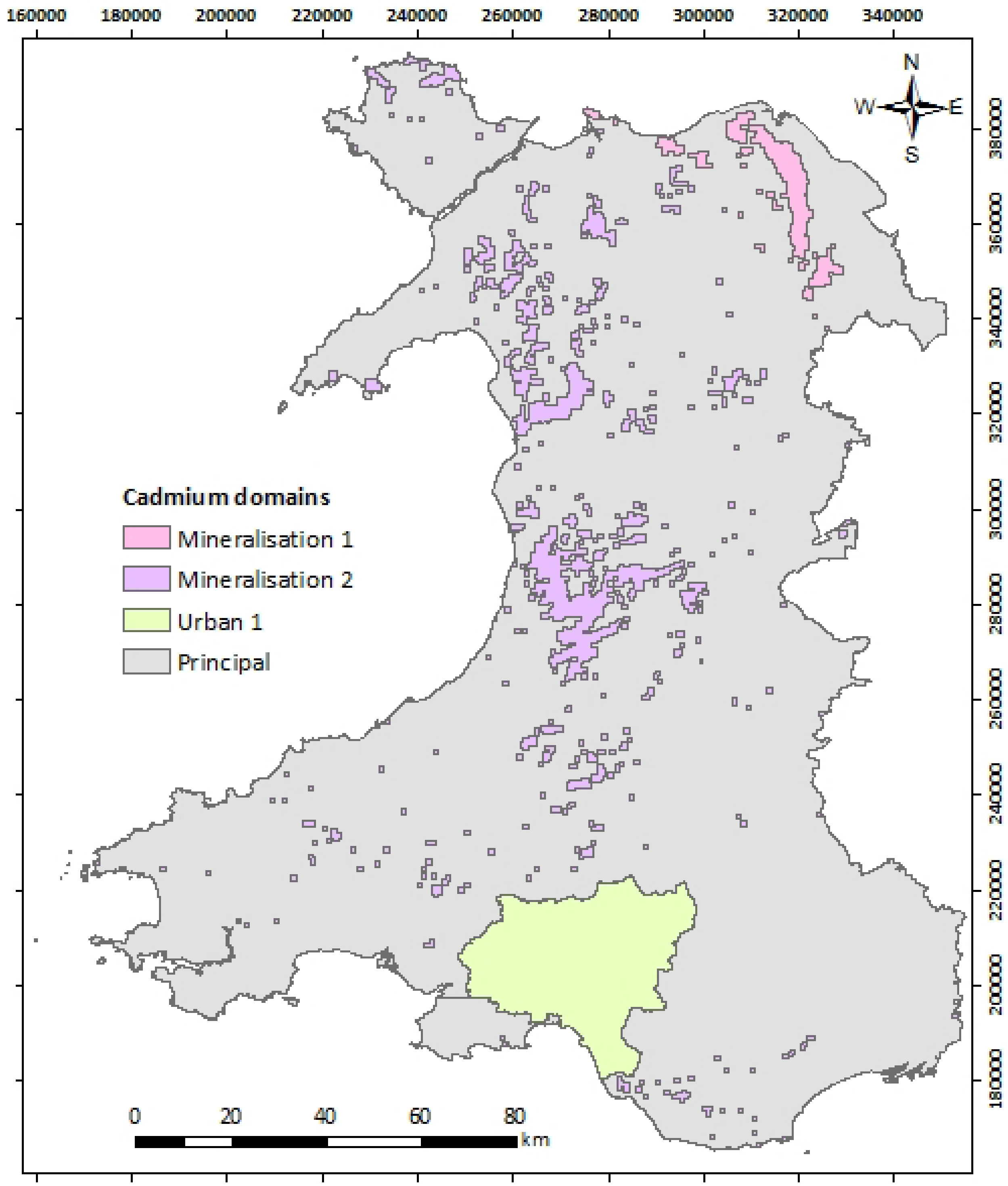


Figure 37: Probability plot of Cd topsoil data classified by domains



Contains Ordnance Survey data © Crown copyright and database rights (2012)
Urban areas defined from OS Strategi® used under OS OpenData™ terms
Mineralisation after the Ove Arup data

Figure 38: Map showing the location of Cd domains

4.4 Cd NBC CALCULATION

The NBC calculations for Cd have been carried out according to the method described in Cave *et al.* (2012).

4.4.1 Principal Domain

Untransformed and $\log_e$ transfomed data have skewness coefficient and octile skew which fail the test criteria for underlying Gaussian distribution (Figure 39). The Box-Cox transform brings the skew coefficient below 1, but the octile skew is still >0.2 (Figure 40 and Figure 41). Therefore empirical percentiles on the untransformed data were used to calculate the NBC (Figure 42 and Table 14).

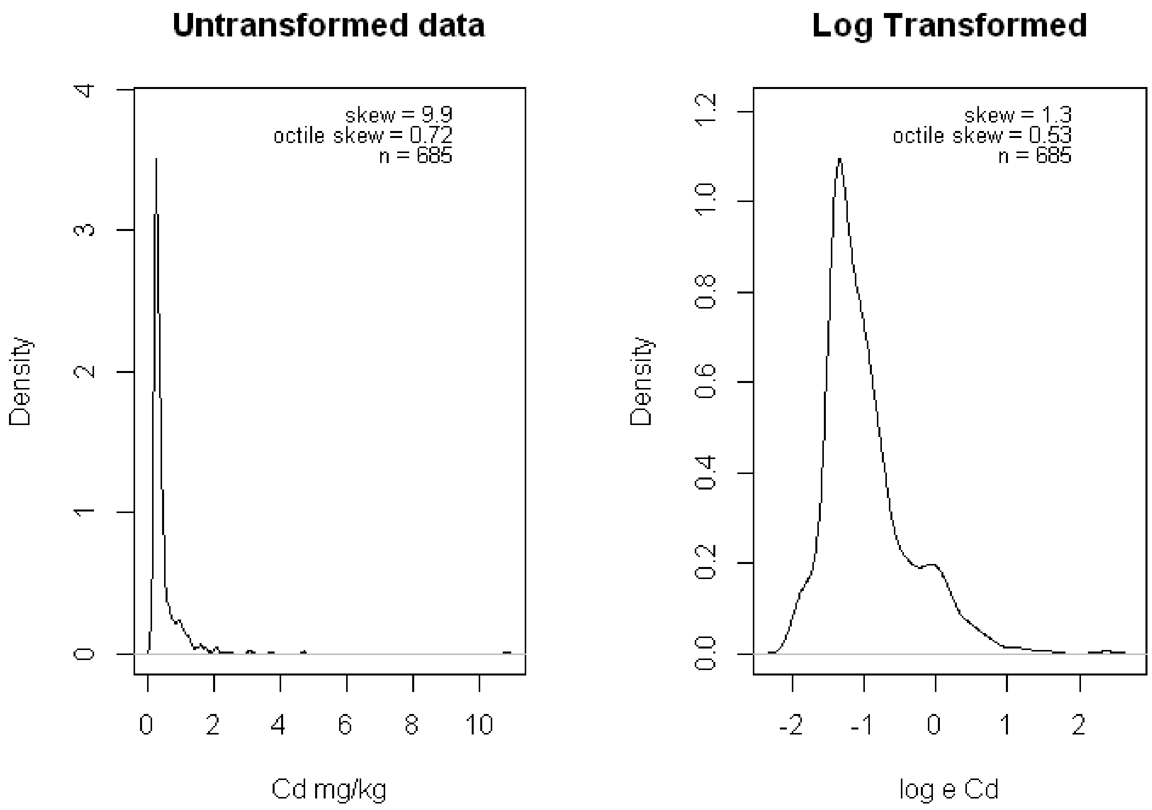


Figure 39: Density distributions for the raw data and the $\log_e$ transformed data for Cd in the Principal Domain (n = number of samples)

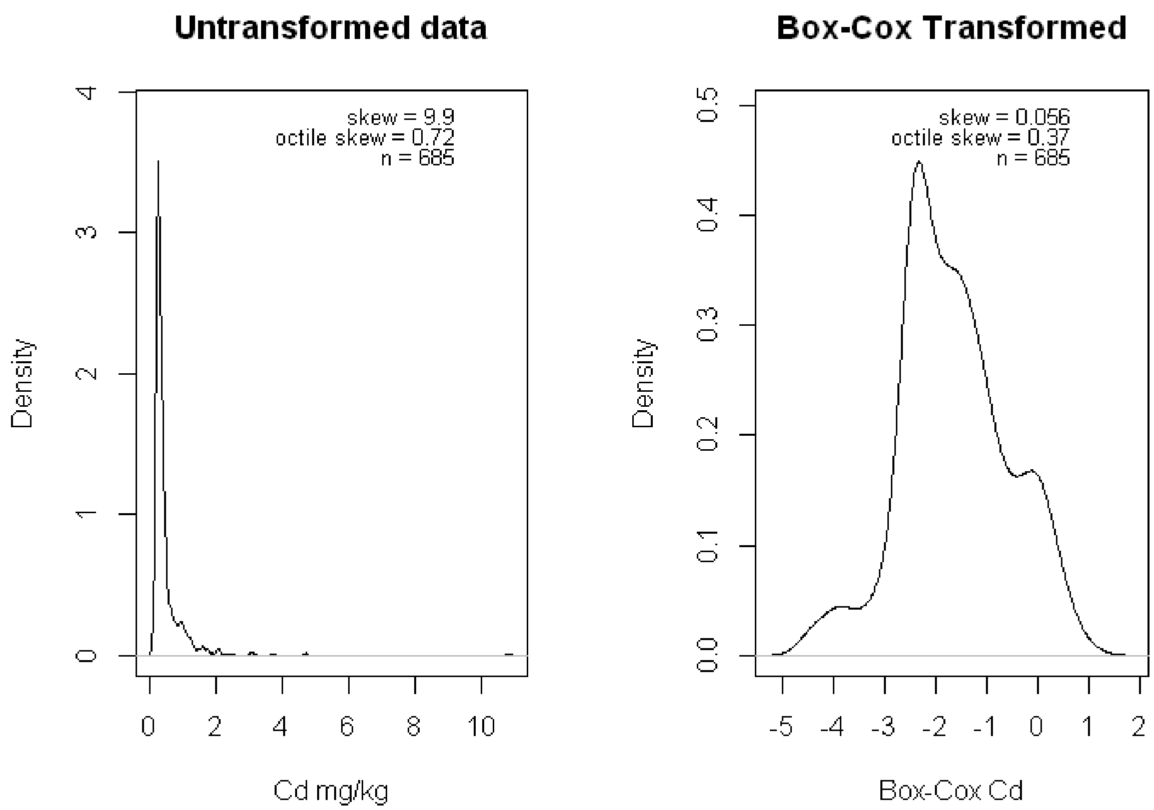


Figure 40: Density distributions for the raw data and the Box-Cox transformed data for Cd in the Principal Domain (n = number of samples)

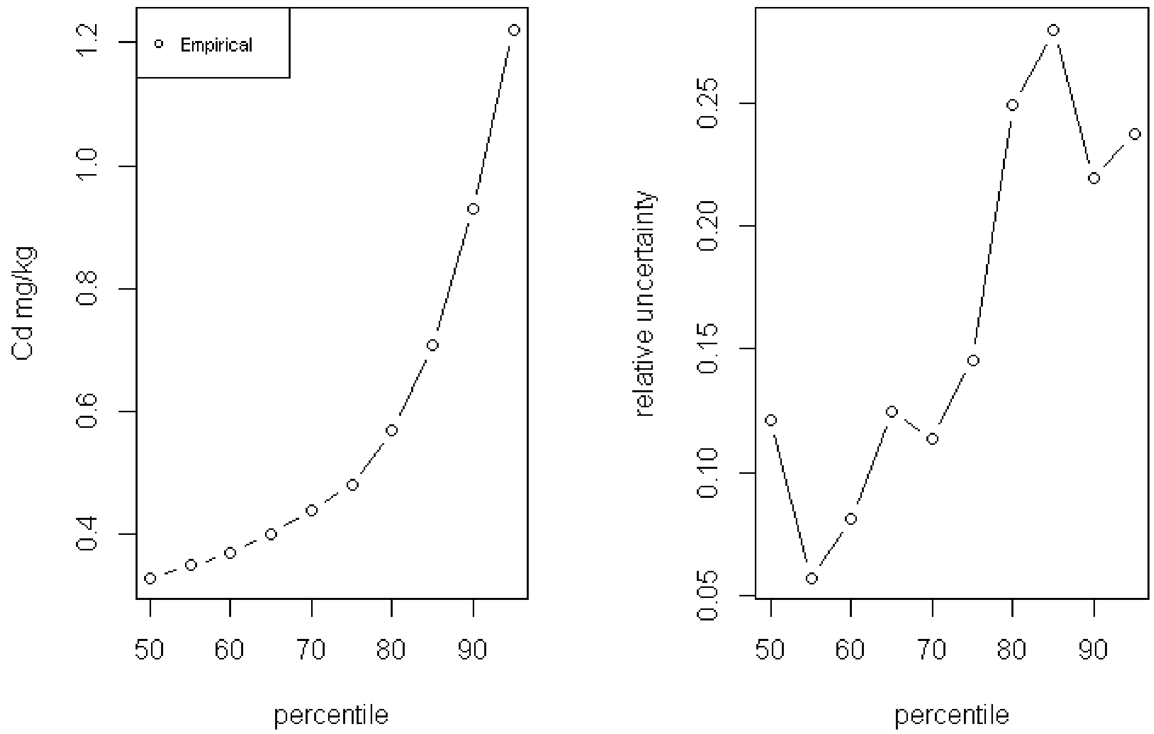


Figure 41: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Cd in the Principal Domain

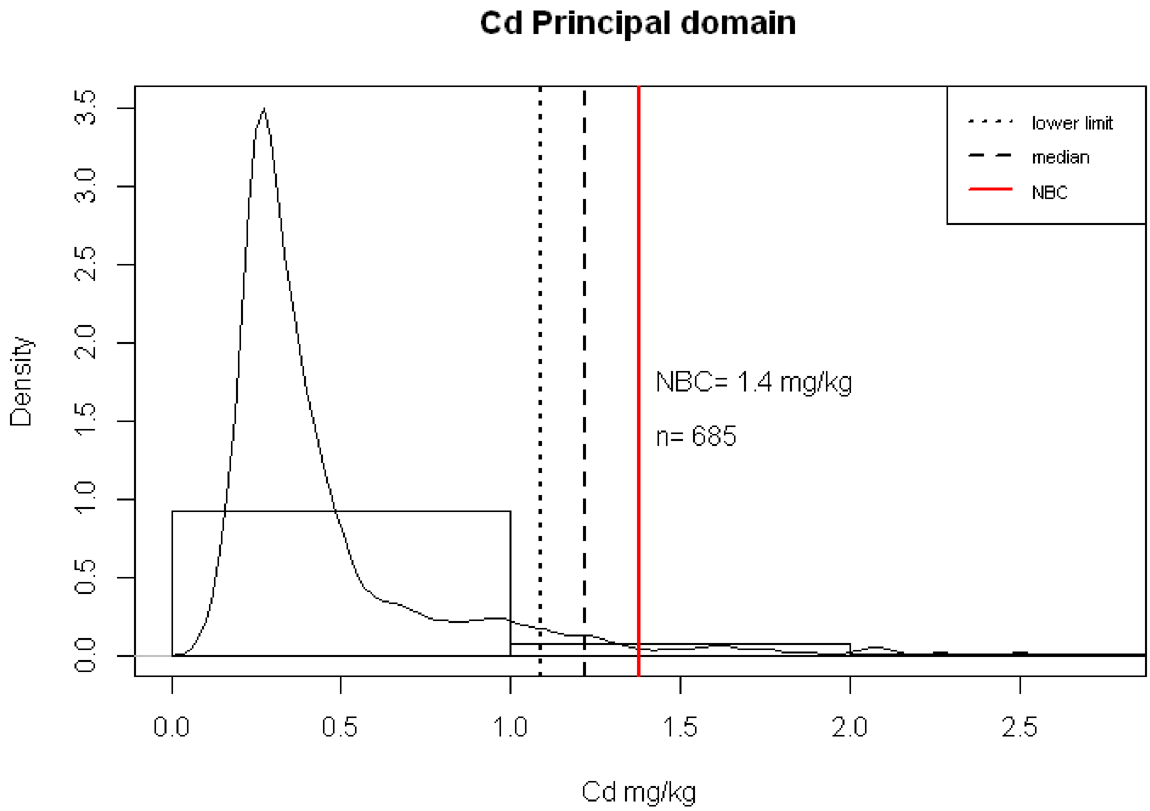


Figure 42: Summary density plot and histogram of the distribution for Cd in the Principal Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H
50	0.31	0.33	0.35
55	0.34	0.35	0.36
60	0.36	0.37	0.39
65	0.38	0.40	0.43
70	0.41	0.44	0.47
75	0.45	0.48	0.52
80	0.51	0.57	0.66
85	0.62	0.71	0.81
90	0.83	0.93	1.00
95	1.10	1.20	**1.40**

Low (L) and High (H) values represent the 95% confidence intervals around the median.

Table 14: Empirical (Emp), percentile values for Cd (mg/kg) in the Principal Domain

4.4.2 Mineralisation 1 Domain

With only fifteen samples in this domain it is not possible to calculate an NBC (Cave *et al.*, 2012)

4.4.3 Mineralisation 2 Domain

Untransformed and $\log_e$ transformed skewness and octile skew coefficients fail the test criteria for underlying Gaussian distribution (Figure 43). Therefore a Box-Cox transformation was applied (Figure 44 and Figure 45), which indicated an underlying Gaussian distribution, for which parametric percentiles were used to calculate a NBC (Figure 46 and Table 15).

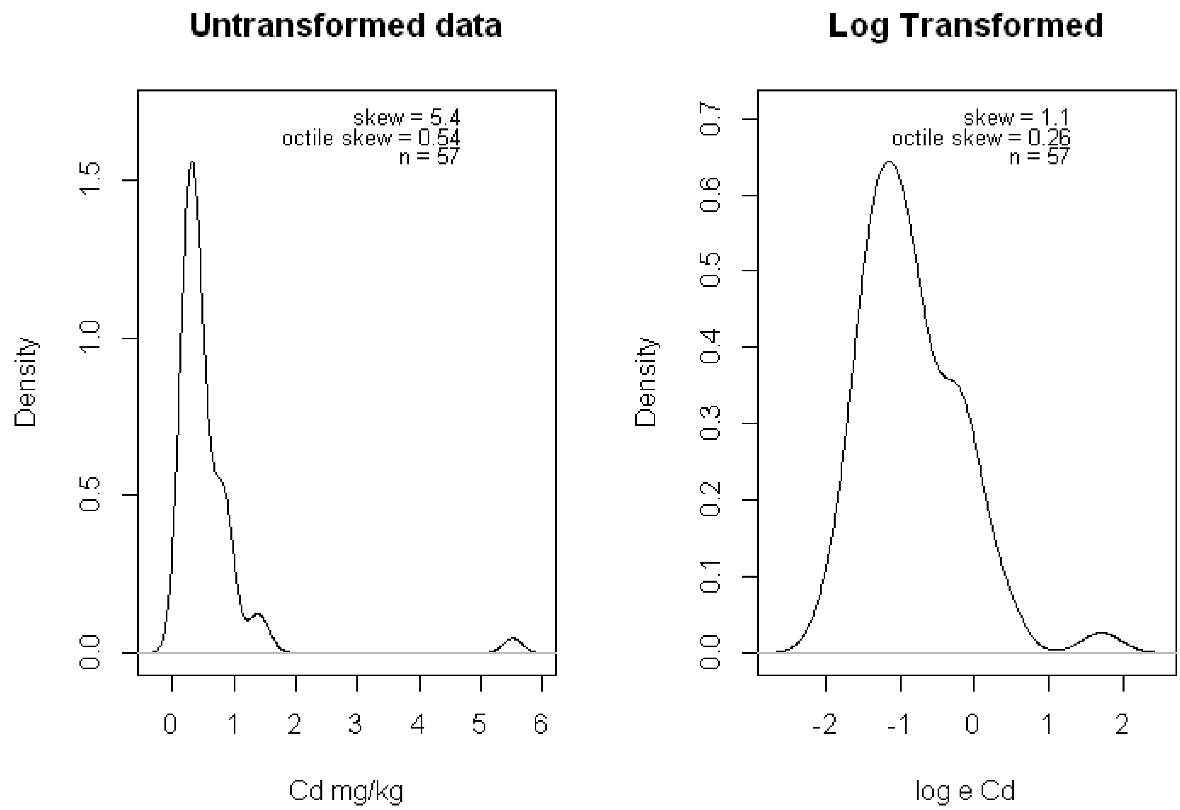


Figure 43: Density distributions for the raw data and the $\log_e$ transformed data for Cd in the Mineralisation 2 Domain (n = number of samples)

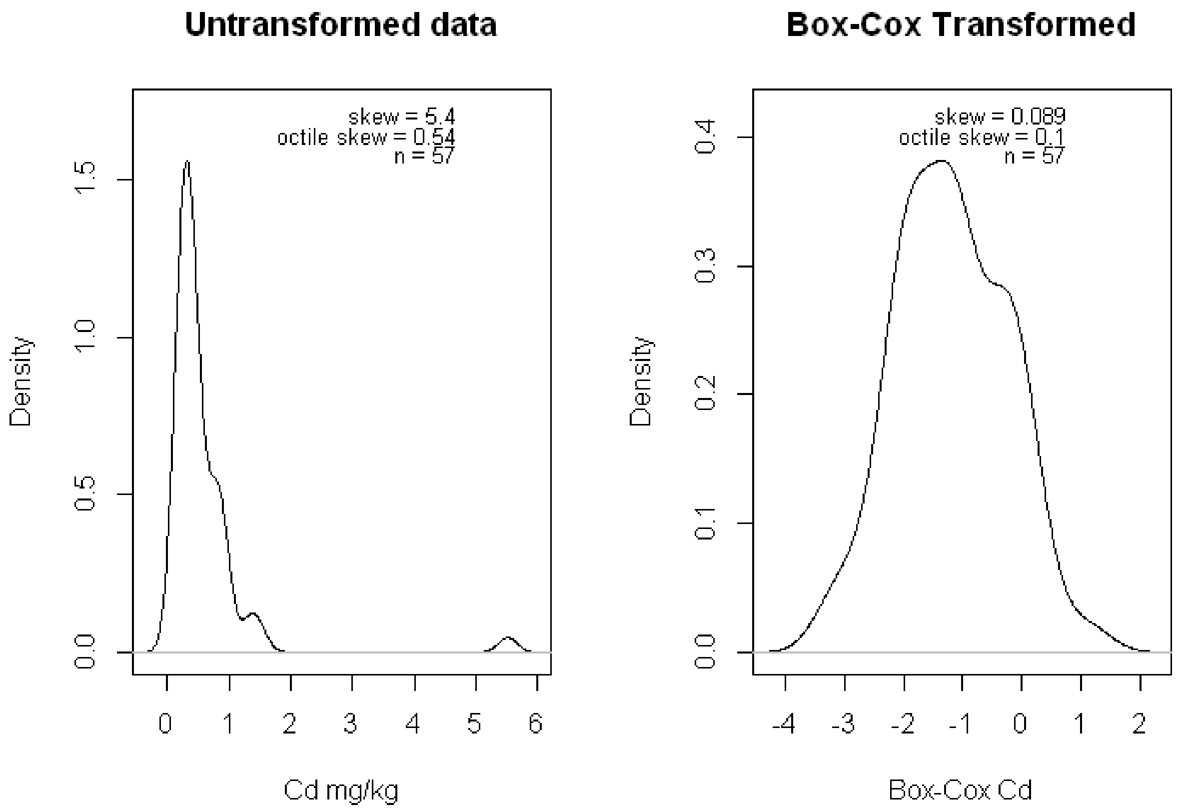


Figure 44: Density distributions for the raw data and the Box-Cox transformed data for Cd in the Mineralisation 2 Domain (n = number of samples)

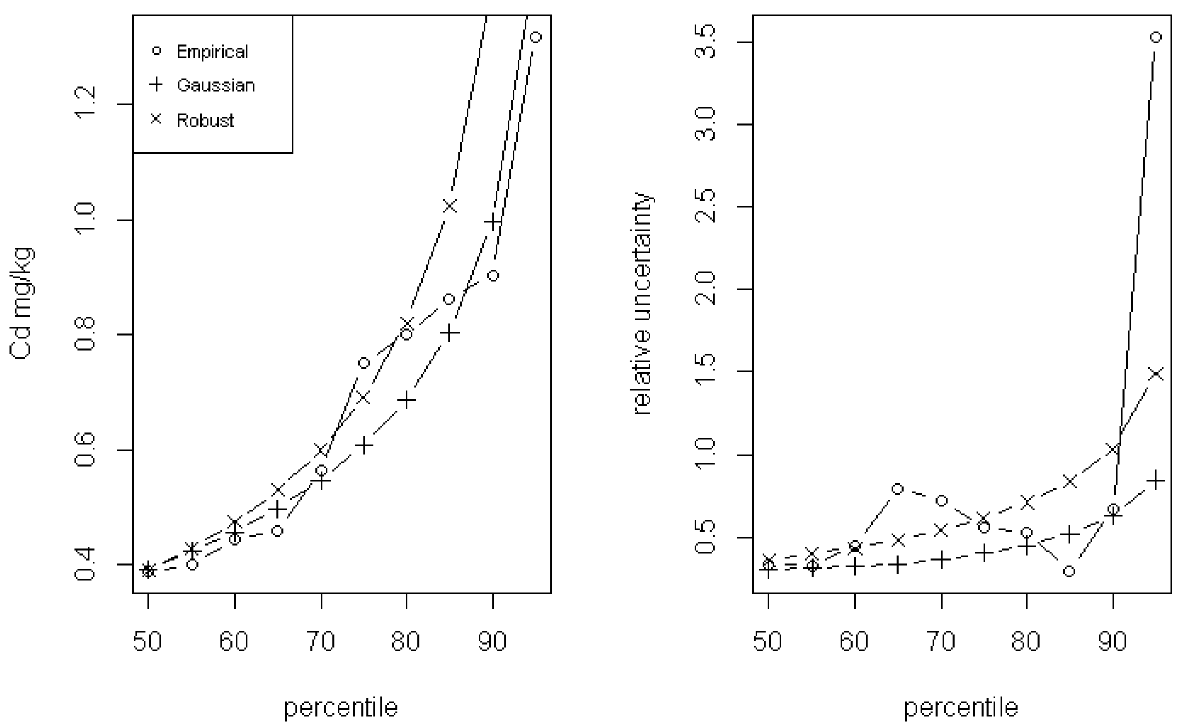


Figure 45: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Cd in the Mineralisation 2 Domain

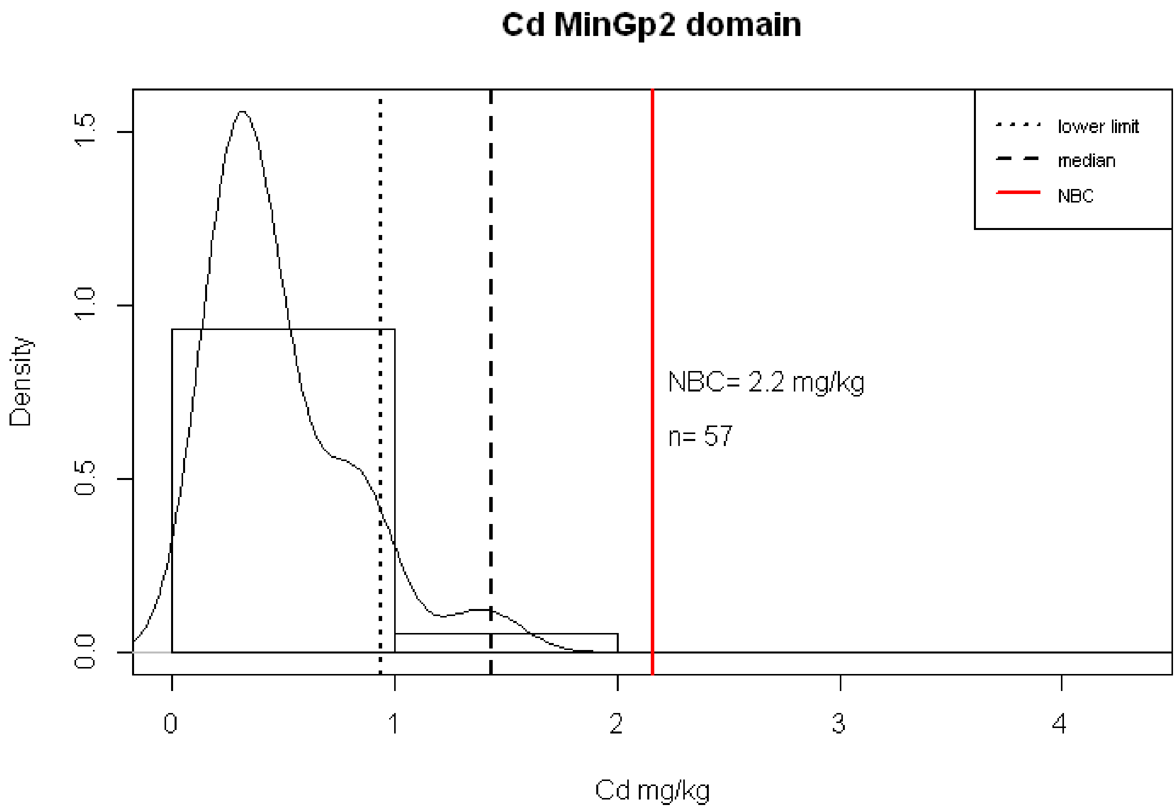


Figure 46: Summary density plot and histogram of the distribution for Cd in the Mineralisation 2 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	0.3	0.4	0.5	**0.3**	**0.4**	**0.5**	0.3	0.4	0.5
55	0.3	0.4	0.5	**0.4**	**0.4**	**0.5**	0.3	0.4	0.5
60	0.4	0.4	0.6	**0.4**	**0.5**	**0.5**	0.4	0.5	0.6
65	0.4	0.5	0.8	**0.4**	**0.5**	**0.6**	0.4	0.5	0.6
70	0.4	0.6	0.8	**0.5**	**0.5**	**0.7**	0.4	0.6	0.7
75	0.5	0.8	0.9	**0.5**	**0.6**	**0.7**	0.4	0.7	0.8
80	0.5	0.8	0.9	**0.5**	**0.7**	**0.9**	0.5	0.8	1.0
85	0.7	0.9	0.9	**0.6**	**0.8**	**1.0**	0.5	1.0	1.4
90	0.8	0.9	1.4	**0.7**	**1.0**	**1.4**	0.6	1.4	2.1
95	0.9	1.3	1.8	**0.9**	**1.4**	**2.2**	0.7	2.5	4.5

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 15: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Cd (mg/kg) in the Mineralisation 2 Domain

4.4.4 Urban 1 Domain

Untransformed skewness and octile skew coefficients fail the test criteria for underlying Gaussian distribution. When $\log_e$ transformed was used, skewness coeficient was <1 and octile skew <0.2, indicating an underlying Gaussian distribution (Figure 47 and Figure 48) for which parametric percentiles were used (Table 16) to calculate an example NBC, as shown in Figure 49.

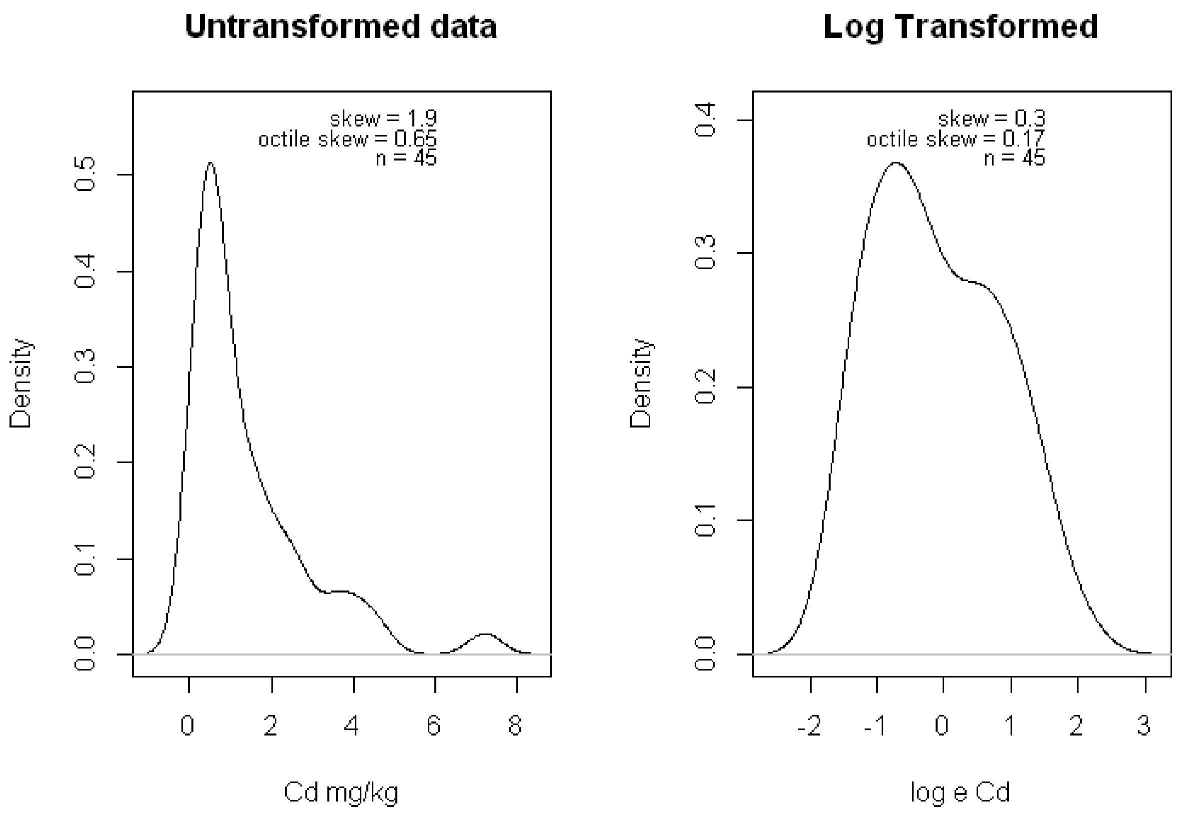


Figure 47: Density distributions for the raw data and the $\log_e$ transformed data for Cd in the Urban 1 Domain (n = number of samples)

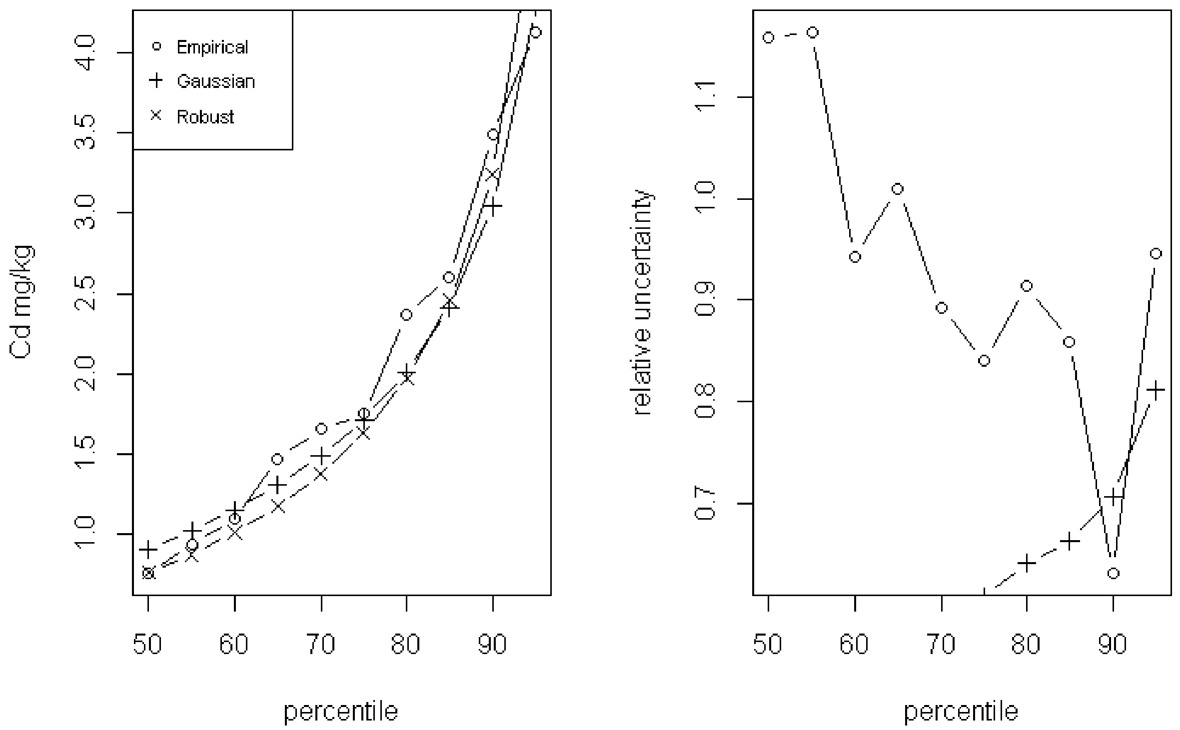


Figure 48: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Cd in the Urban 1 Domain

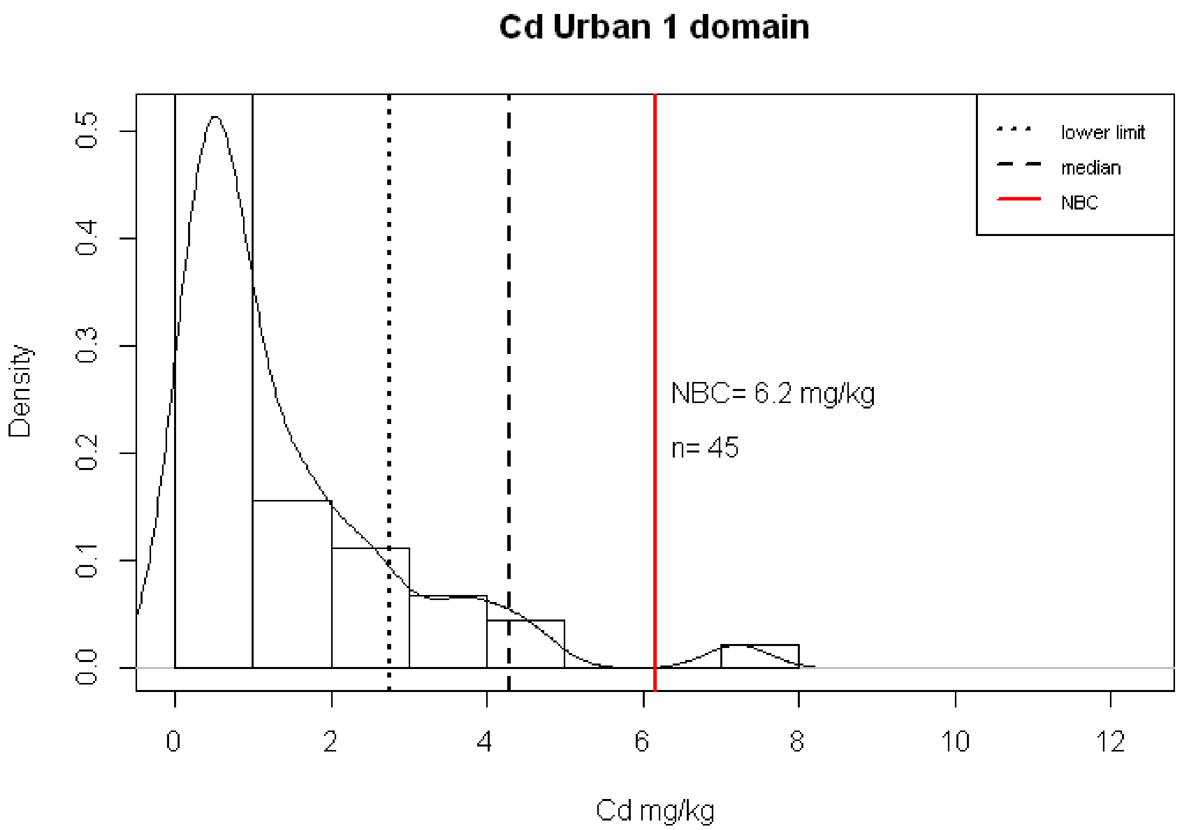


Figure 49: Summary density plot and histogram of the distribution for Cd in the Urban 1 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	0.5	0.8	1.4	0.7	0.9	1.2	0.5	0.8	1.4
55	0.6	0.9	1.7	0.8	1.0	1.3	0.6	0.9	1.7
60	0.7	1.1	1.7	0.9	1.2	1.5	0.6	1.0	2.0
65	0.8	1.5	2.2	1.0	1.3	1.7	0.7	1.2	2.2
70	0.9	1.7	2.4	1.1	1.5	2.0	0.8	1.4	2.6
75	1.3	1.8	2.7	1.2	1.7	2.3	0.9	1.6	3.2
80	1.5	2.4	3.4	1.4	2.0	2.7	1.1	2.0	3.9
85	1.7	2.6	3.9	1.7	2.4	3.3	1.2	2.5	5.0
90	2.2	3.5	4.4	2.1	3.0	4.2	1.5	3.2	6.9
95	2.7	4.1	6.6	2.8	4.3	**6.2**	1.9	4.9	11

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 16: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Cd (mg/kg) in the Urban 1 Domain

4.4.5 Cadmium NBC summary

Summary of the Cd domain NBCs is shown in Table 17, with details of the number of samples used and the transformation applied as well as the area of Wales each covers.

Percentile	Principal Domain n =685 Area = 18,700 km^2 Transformation: none			Mineralisation 1 Domain n = 15 Area = 200 km^2 Transformation: n/a			Mineralisation 2 Domain n = 57 Area = 1200 km^2 Transformation: Box-Cox		
	lower	middle	upper	lower	middle	upper	lower	middle	upper
50	0.31	0.33	0.35	nc	nc	nc	0.3	0.4	0.5
55	0.34	0.35	0.36	nc	nc	nc	0.4	0.4	0.5
60	0.36	0.37	0.39	nc	nc	nc	0.4	0.5	0.6
65	0.38	0.40	0.43	nc	nc	nc	0.4	0.5	0.6
70	0.41	0.44	0.47	nc	nc	nc	0.5	0.6	0.7
75	0.45	0.48	0.52	nc	nc	nc	0.5	0.6	0.8
80	0.51	0.57	0.66	nc	nc	nc	0.6	0.7	0.9
85	0.62	0.71	0.81	nc	nc	nc	0.6	0.8	1.0
90	0.83	0.93	1.00	nc	nc	nc	0.7	1.0	1.4
95	1.10	1.20	**1.40**	nc	nc	**nc**	1.0	1.4	**2.2**

Percentile	Urban 1 Domain n = 45 Area = 1200 km^2 Transformation: $\log_e$		
	lower	**middle**	**upper**
50	0.7	0.9	1.2
55	0.8	1.0	1.3
60	0.9	1.2	1.5
65	1.0	1.3	1.7
70	1.1	1.5	2.0
75	1.2	1.7	2.3
80	1.4	2.0	2.7
85	1.7	2.4	3.3
90	2.1	3.0	4.2
95	2.8	4.3	**6.2**

n = number of samples used in calculation. nc = not calculated (n < 30)

Table 17: Cadmium (mg/kg) NBC summary table

5 Copper (Cu)

5.1 EXPLORATORY DATA ANALYSIS

The Cu data from NSI(XRFS) and G-BASE urban show concentrations which are typically higher in the latter (Table 18, Figure 50 and Figure 51).

The data are mapped in Figure 52, with the higher concentrations around the urban centres of Swansea and Cardiff clearly visible. These data are also represented using k-means cluster analysis thresholds. This method is used to identify data populations higher in concentration than those which are typical across Wales (Figure 53). However, to examine these variations in the NSI(XRFS) data, it is necessary to recalculate and re-plot just the NSI(XRFS) data (Figure 54). This shows more areas of elevated Cu results, including a small number in north and central Wales, as well as those from the west of the South Wales Coalfield.

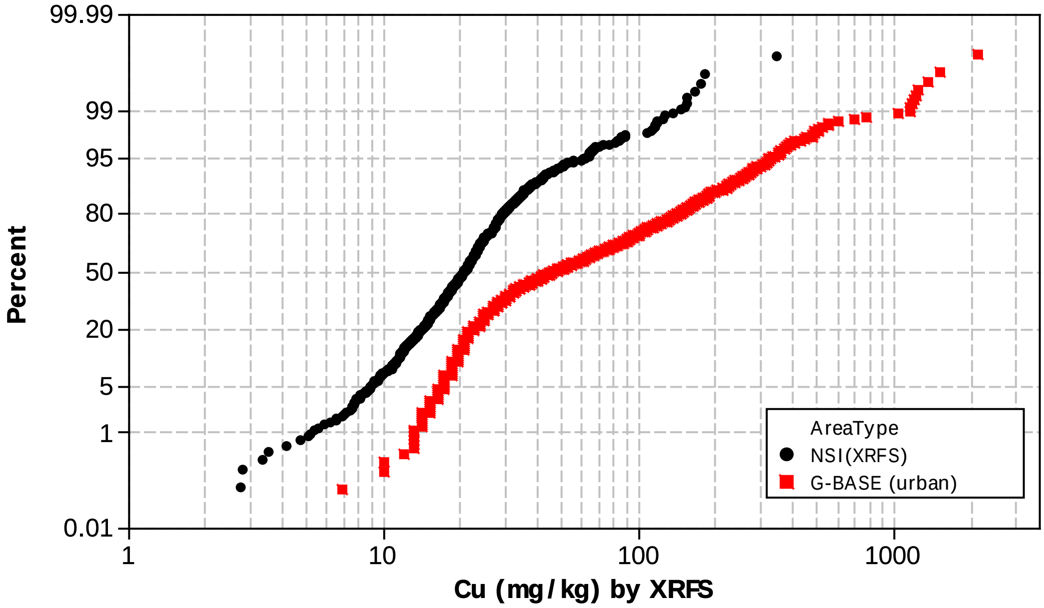


Figure 50: Cumulative probability plot of topsoil Cu concentrations for G-BASE urban and NSI(XRFS) results

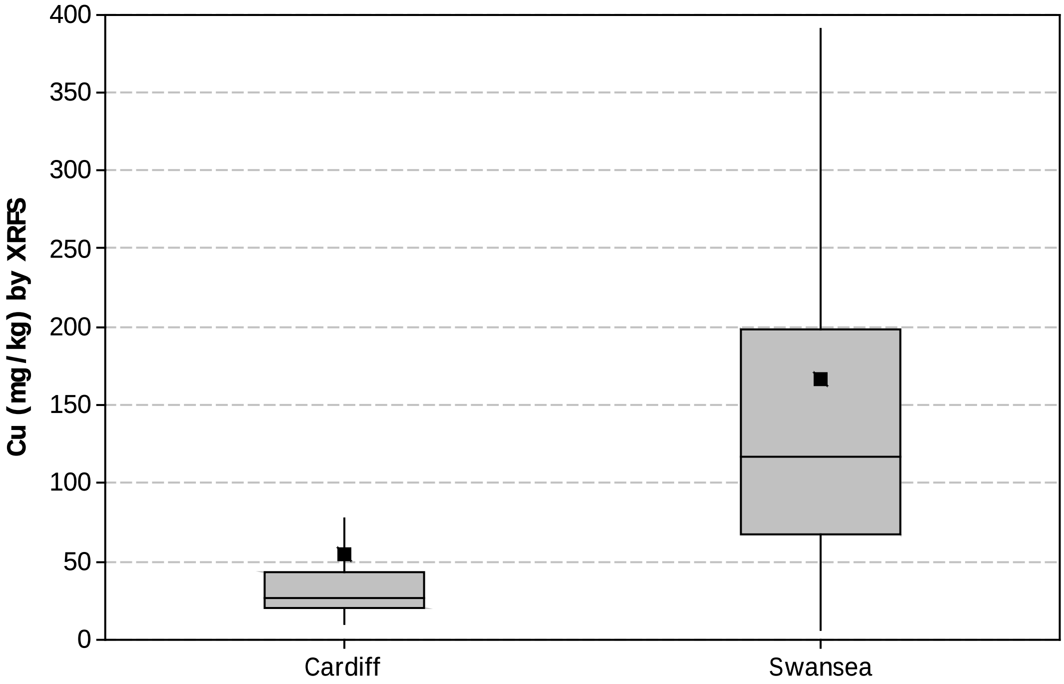


Figure 51: Boxplot of urban topsoil Cu data collected by the G-BASE project

(a) Cu (mg/kg)	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
All data	1675	65.5	2.73	18.8	26.7	60.0	2131	7.44
(b) Cu (mg/kg) Area type	**Number**	**Mean**	**Minimum**	**25th percentile**	**Median**	**75th percentile**	**Maximum**	**Skewness**
NSI(XRFS)	798	25.5	2.73	15.0	20.6	27.3	344	6
Urban	877	102	6.85	24.6	45.4	117	2131	6
(c) Cu (mg/kg) Area name	**Number**	**Mean**	**Minimum**	**25th percentile**	**Median**	**75th percentile**	**Maximum**	**Skewness**
Cardiff	507	54.3	9.98	20.4	26.7	43.3	2131	10
Swansea	370	167	6.85	67.3	117	199	1507	4

Table 18: Statistical summary of topsoil Cu in the main data sets (a) whole data set (b) by sampling density (c) by urban centres sampled by G-BASE

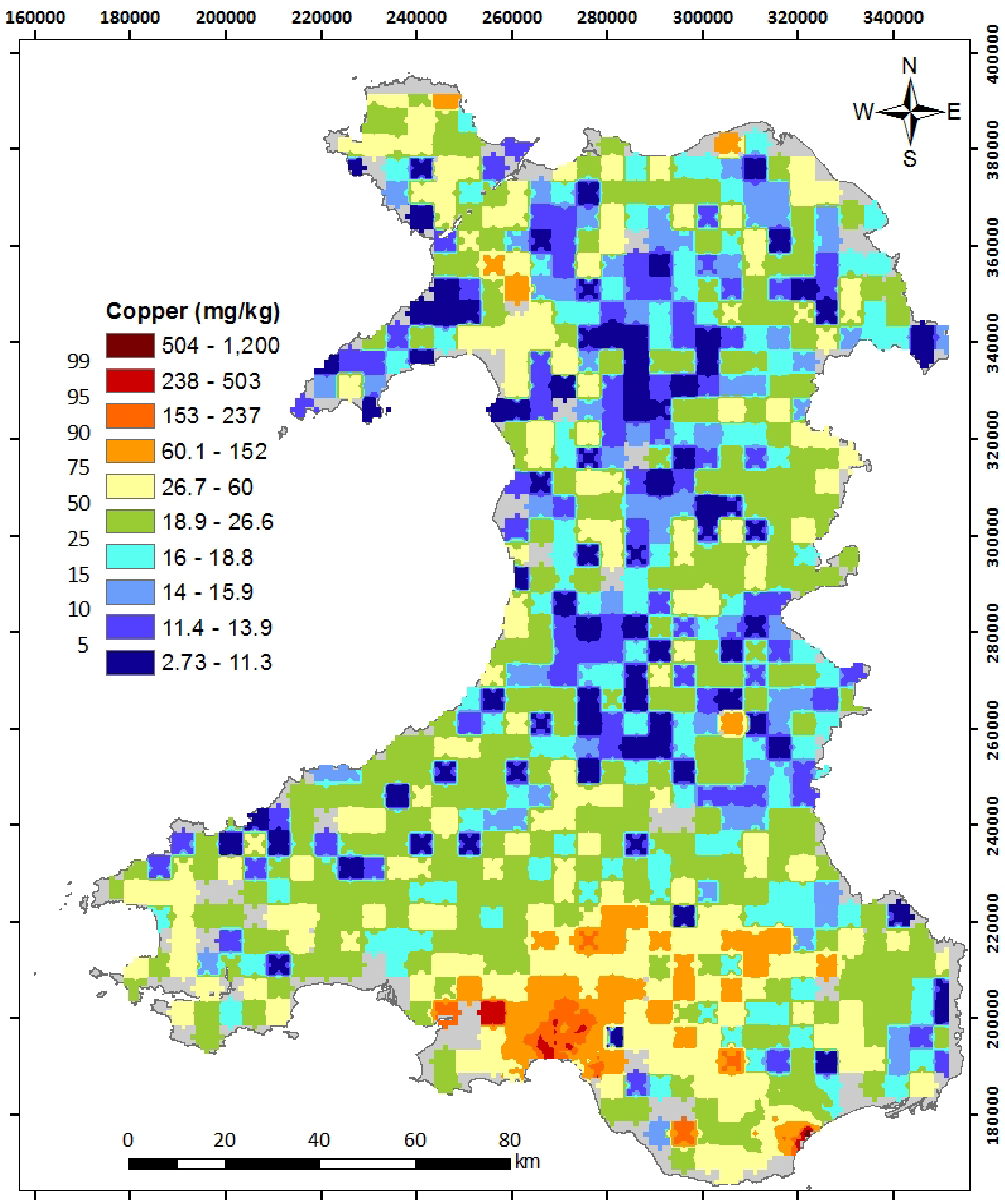


Figure 52: Interpolated map of topsoil Cu data. Colour thresholds are designed for highly skewed data

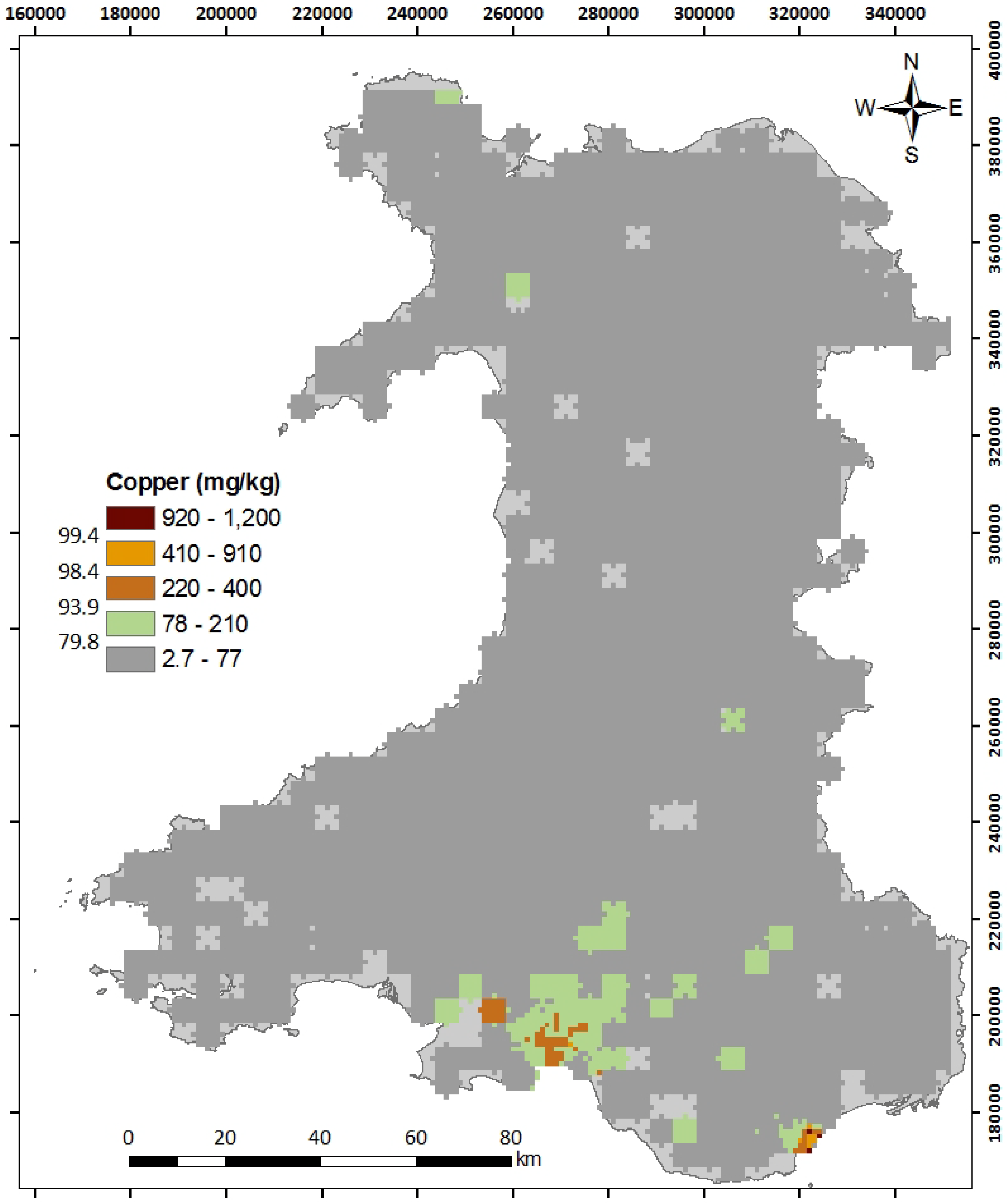


Figure 53: Interpolated map of topsoil Cu concentrations. Thresholds determined by using k-means cluster analysis

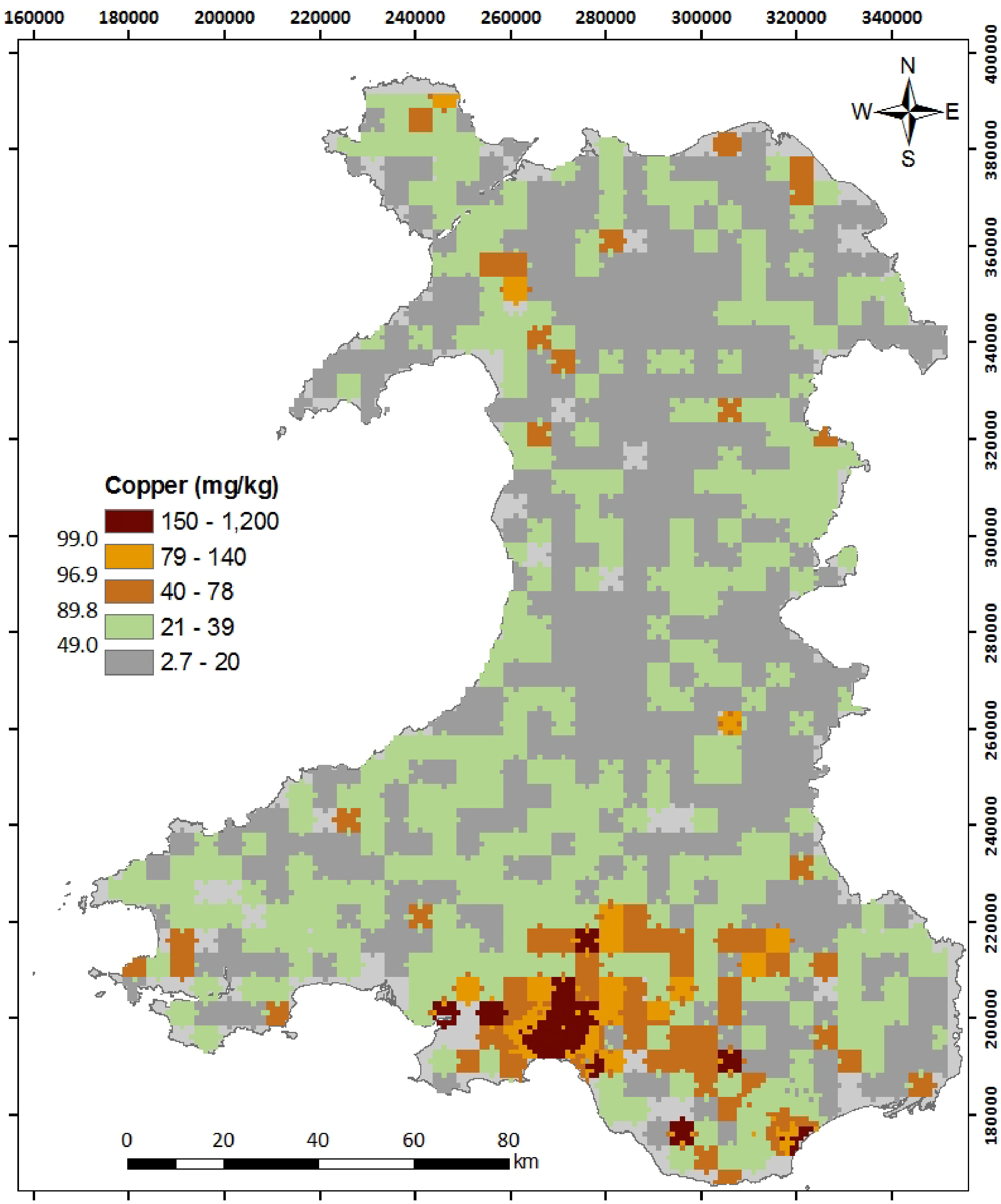
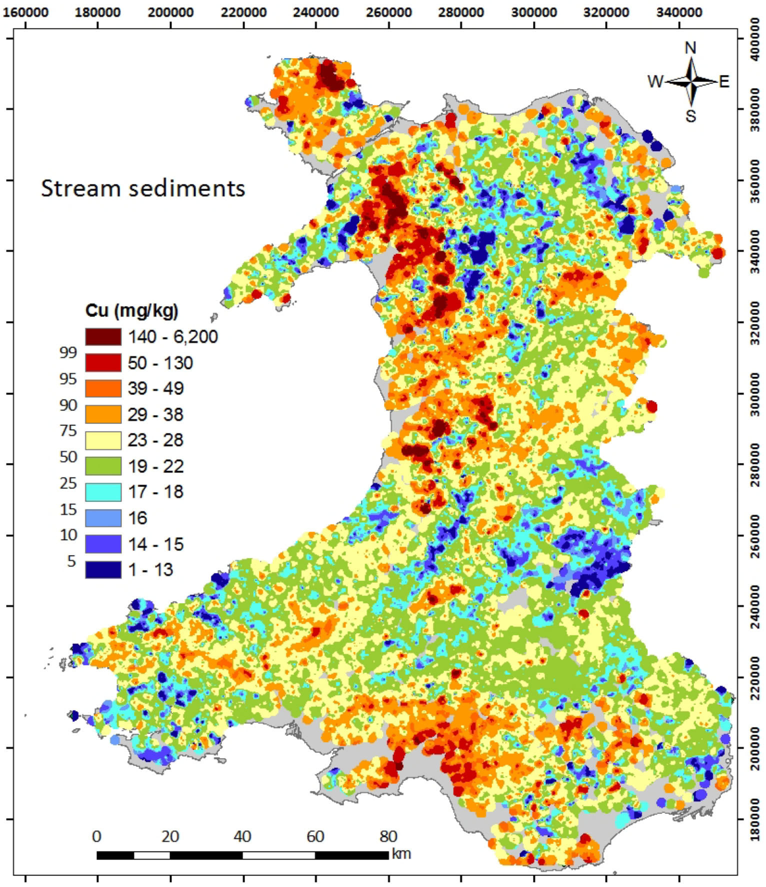
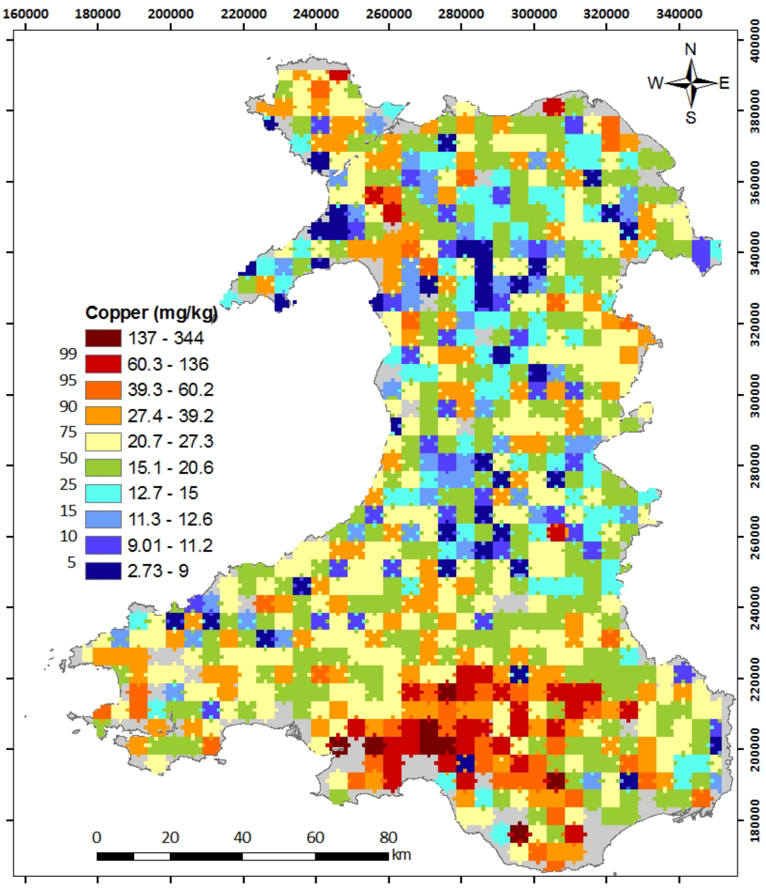


Figure 54: Interpolated map of topsoil Cu concentrations from the NSI(XRFS) data set only. Thresholds determined by using k-means cluster analysis



(a)



(b)

Figure 55: Copper in (a) G-BASE stream sediments and (b) NSI(XRFS) soils

5.2 DOMAIN SELECTION

The information shown below was used to compile the data for domains, with higher topsoil Cu concentrations than typical, in:

- non-ferrous metalliferous mineralisation and mining areas; and
- urban areas.

5.2.1 Non-ferrous metalliferous mineralisation and mining

Copper was historically exploited at many locations in Wales and there are also unexploited ores at areas such as Coed-y-Brenin (Rice and Sharp, 1976); these result in typically higher concentrations which can be clearly identified on the stream sediment data for Wales (Figure 55) (British Geological Survey, 2000).

A study of base metal mining areas in Wales (Alloway and Davies, 1971) used the Aeron catchment to calculate background concentrations in comparison to mineralised catchments of Ceredigion. A 'Normal range' of Cu 5.0 - 33 mg/kg is reported, compared to alluvium in the Rheidol (Cu range 17 - 42 mg/kg), a drainage catchment affected by mineralisation and mine wastes. Parys Mountain has a long copper mining history with major production being between 1768-1883. It was once the largest copper producer in Europe. Nearby, pasture soil is reported with very high Cu (116 - 1405 mg/kg) (Alloway and Davies, 1971).

However, Cu was noted as not being elevated in soils (mean of 28 mg/kg) in a survey of the worked Pb-Zn ore-field at Halkyn Mountain (Davies and Roberts, 1975), which is close to the background values reported in a study undertaken in northwest Pembrokeshire, which found a mean of 34 mg/kg (261 samples) (Bradley *et al.*, 1978). Archer and Hodgson (1987) report a median concentration of 18 mg/kg for England and Wales (1468 samples).

Using the regional soil data to intersect the non-ferrous metalliferous mineralisation and mining polygons gives very few samples in each mineralised area (Table 19). It is rather surprising that those mineralised areas which stand out in the G-BASE stream sediment data, and where Cu is known to be present as either an ore or widespread accessory mineral, are not indicated by the NSI(XRFS) soil data, and is most likely to be due to the sampling interval (1 per 25 km^2).

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Anglesey	3	49.7	5.4	5.4	26.9	117	117	1.5
north	22	25.1	6.9	12.0	17.2	28.5	116	3.0
Halkyn	15	25.2	8.9	16.1	22.1	27.1	64.8	1.8
central	31	18.8	9.7	12.3	17.7	22.7	41.1	1.2
south	7	52.6	16.2	17.3	18.3	80.8	165	1.8
none	1597	67.4	2.7	19.4	26.9	63.1	2131	7.3

Table 19: Summary statistics of topsoil Cu (mg/kg) for metalliferous mineralisation and mining areas

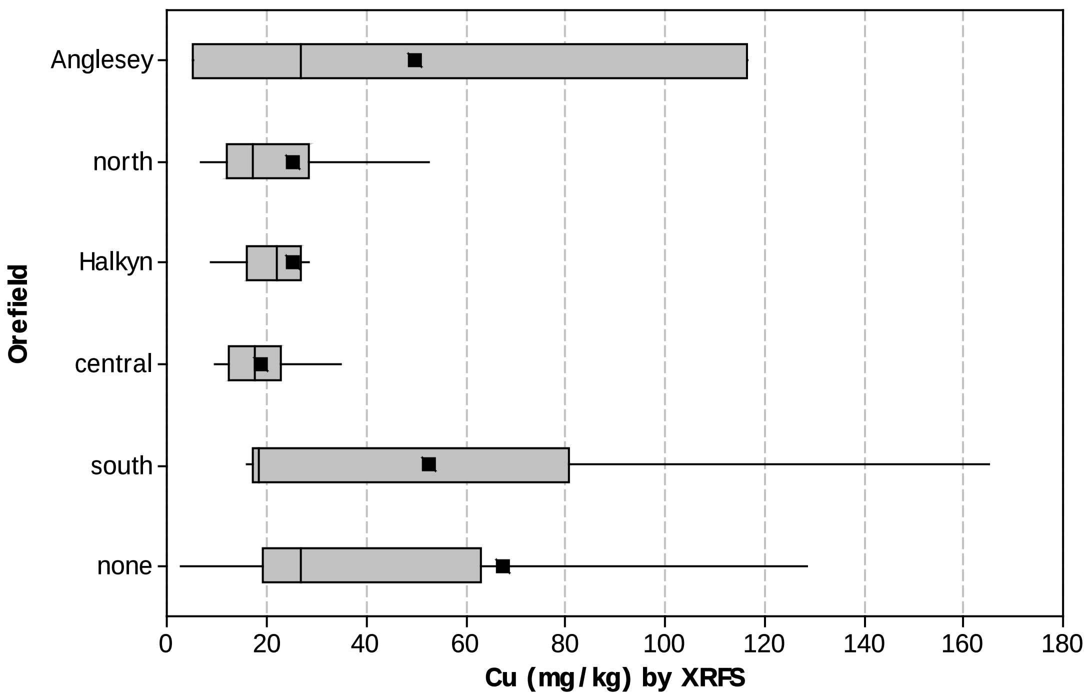


Figure 56: Boxplot of topsoil Cu concentrations in the mineralisation domains

5.2.2 Urban areas

The data from Cardiff and Swansea studies can be seen to be of higher concentration than the regional data in Table 18. In addition, the sample sites have been attributed to the extent of urbanisation and the positive relationship can again be seen in Table 20 and Figure 57. These clearly show a much higher concentration in the urban data set compared to all the other data. However, the topsoil Cu concentrations from the Loughor, Tawe and Neath valleys are much higher again (Figure 58), and more consistent with those data from Swansea (Table 18). This is consistent with the industrial history of the Lower Swansea valley, which at the height of Cu smelting had five smelters in operation (Hughes and Reynolds, 1992). Soils in part of the Swansea valley were found to have a median Cu concentration of 65 mg/kg, using a partial extraction (maximum 953 mg/kg; 70 samples) (Davies, 1997).

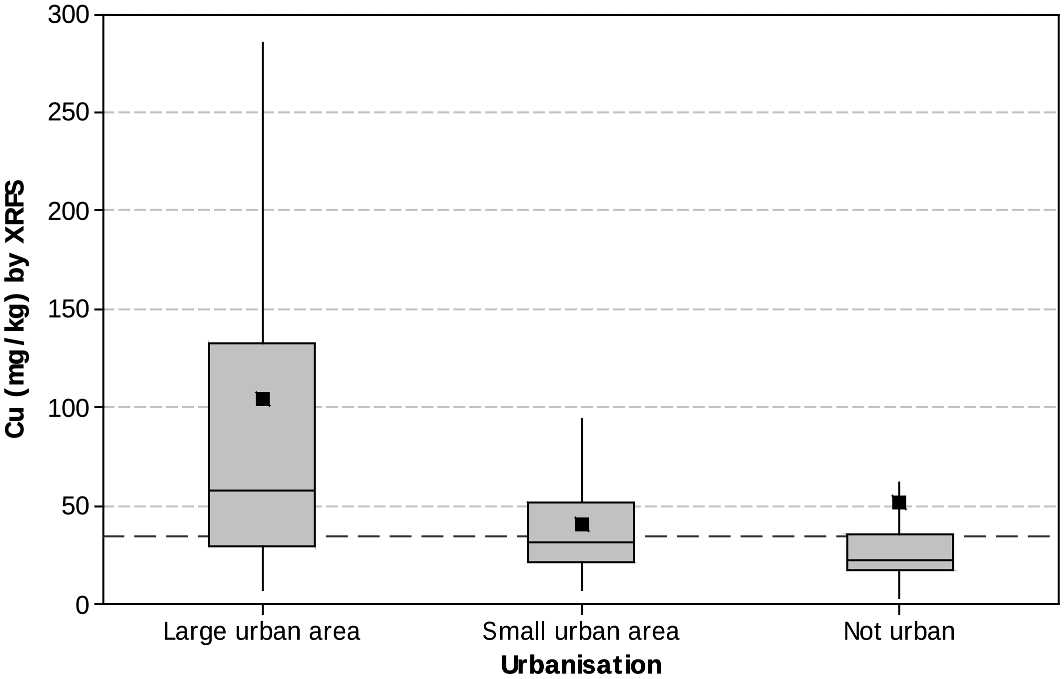


Figure 57: Boxplot of topsoil Cu concentrations over urban and non-urban areas

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Large urban area	452	105	6.9	28.9	57.9	132	1507	4
Small urban area	40	40.8	6.9	21.2	31.8	51.4	117	1
Not urban	1183	51.4	2.7	17.1	22.7	35.3	2131	9

Table 20: Summary statistics for topsoil Cu (mg/kg), over urban and non-urban areas

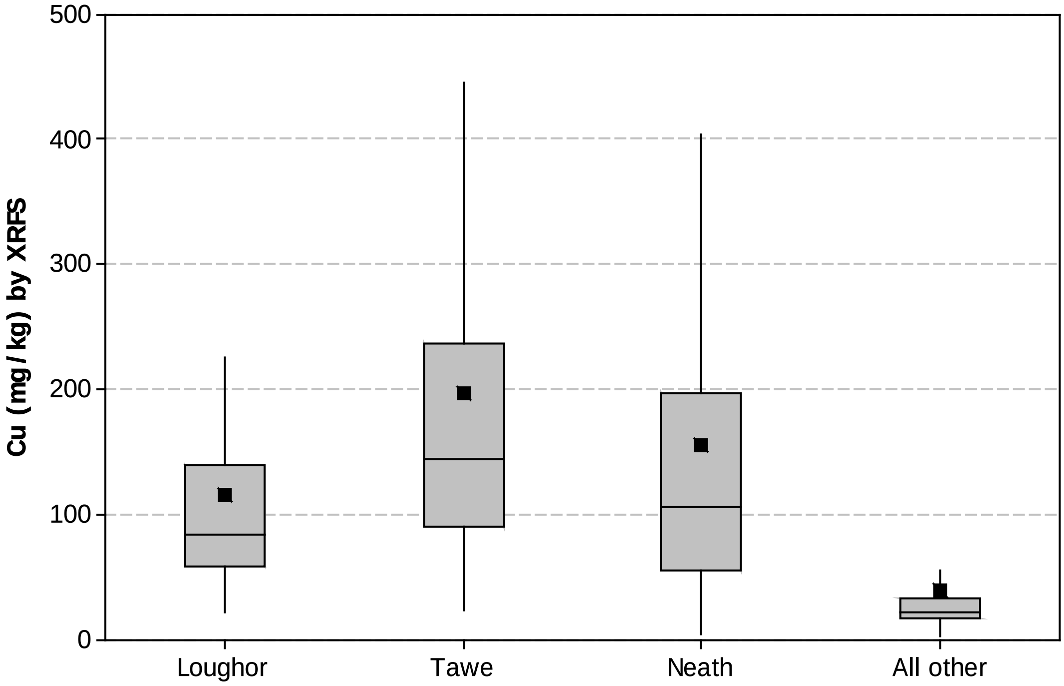


Figure 58: Boxplot of topsoil Cu data from the Loughor, Tawe and Neath valleys

5.3 DOMAIN DATA SUMMARY

The data have been ascribed to mineralisation and urban categories. The Mineralisation Domain comprises of all the non-ferrous metalliferous mineralisation and mining, other than Halkyn/Minera. This domain has been included despite the relative lack of elevated concentrations of soil Cu which would be expected in these areas. The urban categories have been divided into the Urban 1 Domain, which comprises the areas of Loughor, Tawe and Neath valleys. The Urban 2 Domain comprises all other 'Large urban area' extents in Wales (outside the Urban 1 area) based on the OS Stategi® data.

It can be seen that the two urban domains are higher in typical concentration than the remaining results (Table 21 and Figure 59) which are assigned to the Principal Domain. However, the difference between the Mineralisation and Principal Domains is small in the context of the urban areas (Figure 60). The domains for Cu are shown in Figure 61.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Mineralisation	61	26.0	5.4	12.4	17.8	25.0	165	3
Urban 1	342	167	5.1	66.2	121	200	1507	4
Urban 2	291	58.3	6.9	24.6	36.0	69.4	554	4
Principal	981	34.6	2.7	16.6	21.4	28.0	2131	14

Table 21: Summary statistics of topsoil Cu (mg/kg) concentrations in the domains

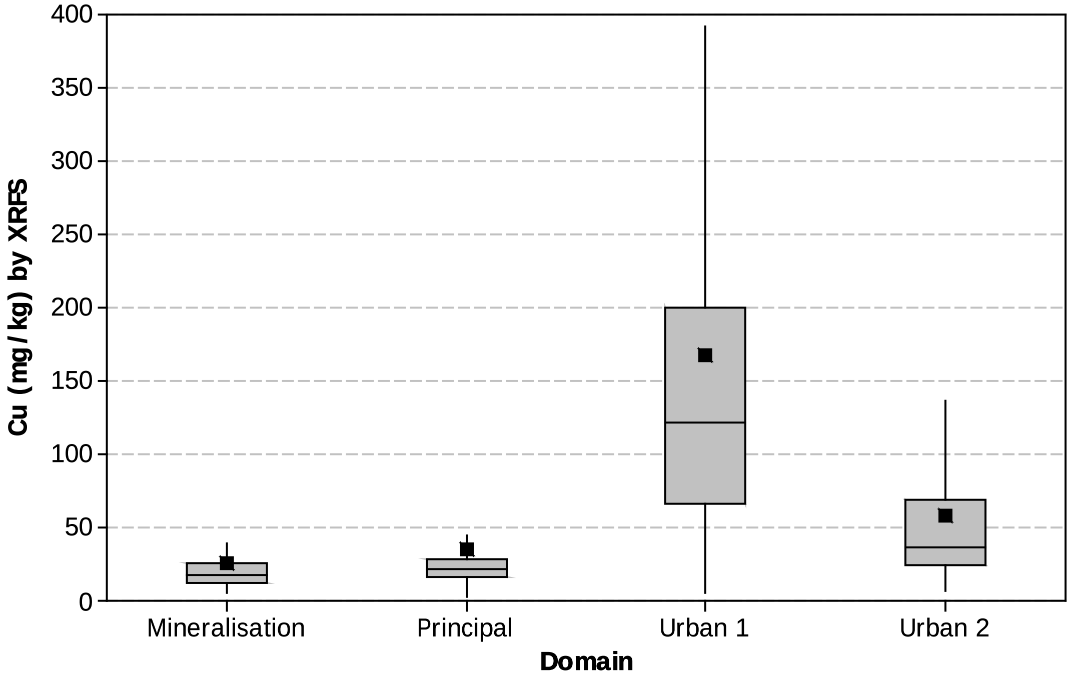


Figure 59: Boxplot of Cu data classified by domains

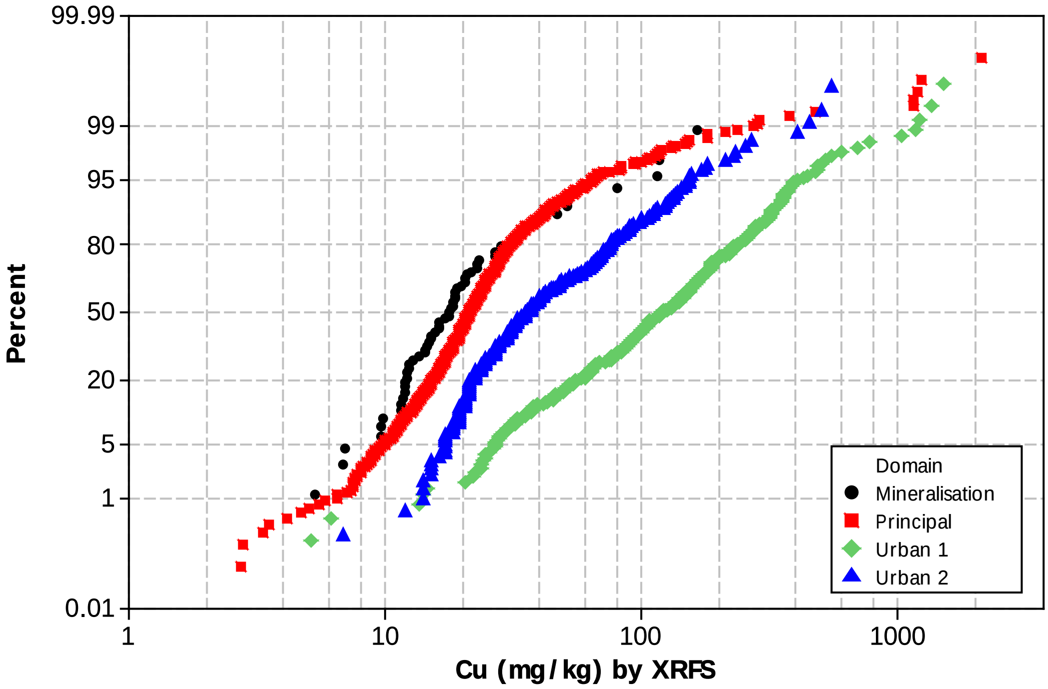
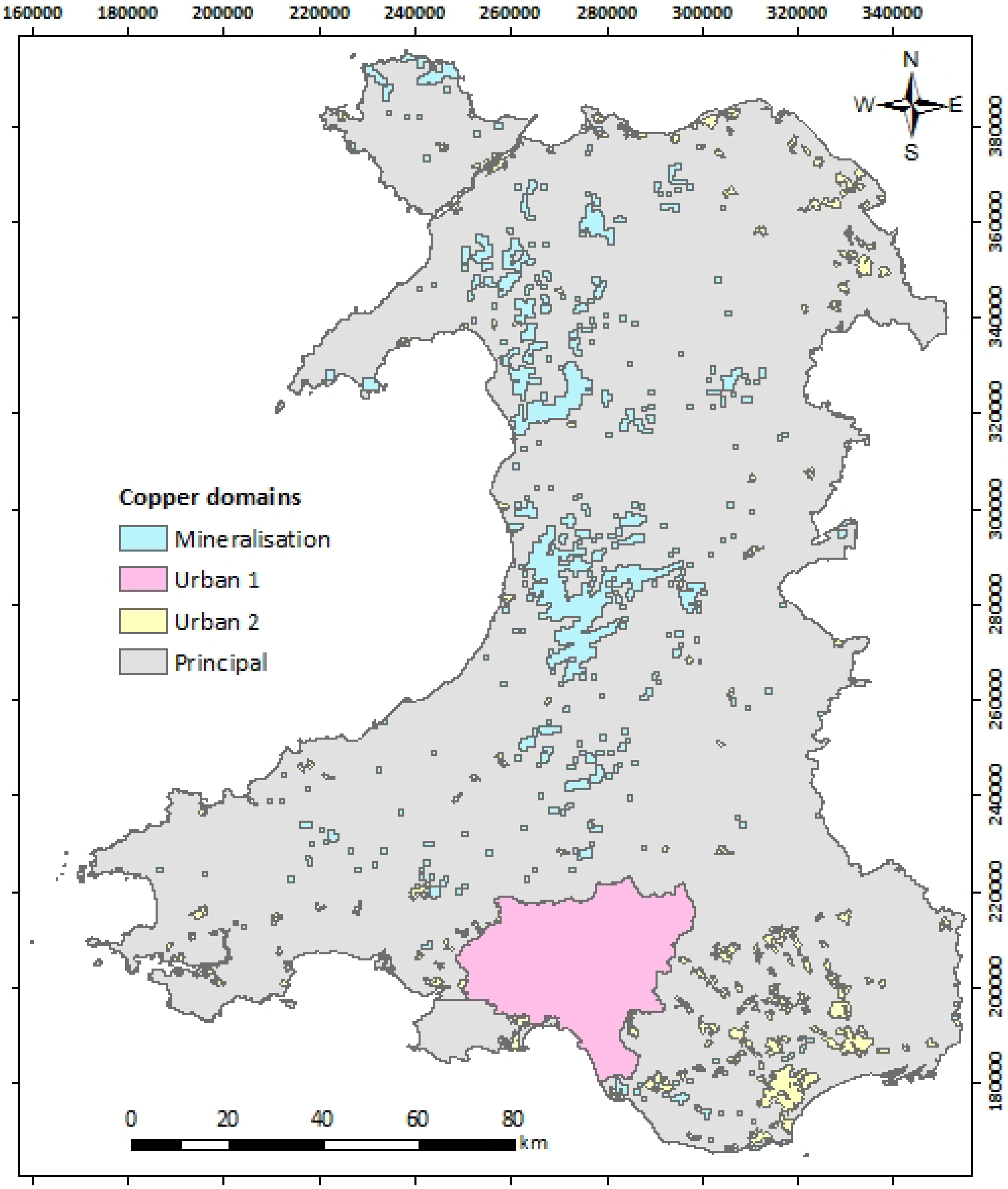


Figure 60: Probability plot of Cu topsoil data classified by domains



Contains Ordnance Survey data © Crown copyright and database rights (2012)
Urban areas defined from OS Strategi® used under OS OpenData™ terms
Mineralisation after the Ove Arup data

Figure 61: Map showing the location of Cu domains

5.4 Cu NBC CALCULATION

The NBC calculations for Cu have been carried out according to the method described in Cave *et al*. (2012).

5.4.1 Principal Domain

Untransformed skewness and octile skew coefficients fail the test criteria for underlying Gaussian distribution. The $\log_e$ transformation produces a skewness coefficient (SC) that is >1, but octile skew (OS) is <0.2, suggesting a log distribution with outliers (Figure 62 and Figure 63), requiring a robust percentile calculation. Percentiles are given in Table 22 and an example NBC is shown in Figure 64.

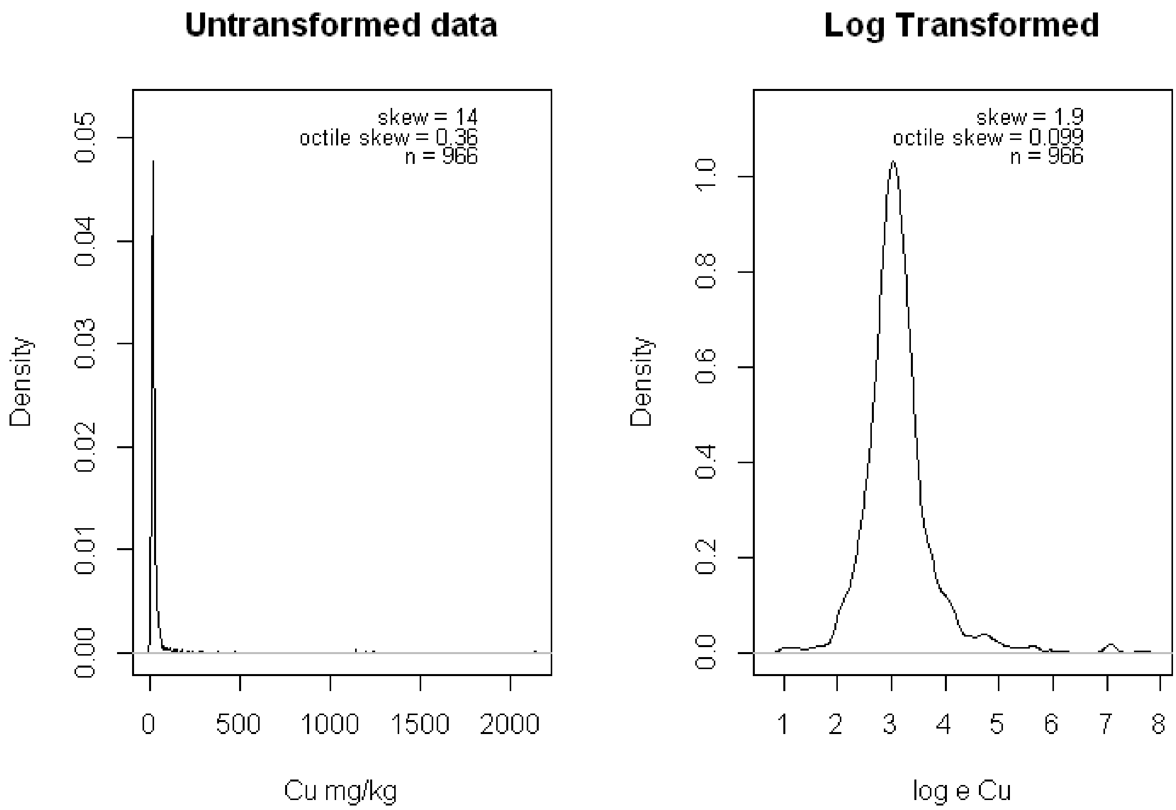


Figure 62: Density distributions for the raw data and the $\log_e$ transformed data for Cu in the Principal Domain (n = number of samples)

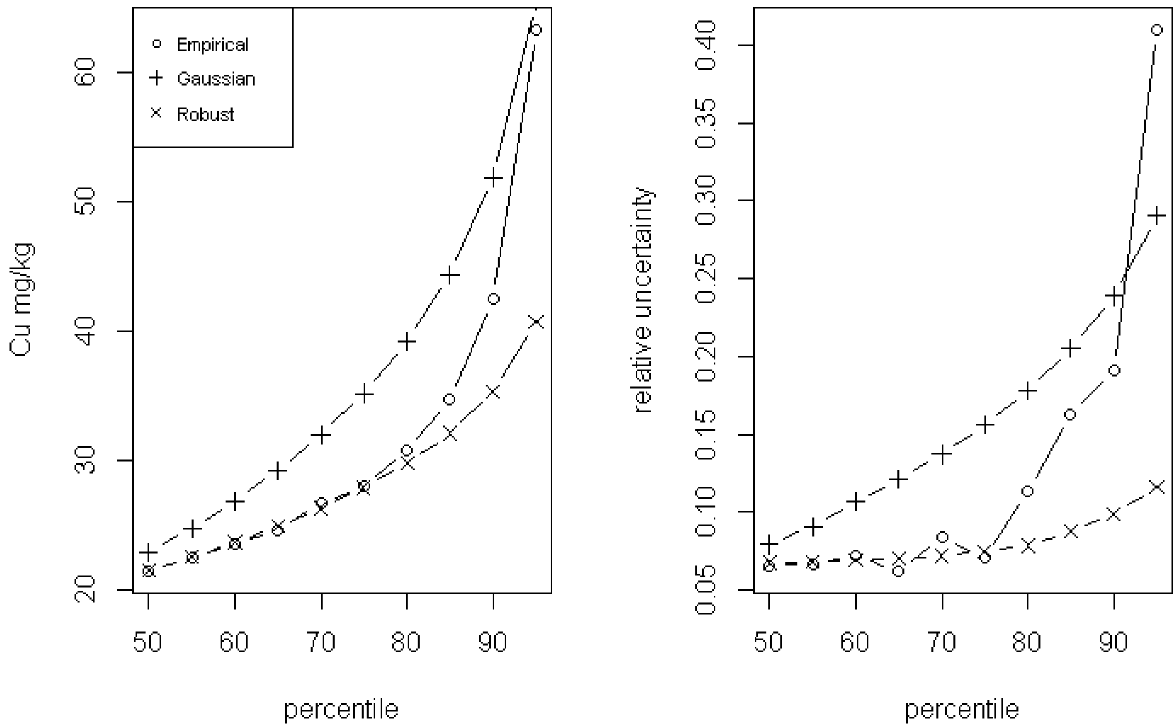


Figure 63: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Cu in the Principal Domain

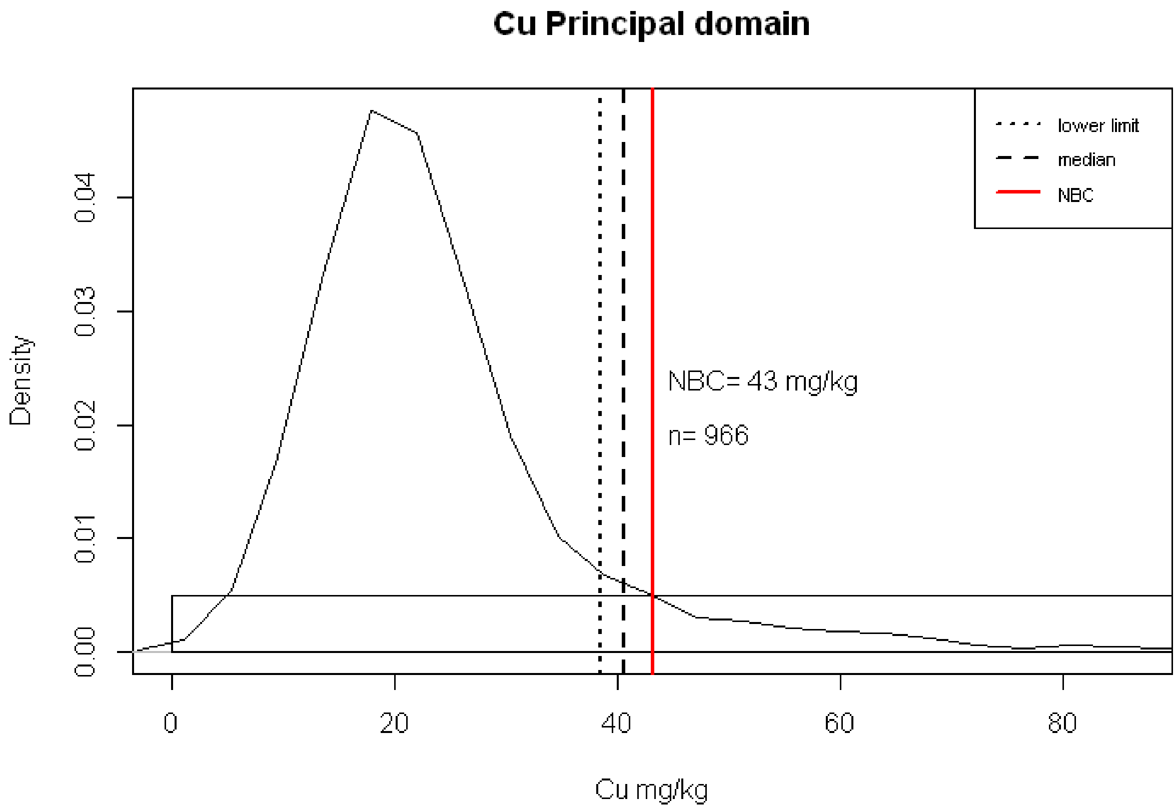


Figure 64: Summary density plot and histogram of the distribution for Cu in the Principal Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	20.7	21.4	22.2	21.9	22.8	23.8	**20.6**	**21.4**	**22.0**
55	21.8	22.5	23.1	23.7	24.7	25.9	**21.7**	**22.5**	**23.1**
60	22.8	23.5	24.4	25.5	26.9	28.3	**22.8**	**23.6**	**24.3**
65	24.2	24.6	25.6	27.5	29.2	31.0	**24.0**	**24.9**	**25.7**
70	25.4	26.6	27.6	29.7	32.0	34.2	**25.3**	**26.3**	**27.1**
75	27.4	28.0	29.2	32.4	35.2	38.0	**26.8**	**27.9**	**28.8**
80	29.0	30.7	32.5	35.7	39.2	42.7	**28.6**	**29.7**	**30.9**
85	32.5	34.7	38.3	39.9	44.4	49.0	**30.8**	**32.1**	**33.5**
90	39.9	42.4	48.0	45.9	51.9	58.3	**33.8**	**35.3**	**37.1**
95	54.5	63.4	81.8	56.4	65.6	75.3	**38.7**	**40.7**	**43.3**

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 22: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Cu (mg/kg) in the Principal Domain

5.4.2 Mineralisation domain

Untransformed and $\log_e$ transformed skewness and octile skew coefficients fail the test criteria for an underlying Gaussian distribution (Figure 65). Therefore a Box-Cox transformation was applied (Figure 66 and Figure 67). Percentiles are summarised in Table 23 and an example NBC is given in Figure 68.

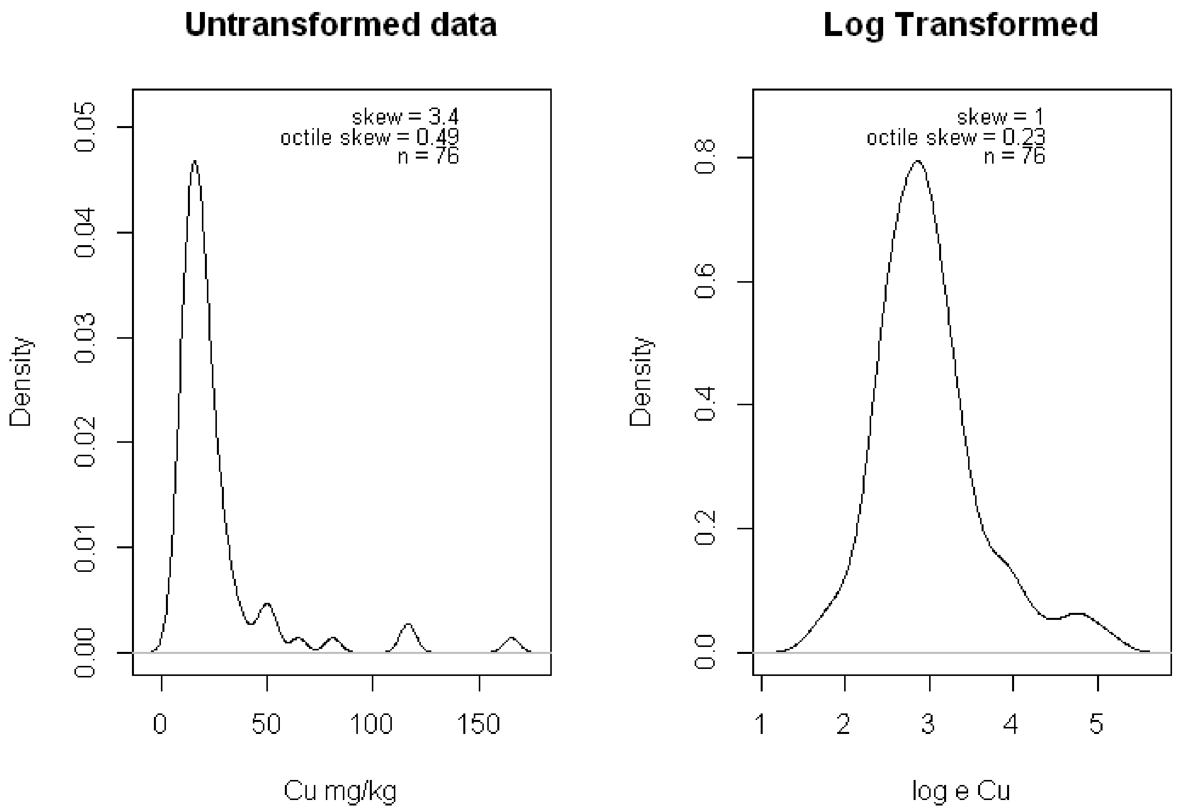


Figure 65: Density distributions for the raw data and the $\log_e$ transformed data for Cu in the Mineralisation Domain (n = number of samples)

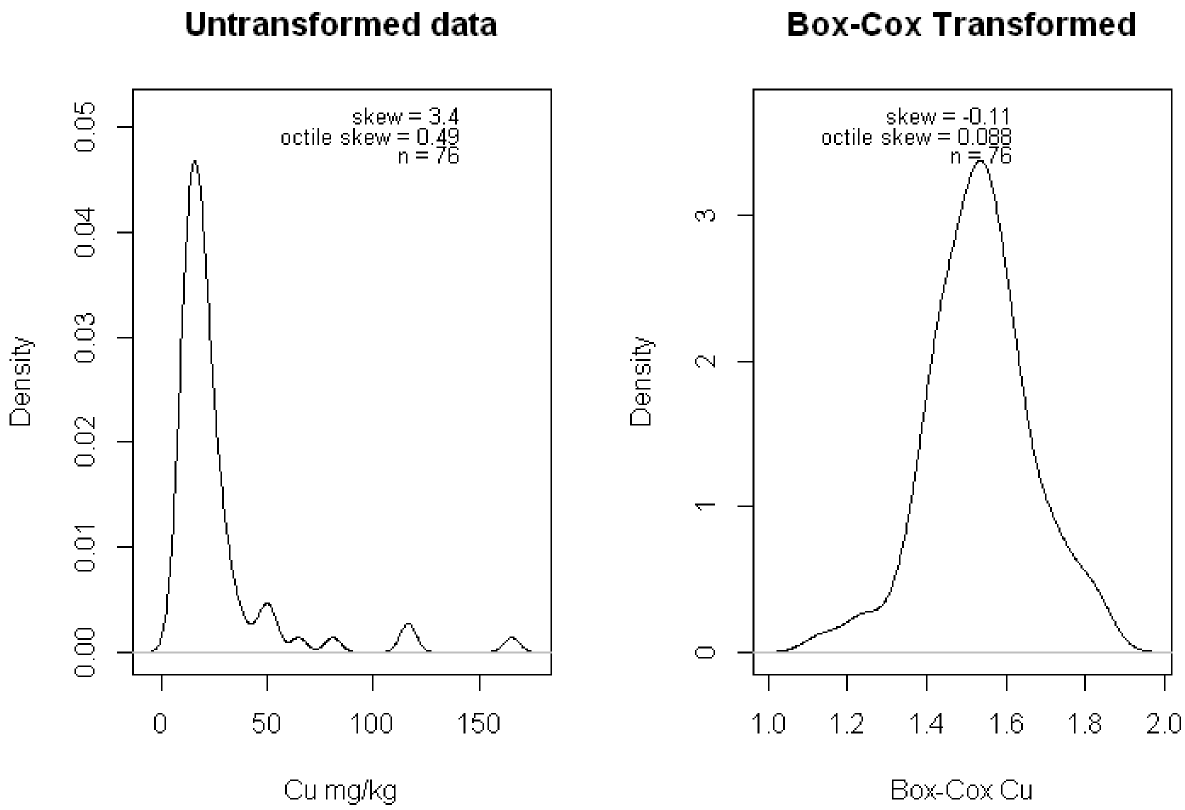


Figure 66: Density distributions for the raw data and the Box-Cox transformed data for Cu in the Mineralisation Domain (n = number of samples)

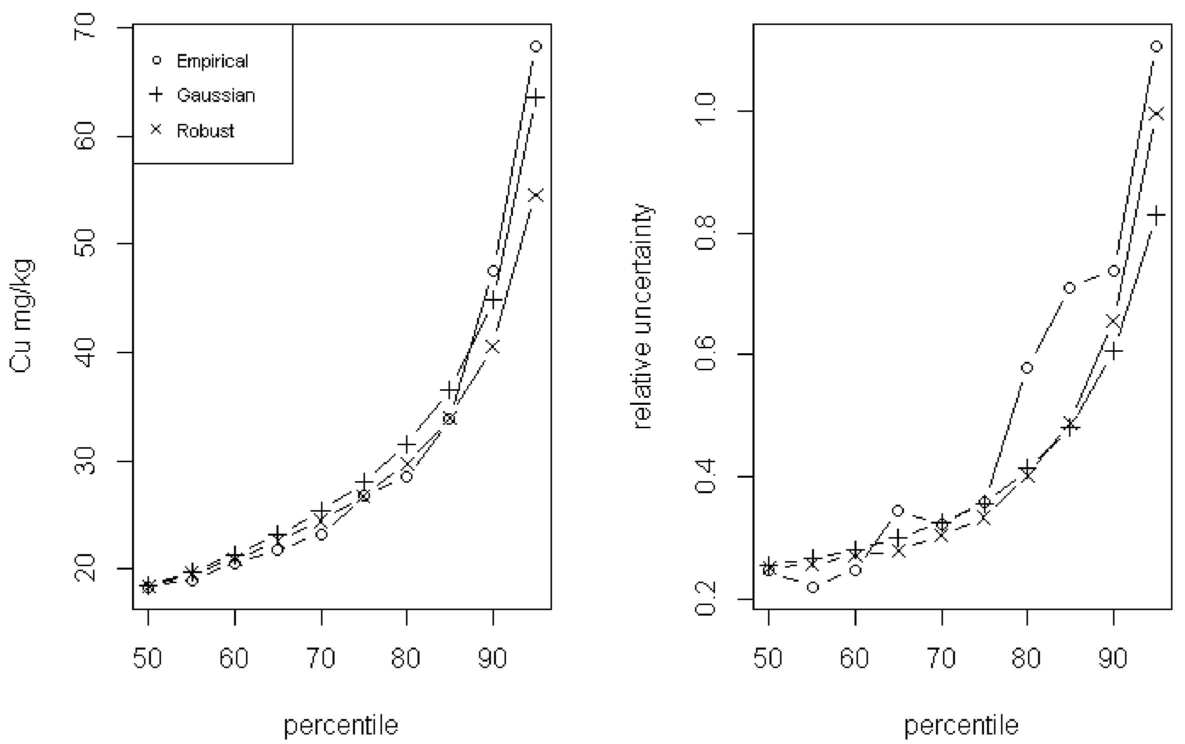


Figure 67: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Cu in the Mineralisation Domain

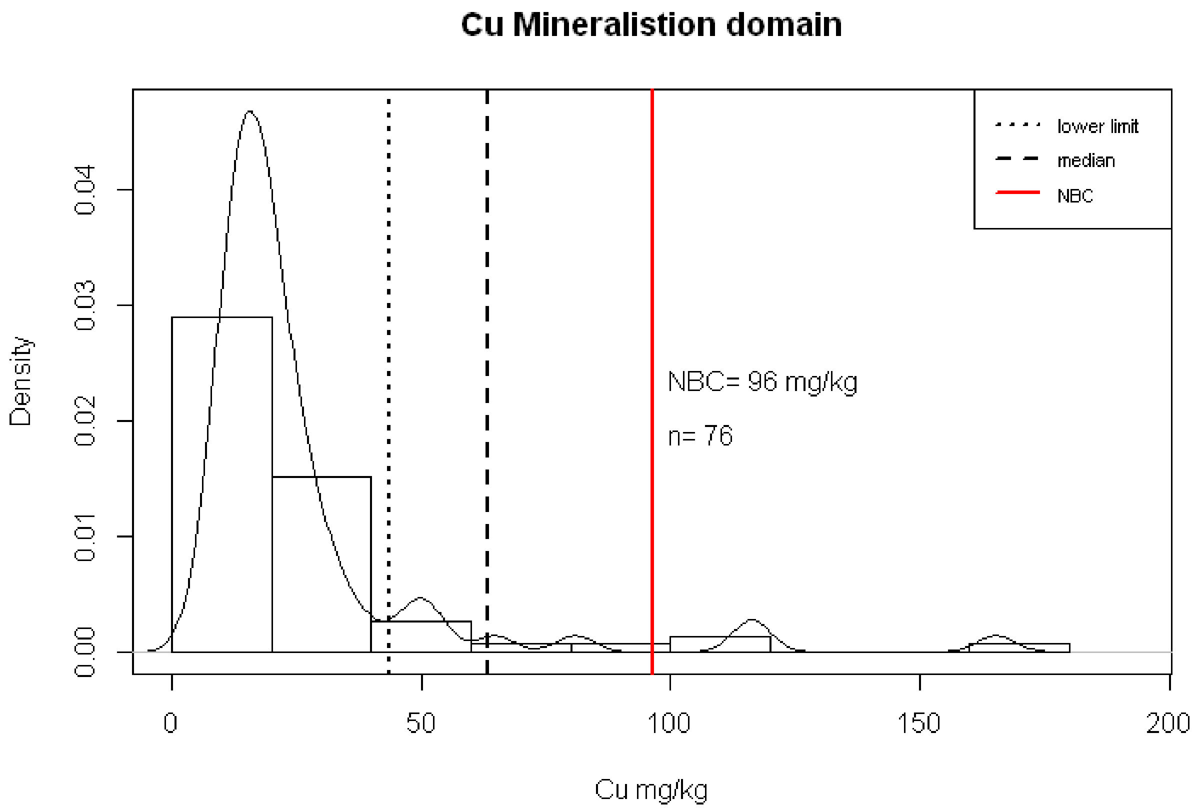


Figure 68: Summary density plot and histogram of the distribution for Cu (mg/kg) in the Mineralisation Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	16	18	21	**16**	**18**	**21**	16	18	21
55	18	19	22	**17**	**20**	**23**	17	20	22
60	18	21	23	**19**	**21**	**25**	18	21	24
65	19	22	26	**20**	**23**	**27**	19	23	26
70	20	23	28	**22**	**25**	**30**	21	24	28
75	22	27	31	**24**	**28**	**34**	22	27	31
80	23	28	39	**26**	**32**	**39**	24	30	36
85	27	34	51	**29**	**37**	**47**	26	34	43
90	30	48	65	**34**	**45**	**61**	29	41	56
95	41	68	117	**44**	**64**	**96**	34	55	88

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 23: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Cu in the Mineralisation Domain

5.4.3 Urban 1 Domain

Untransformed skewness coefficient and octile skew fail the test criteria for underlying Gaussian distribution. When $\log_e$ transformed was used, the skewness coefficient was <1 and octile skew <0.2, indicating an underlying Gaussian distribution for which parametric percentiles were used (Figure 69 and Figure 70). Percentiles are given in Table 24 and an example NBC shown in Figure 71.

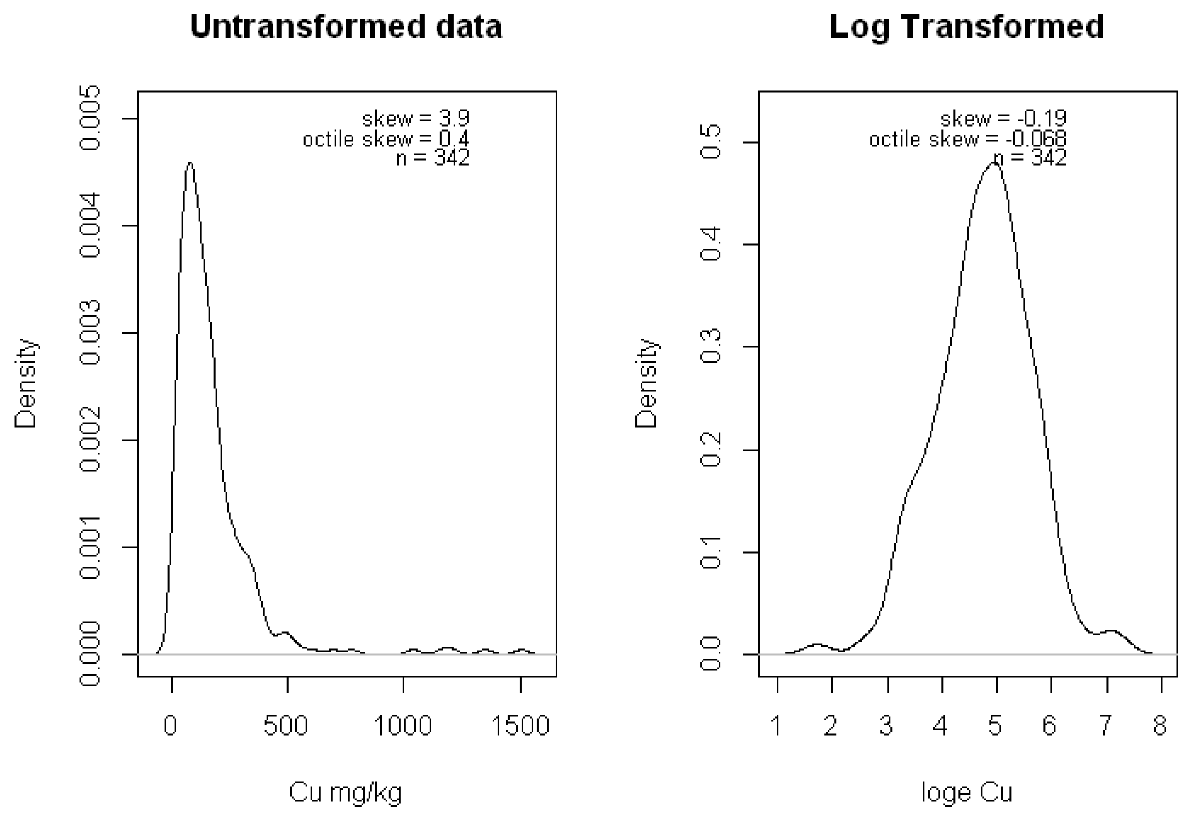


Figure 69: Density distributions for the raw data and the $\log_e$ transformed data for Cu in the Urban 1 Domain (n = number of samples)

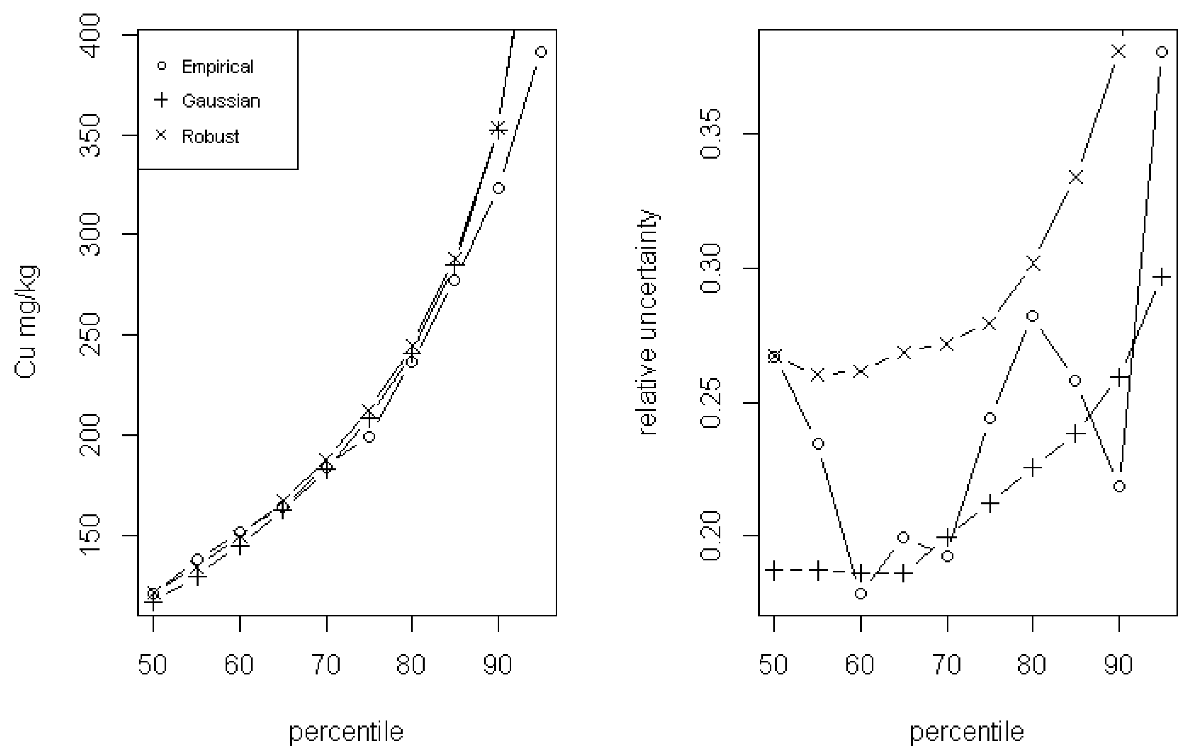


Figure 70: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Cu in the Urban 1 Domain

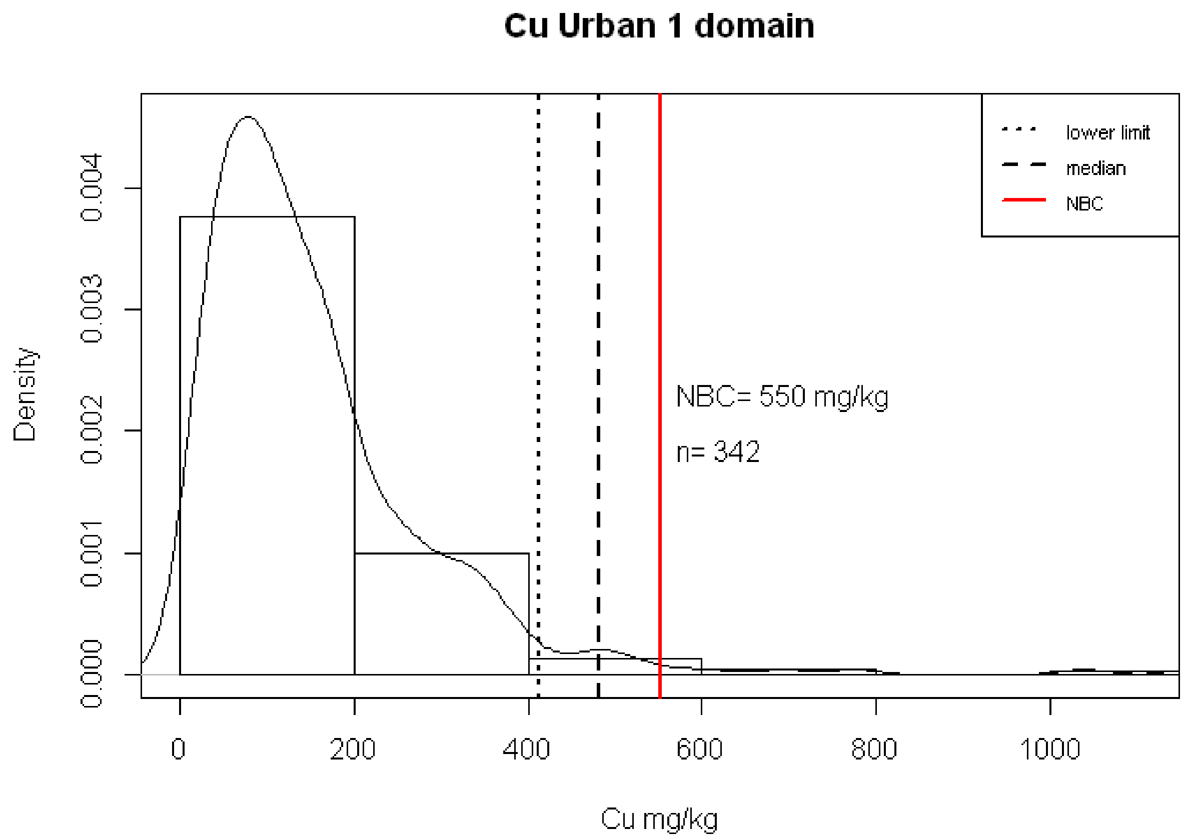


Figure 71: Summary density plot and histogram of the distribution for Cu in the Urban 1 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	107	121	138	**107**	**116**	**127**	107	121	139
55	120	138	152	**119**	**130**	**142**	119	134	153
60	137	151	164	**133**	**145**	**159**	131	149	170
65	152	164	184	**149**	**162**	**178**	147	167	190
70	164	184	199	**167**	**183**	**202**	165	188	215
75	184	199	228	**189**	**208**	**230**	186	213	245
80	201	237	267	**216**	**241**	**267**	211	244	283
85	243	278	313	**254**	**285**	**318**	242	288	336
90	286	324	356	**310**	**352**	**398**	288	353	420
95	356	391	485	**417**	**483**	**553**	369	479	584

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 24: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Cu (mg/kg) in the Urban 1 Domain

5.4.4 Urban 2 Domain

Untransformed skewness and octile skew coefficients fail the test criteria for underlying Gaussian distribution. Although Log_e transform gave an octile skew >0.2, the skewness coefficient was less than one (Figure 72), which indicated an underlying Gaussian distribution, for which parametric percentiles were used to calculate the NBC (Figure 72 and Figure 73). Percentiles are given in Table 25 and an example NBC shown in Figure 74.

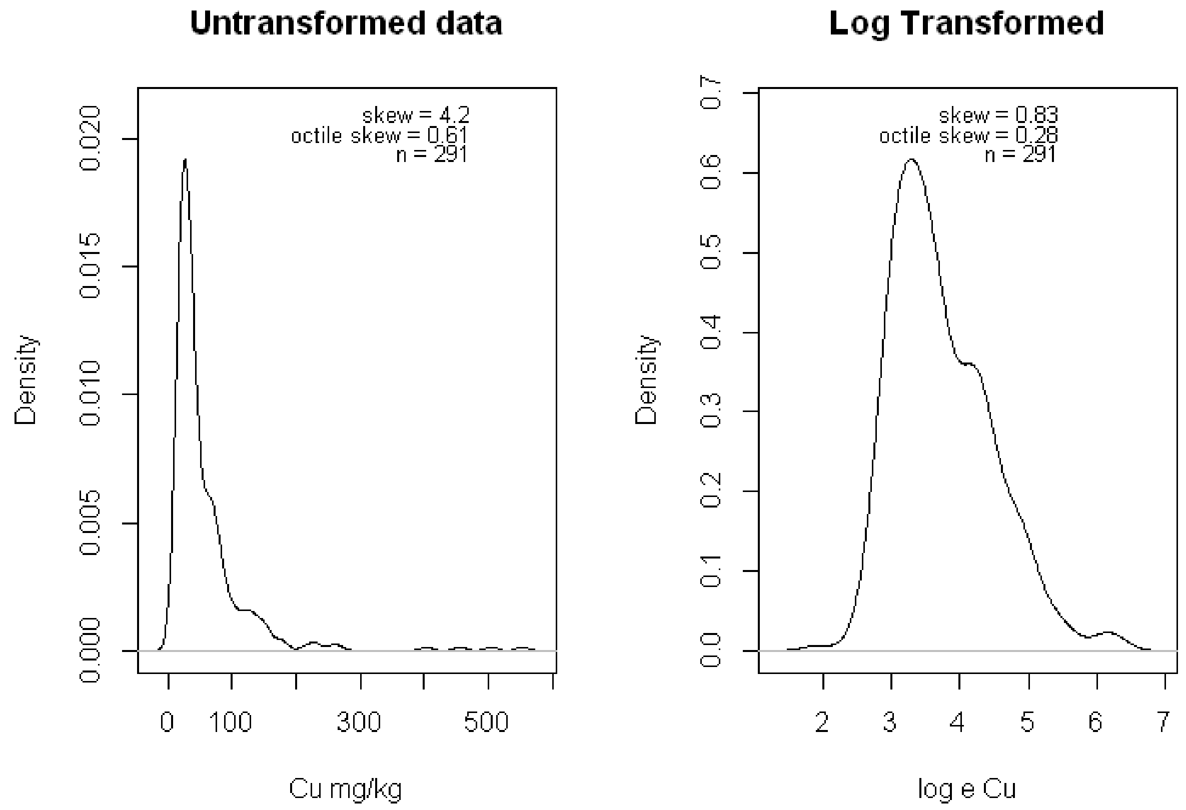


Figure 72: Density distributions for the raw data and the log$_e$ transformed data for Cu in the Urban 2 Domain (n = number of samples)

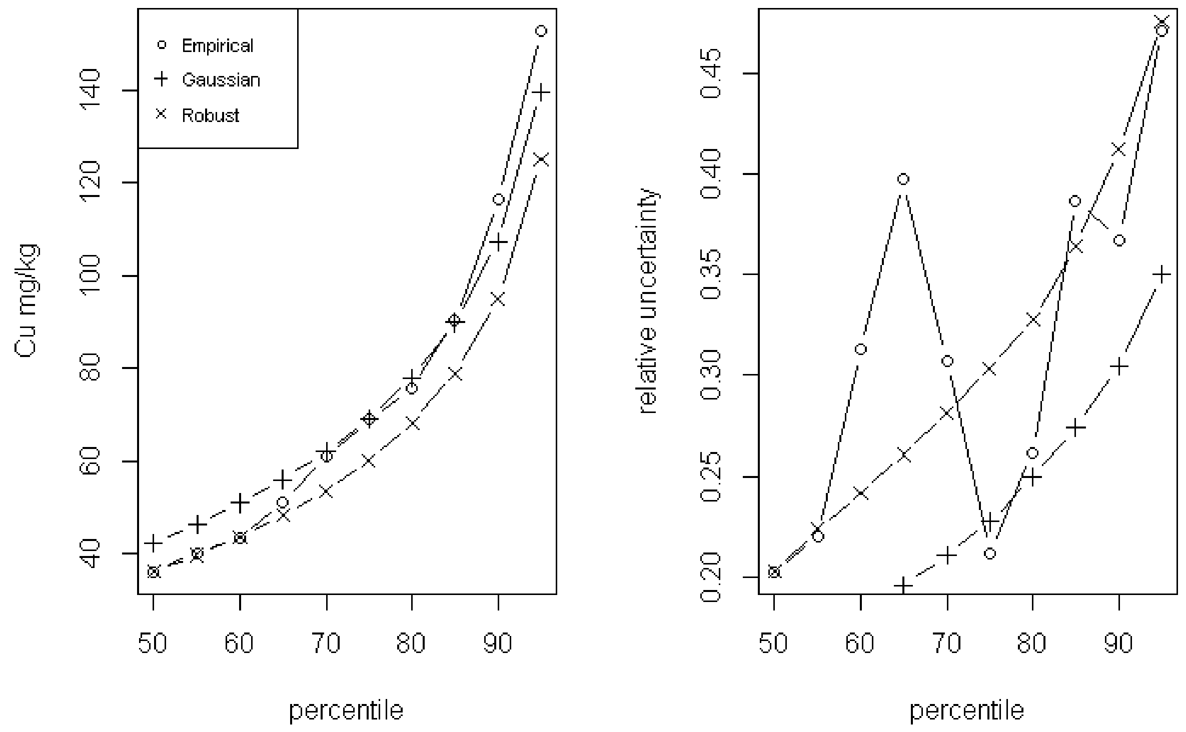


Figure 73: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Cu in the Urban 2 Domain

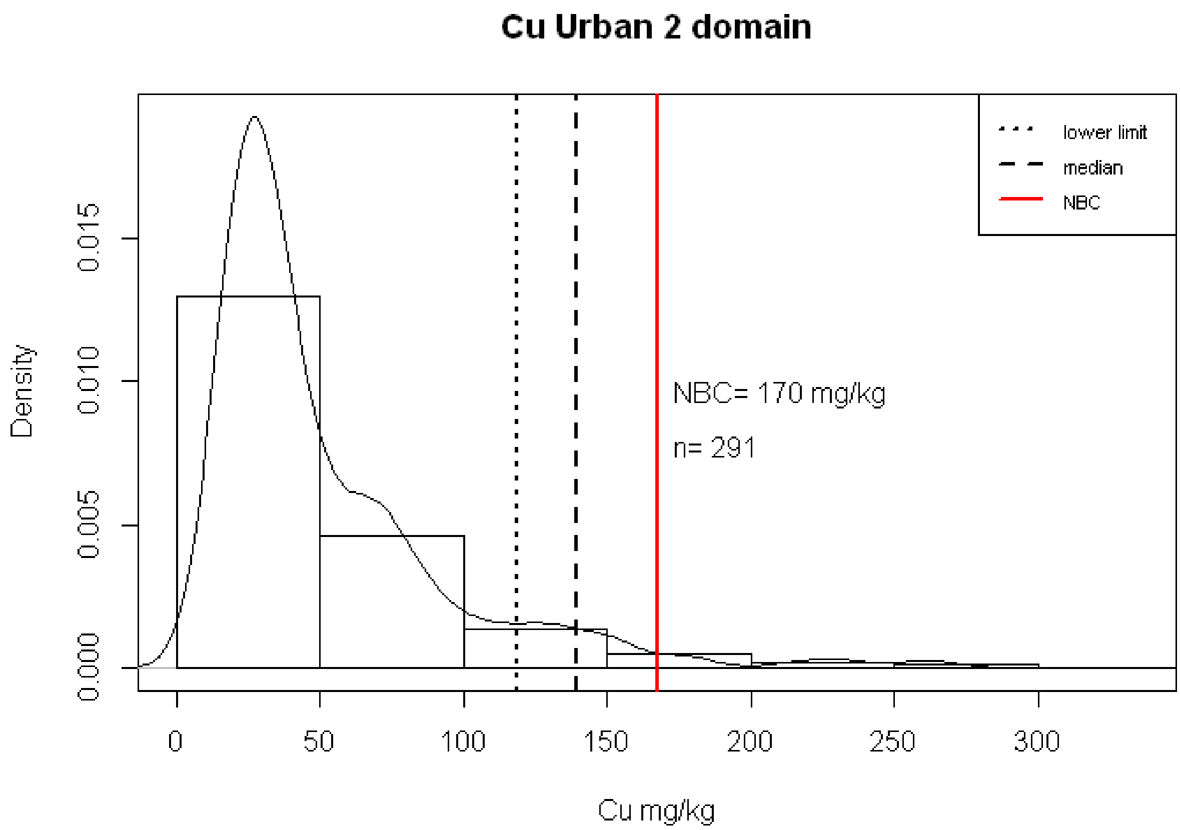


Figure 74: Summary density plot and histogram of the distribution for Cu in the Urban 2 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	32.9	36.0	40.2	39.1	42.3	46.2	32.9	36.0	40.2
55	35.5	40.2	44.4	42.6	46.4	50.8	35.6	39.6	44.5
60	39.2	43.3	52.7	46.6	50.9	55.9	38.6	43.6	49.2
65	42.8	51.1	63.1	50.9	56.0	61.8	42.0	48.2	54.6
70	50.6	61.0	69.4	55.9	61.9	68.9	45.9	53.6	61.0
75	60.5	68.8	75.1	61.7	69.0	77.4	50.5	60.0	68.7
80	69.4	75.6	89.2	69.0	77.9	88.5	56.1	68.1	78.4
85	76.1	90.2	111	78.6	89.8	103	63.5	78.9	92.3
90	93.3	116	136	92.8	107	126	74.2	94.9	113
95	131	153	203	119	140	**167**	92.7	125	152

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 25: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Cu (mg/kg) in the Urban 2 Domain

5.4.5 Copper NBC summary

Summary of all Cu percentiles is shown in Table 26, with details of the number of samples used and the transformation applied as well as the area of Wales each covers.

Percentile	Principal Domain n = 966 Area = 18,400 km^2 Transformation: $\log_e$			Mineralisation Domain n = 76 Area = 1100 km^2 Transformation: Box-Cox		
	lower	middle	upper	lower	middle	upper
50	20.6	21.4	22.0	**16**	**18**	**21**
55	21.7	22.5	23.1	**17**	**20**	**23**
60	22.8	23.6	24.3	**19**	**21**	**25**
65	24.0	24.9	25.7	**20**	**23**	**27**
70	25.3	26.3	27.1	**22**	**25**	**30**
75	26.8	27.9	28.8	**24**	**28**	**34**
80	28.6	29.7	30.9	**26**	**32**	**39**
85	30.8	32.1	33.5	**29**	**37**	**47**
90	33.8	35.3	37.1	**34**	**45**	**61**
95	38.7	40.7	**43.3**	**44**	**64**	**96**

Percentile	Urban 1 Domain n = 342 Area = 1200 km^2 Transformation: $\log_e$			Urban 2 Domain n =291 Area = 500 km^2 Transformation: $\log_e$		
	lower	middle	upper	lower	middle	upper
50	107	116	127	39.1	42.3	46.2
55	119	130	142	42.6	46.4	50.8
60	133	145	159	46.6	50.9	55.9
65	149	162	178	50.9	56.0	61.8
70	167	183	202	55.9	61.9	68.9
75	189	208	230	61.7	69.0	77.4
80	216	241	267	69.0	77.9	88.5
85	254	285	318	78.6	89.8	103
90	310	352	398	92.8	107	126
95	417	483	**553**	119	140	**167**

n = number of samples used in calculation

Table 26: Copper (mg/kg) NBC summary table

6 Mercury (Hg)

6.1 EXPLORATORY DATA ANALYSIS

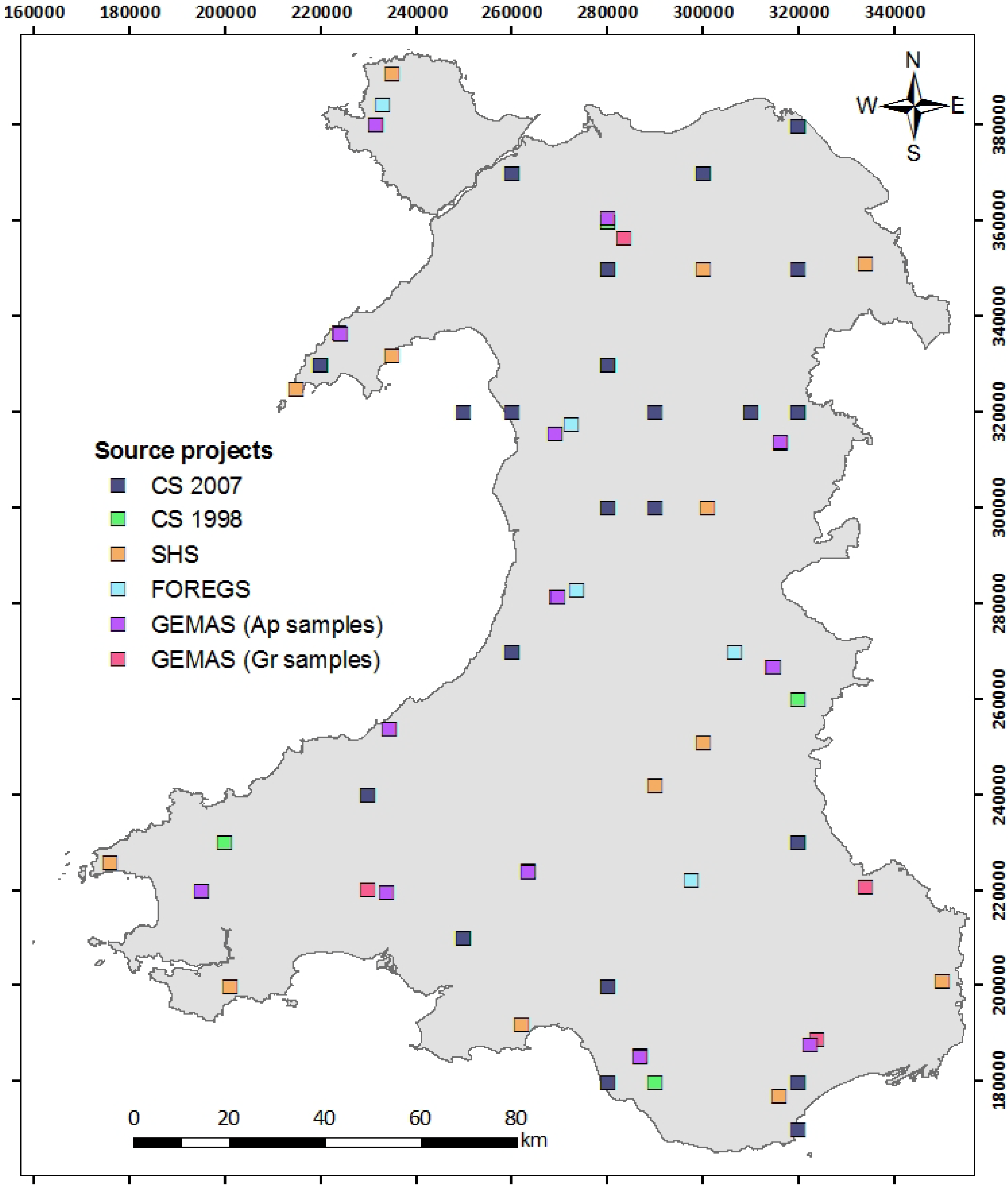
There is no NSI or G-BASE topsoil Hg data for any of the data sets used in the assessment of other contaminants in this report. Although Hg was determined on both phases of NSI original sample analysis, the method had a relatively high detection limit (0.1 mg/kg), with over 50% of the data falling below that detection limit, and the remainder of the data being reported to only one significant figure. Therefore, these data were not used here.

Results from a variety of different studies have been utilised, all with a much lower sample density and poorer systematic coverage. These data sets are summarised in Table 27. Project data sources are reviewed in more detail in Ander *et al*. (2011) and sample locations are shown in Figure 75. It can be seen that detection limits vary by over ten times (Table 27), and that the different project data sources have different statistical properties, with considerable variation in reported central tendency and range values (Table 28). The GEMAS data have been reported within the full data set by Ottesen *et al.* (2013) to have a median of 0.03 mg/kg, and show a strong relationship on the continental scale to soil organic matter.

There is an inherent problem in that these studies have targeted specific geographical areas and/or land uses – most particularly there are only three samples which have been collected which were classified as the source project as 'urban', with all other data specifically targeting rural land uses. It is therefore very difficult to establish the extent to which systematic bias arising from sample collection, preparation, digestion and analytical methods contribute to the apparent difference between these data sources in Figure 76 and Table 29.

Data source	Depth, support and preparation	Digestion step	Instrument and Laboratory	Detection limit (mg/kg)
CS 1998	15 cm deep (× 8 cm diameter) core. Single core. Drying temperature/ sieving not recorded.	not specified.	not specified.	not specified.
CS 2007	15 cm deep. 15 cm deep (× 8 cm diameter) core. Single core. Air dried. Sieving not recorded.	*Aqua-regia* microwave digestion.	ICP-MS. Laboratory not reported.	0.067
SHS	5 cm deep. 3 cores per sample, support not specified but 3 samples collected within a 20×20 m square. Stored 4°C. Not specified whether sieved/dried.	*Aqua-regia*.	Cold-vapour atomic absorption spectrometry (CV-AAS). EA laboratory.	0.07
FOREGS	0-25 cm. 3-5 sample composite. Dried at 40°C, sieved <2 mm.	n/a. Heated to 850°C to drive off Hg.	Hg analyser. Hungarian Geological Survey laboratory.	0.0001
GEMAS	0-20 cm on arable; 0-10 cm permanent pasture. Five spade-dug pits in 10 × 10 m, ~3.5kg sample collected. Air-dried and <2 mm sieved.	n/a. Heated to 850°C to drive off Hg.	Hg analyser. Hungarian Geological Survey laboratory.	0.0001

Table 27: Summary of available Hg data



Note that locations from SHS (Soil and Herbage Survey) and CS (Countryside Survey) are rounded to ±10 km, and locations cannot be checked any more closely than that distance

Figure 75: Location of samples, classified by originating project, used to establish Hg normal background concentrations

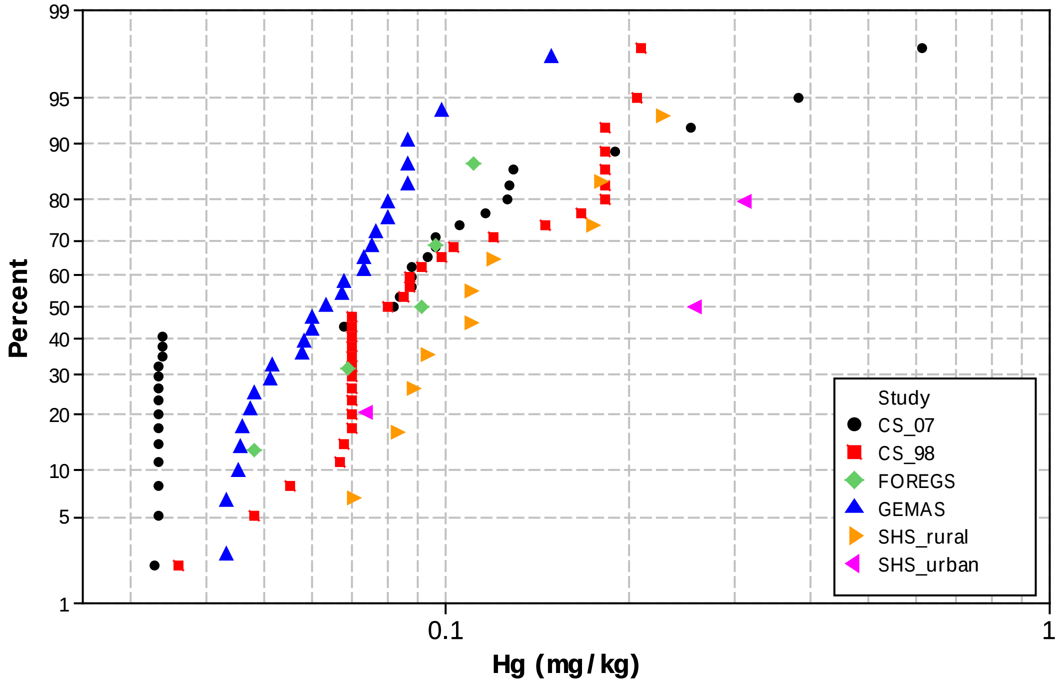
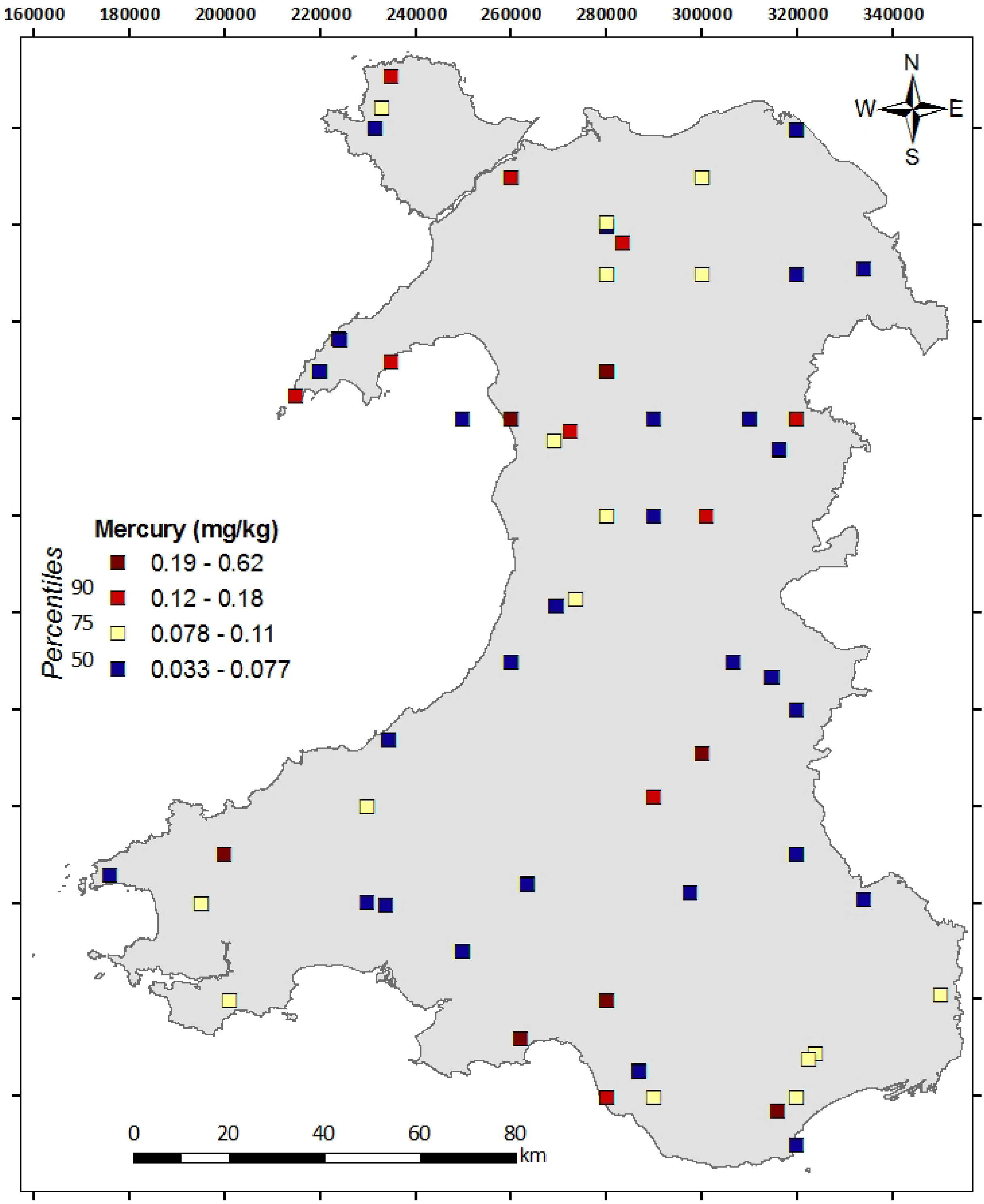


Figure 76: Probability plot of Hg data by source data set

Data source	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
CS 1998	CS_98	33	0.10	<0.07	<0.07	0.08	0.16	0.21
CS 2007	CS_07	33	0.10	<0.067	<0.067	0.08	0.11	0.62
SHS (rural)	SHS_rural	10	0.13	<0.07	0.09	0.11	0.18	0.23
SHS (urban)	SHS_urban	3	0.22	<0.07	<0.07	0.26	0.31	0.31
FOREGS	FOREGS	5	0.083	0.048	0.059	0.091	0.104	0.111
GEMAS	GEMAS	27	0.067	0.043	0.048	0.064	0.080	0.149

Summary statistics are presented with significant figures proportional to the detection limit sensitivity (Table 27).

Table 28: Summary statistics of Hg (mg/kg) data by source data set



Note that locations from SHS (Soil and Herbage Survey) and CS (Countryside Survey) are rounded to ±10 km, and locations cannot be checked any more closely than that distance.

Figure 77: Map of Hg concentrations in topsoil

6.2 DOMAIN SELECTION

Urbanisation and some instances of ore-field mineralisation and have been shown by previous studies to result in a typically higher concentration of topsoil Hg (see Ander *et al.*, 2012).

6.2.1 Urban areas

Results of analysis show that topsoil Hg concentrations are likely to have a positive relationship with the extent of urbanisation (Figure 78, Figure 79 and Figure 80), but the very small number of urban analyses available (Table 29 and Table 30), as classified by source project or the urbanisation extent, mean that this cannot be formally tested.

Coal-burning is considered major atmospheric source of Hg in the EU, and other industrial/domestic sources exist (Rodrigues *et al*., 2006). These authors show high short-distance variation in concentrations (more than ten times within 50 m) and also that concentrations vary greatly between two parks in Glasgow, that is significant in the context of the other urban centre parks in Europe which were sampled.

Data source	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Urban	3	0.22	<0.07	0.07	0.26	0.31	0.31	-1
Rural	108	0.10	<0.07	0.06	0.08	0.10	0.62	4

Table 29: Summary statistics of topsoil Hg (mg/kg) categorised by source project classification

Data source	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Large urban area	7	0.14	<0.07	0.07	0.09	0.26	0.31	1
Small urban area	1	0.15	0.15	*	0.15	*	0.15	*
Not urban	103	0.10	<0.07	0.05	0.07	0.10	0.62	4

Table 30: Summary statistics of topsoil Hg (mg/kg) categorised by urbanisation extent

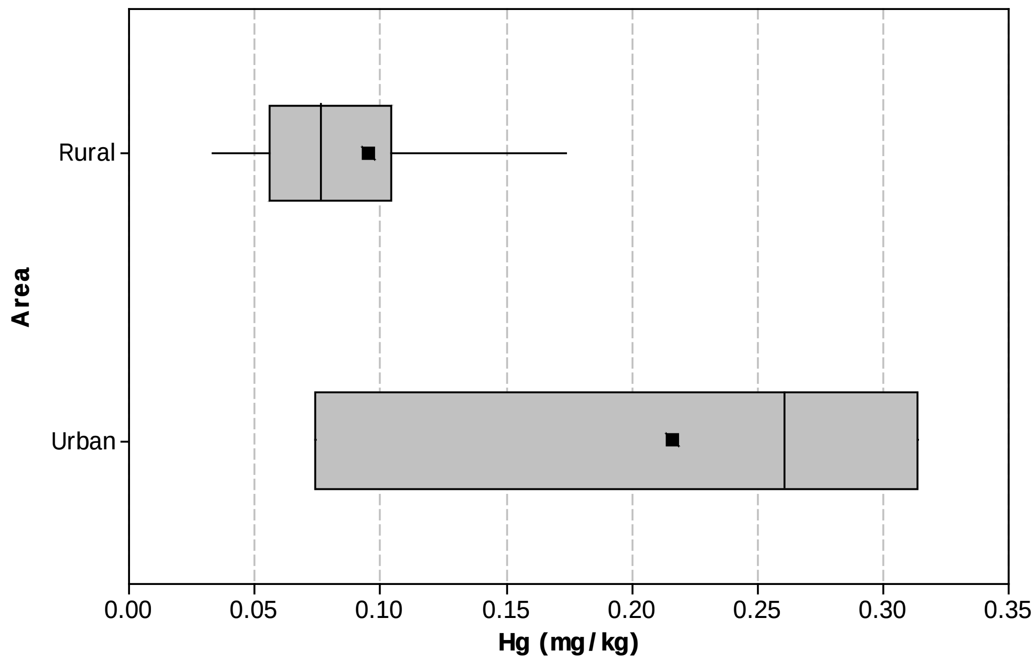


Figure 78: Boxplot of topsoil Hg data categorised by source project classification

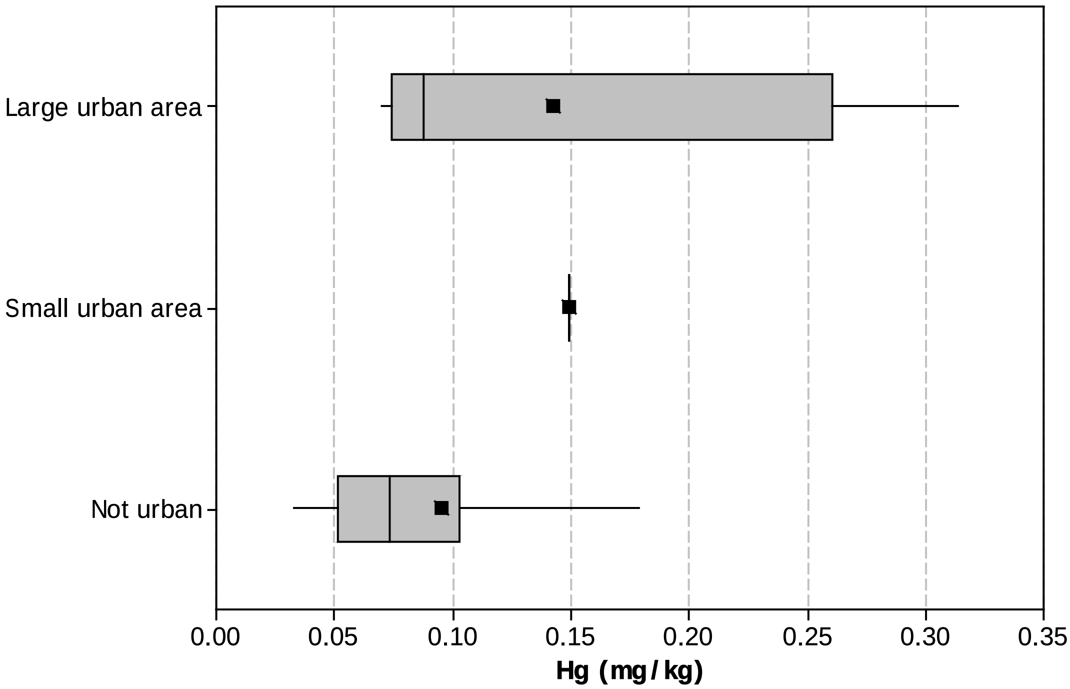


Figure 79: Boxplot of topsoil Hg data categorised by extent of urbanisation

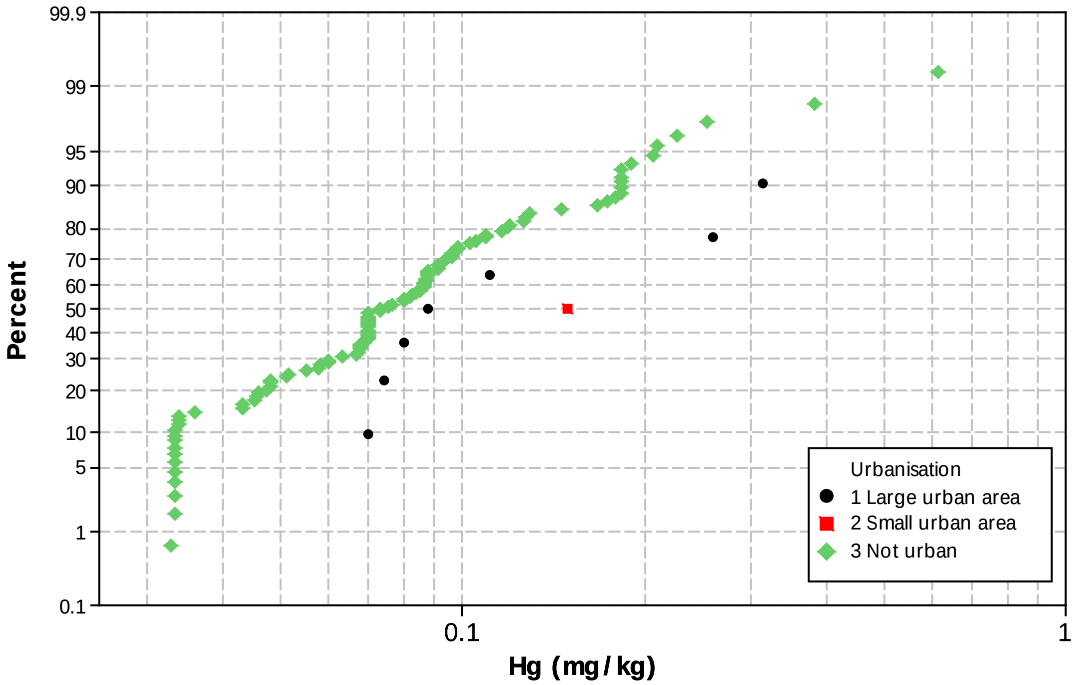


Figure 80: Probability plot of topsoil Hg categorised by extent of urbanisation

6.2.2 Non-ferrous metalliferous mineralisation and mining

A positive relationship between topsoil Hg concentrations and metalliferous mineralisation may be expected from an understanding of ore-field mineral chemistry and from previous work (Davies, 1976; Thornton, 1991), although Davies does note that the relative increase of Hg concentration he observed (maximum of nineteen times the background) is low in relation to the relative increase seen in ore-forming elements (*e.g.* Pb, As, Cu) or major accessory elements (*e.g*. Cd).

Since there are only four samples which intersect the mineralisation/mining occurrence data set (Table 31), it is impossible to discern whether this should form a separate domain for Hg in topsoils. It is unlikely, from the mineral assemblages found, that all of the Welsh ore-fields will have higher Hg concentrations. However, there are records of higher soil Hg in the Ystwyth valley (range 0.08 - 1.8 mg/kg) (Davies, 1976).

Therefore, no evidence within this data set is available to suggest that mining areas form a separate domain. Further Hg data is required to test whether ore-field areas in Wales are associated with elevated levels of Hg.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Metalliferous	4	0.05	<0.07	<0.07	<0.07	0.07	0.08	0.7
Non-metalliferous	107	0.10	<0.07	<0.07	0.08	0.11	0.62	3

Table 31: Summary statistics of topsoil Hg (mg/kg) classified by metalliferous mineralisation/mining areas

6.3 DOMAIN DATA SUMMARY

It can be seen that the Urban Domain has a higher typical concentration than the residual sample data (Figure 81, Figure 82 and Table 32), which form the Principal Domain data set – although there are too few samples to calculate an NBC. The location of these domains is shown in Figure 83.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Urban	7	0.14	<0.07	0.07	0.09	0.26	0.31	1
Principal	104	0.10	<0.07	<0.07	0.07	0.10	0.62	4

Table 32: Summary statistics of topsoil Hg (mg/kg) in the domains

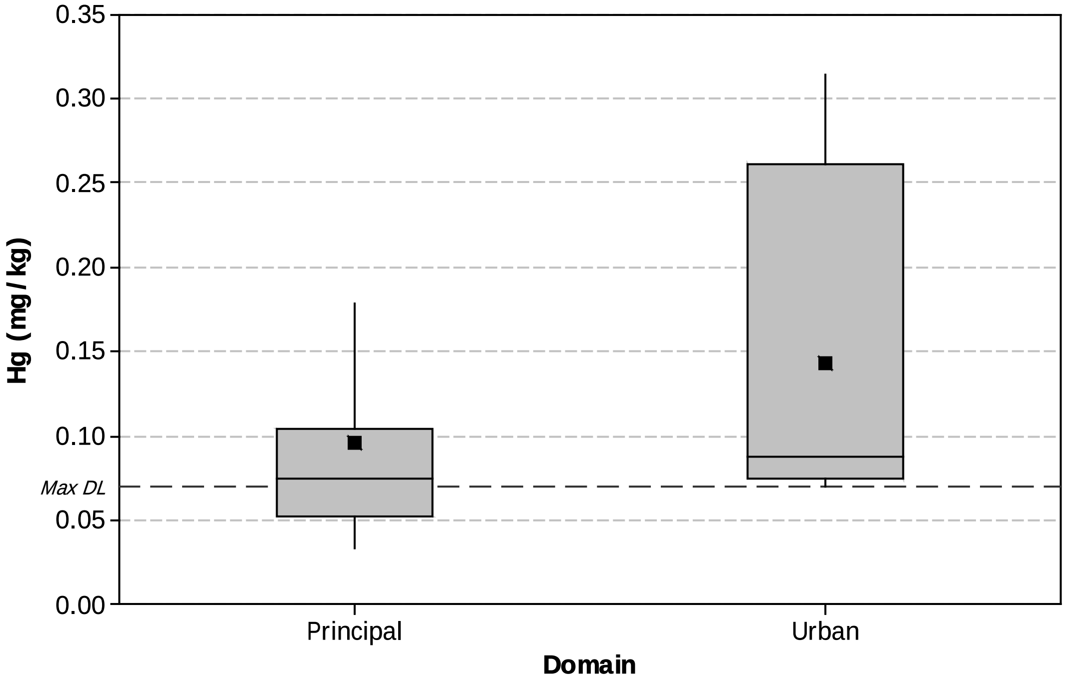


Figure 81: Boxplot of Hg topsoil data classified by domains

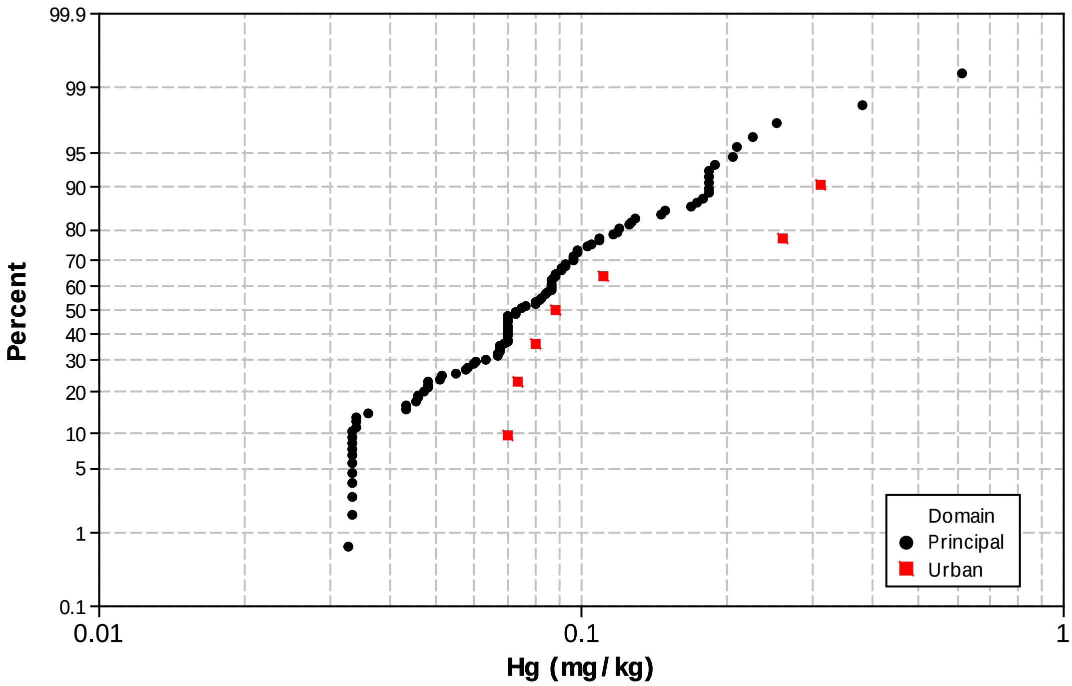
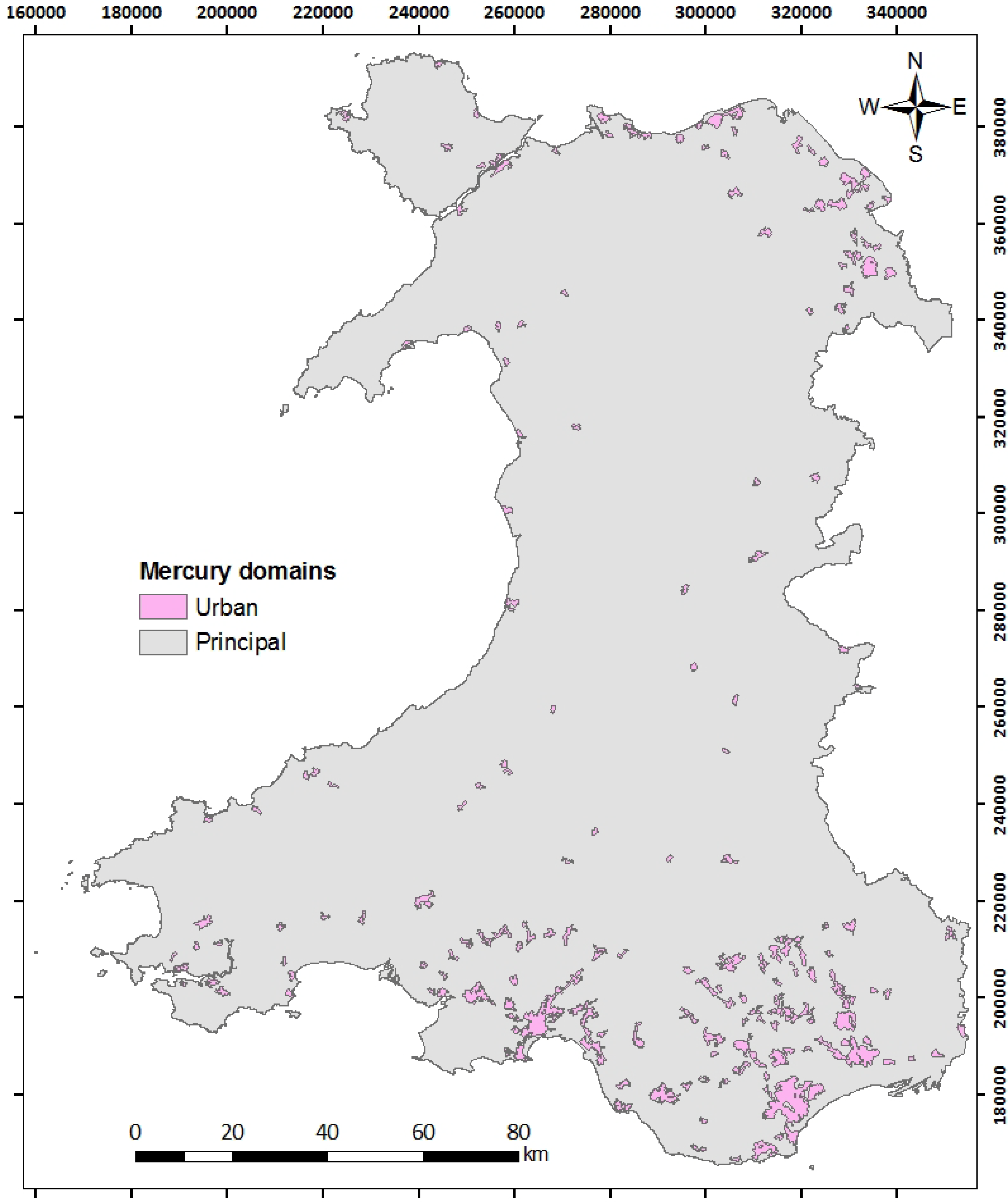


Figure 82: Probability plot of Hg topsoil data classified by domains



Contains Ordnance Survey data © Crown copyright and database rights (2012)
Urban areas defined from OS Strategi® used under OS OpenData™ terms

Figure 83: Map showing the location of Hg domains

6.4 Hg NBC CALCULATION

The NBC calculations for Hg have been carried out according to the method described in Cave *et al.* (2012).

6.4.1 Principal Domain

Untransformed skewness coefficient and octile skew fail the test criteria for underlying Gaussian distribution. When $\log_e$ transformed was used, the skewness coefficient was <1 and octile skew <0.2, indicating an underlying Gaussian distribution (Figure 84 and Figure 85) for which parametric percentiles were used (Table 33) to calculate an example NBC as shown in (Figure 86).

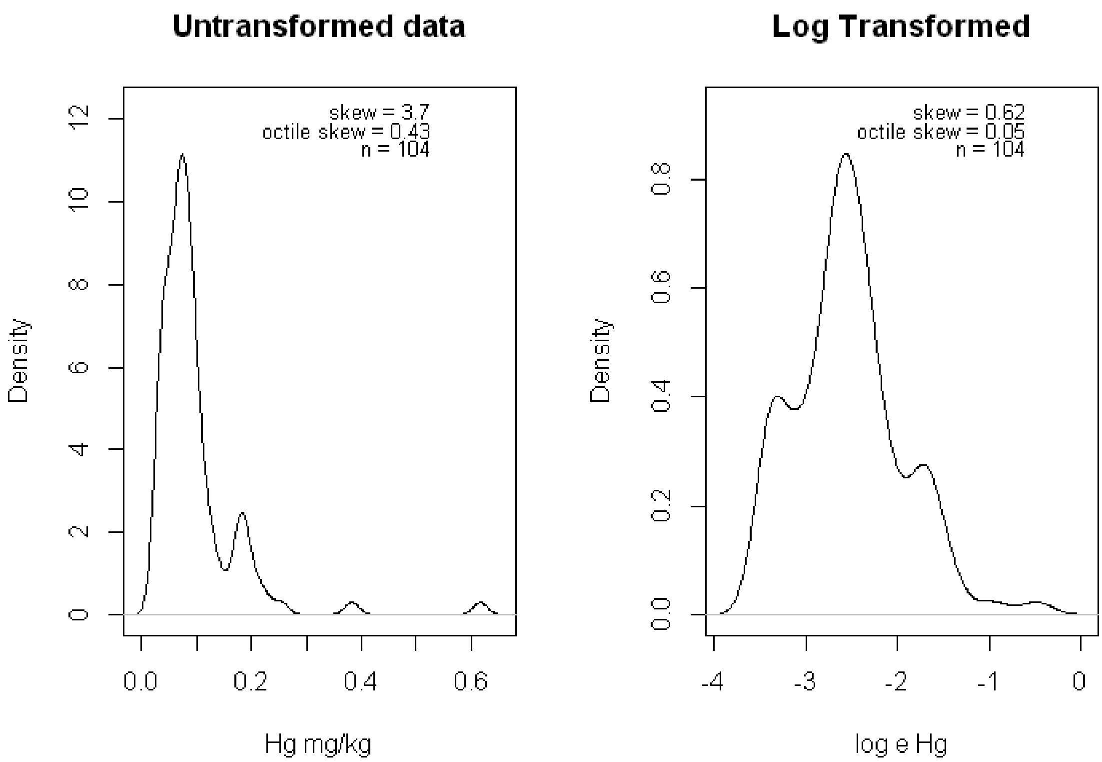


Figure 84: Density distributions for the raw data and the $\log_e$ transformed data for Hg in the Principal Domain (n = number of samples)

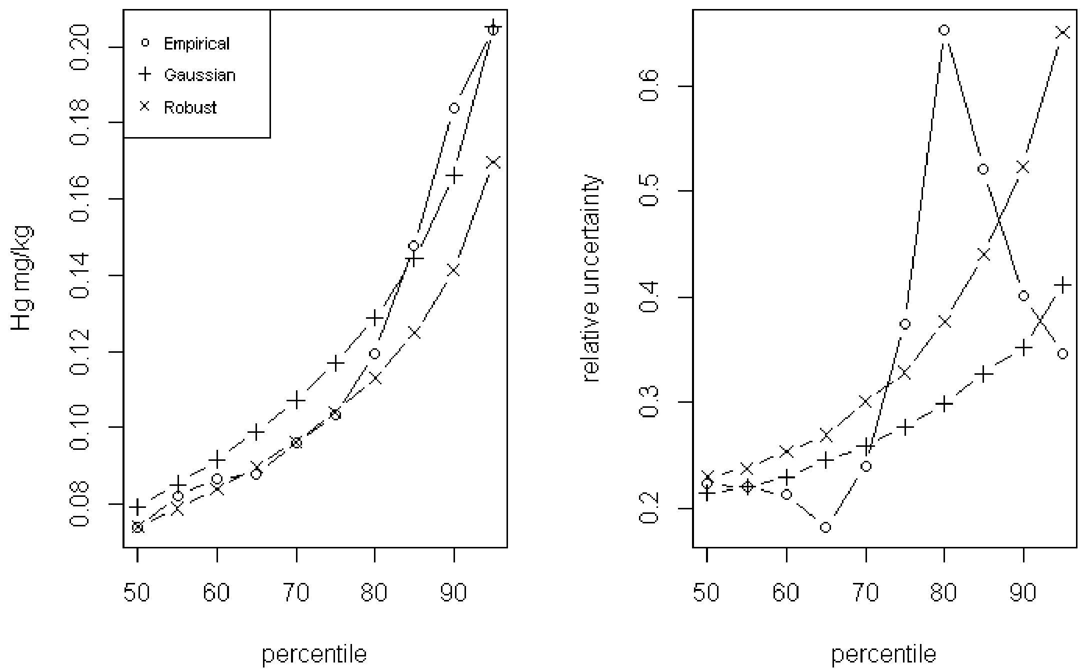


Figure 85: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Hg in the Principal Domain

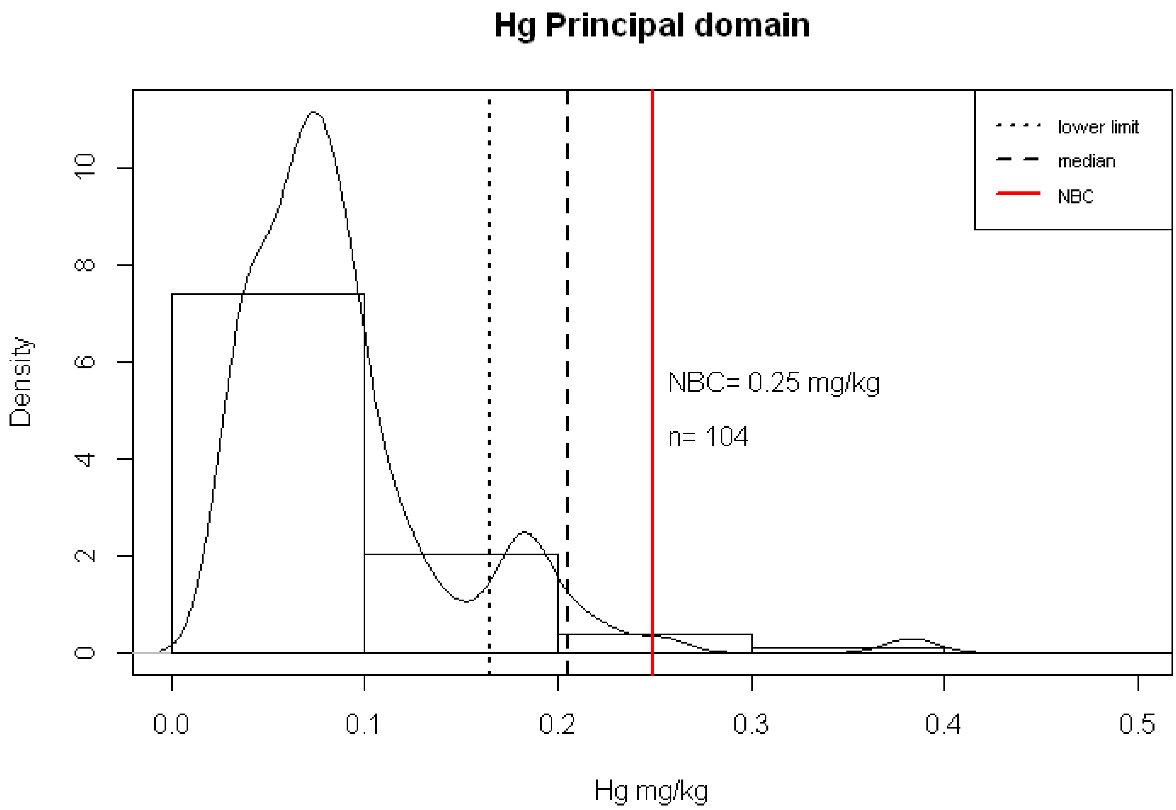


Figure 86: Summary density plot and histogram of the distribution for Hg in the Principal Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	0.07	0.07	0.09	0.07	0.08	0.09	0.07	0.07	0.09
55	0.07	0.08	0.09	0.08	0.09	0.10	0.07	0.08	0.09
60	0.07	0.09	0.09	0.08	0.09	0.10	0.08	0.08	0.10
65	0.08	0.09	0.10	0.09	0.10	0.11	0.08	0.09	0.11
70	0.09	0.10	0.12	0.09	0.11	0.12	0.09	0.10	0.12
75	0.09	0.10	0.13	0.10	0.12	0.13	0.09	0.10	0.13
80	0.10	0.12	0.17	0.11	0.13	0.15	0.10	0.11	0.14
85	0.11	0.15	0.18	0.12	0.14	0.17	0.10	0.12	0.16
90	0.13	0.18	0.20	0.14	0.17	0.20	0.11	0.14	0.19
95	0.18	0.20	0.25	0.16	0.21	**0.25**	0.13	0.17	0.23

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 33: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Hg (mg/kg) in the Principal Domain

6.4.2 Urban Domain

The Hg NBC for the Urban Domain cannot be calculated, since only seven sample sites are available. Until such time as sufficient data are collected to permit a NBC calculation, the English NBC may be used as a guide (Ander *et al.*, 2012).

6.4.3 Mercury NBC summary

Summary of NBC calculations for Hg and the associated information is shown in Table 34.

Percentile	Principal Domain n = 104 Area = 20,600 km^2 Transformation: $\log_e$			Urban Domain n = 7 Area = 600 km^2 Transformation: n/a		
	lower	middle	upper	lower	middle	upper
50	0.07	0.08	0.09	nc	nc	nc
55	0.08	0.09	0.10	nc	nc	nc
60	0.08	0.09	0.10	nc	nc	nc
65	0.09	0.10	0.11	nc	nc	nc
70	0.09	0.11	0.12	nc	nc	nc
75	0.10	0.12	0.13	nc	nc	nc
80	0.11	0.13	0.15	nc	nc	nc
85	0.12	0.14	0.17	nc	nc	nc
90	0.14	0.17	0.20	nc	nc	nc
95	0.16	0.21	**0.25**	nc	nc	nc

n = number of samples used in calculation. nc - Not calculated.

Table 34: Mercury NBC summary table

7 Nickel (Ni)

7.1 EXPLORATORY DATA ANALYSIS

A systematic difference between the urban and rural data can be seen in Figure 87 and Table 35. However, Figure 88 shows that this is dominated by the much higher concentrations in the Swansea G-BASE data, than those in the Cardiff data set. It is also noticeable that Ni has a relatively restricted absolute range (less than three orders of magnitude) and very few high concentration outliers, even in the urban G-BASE data.

The data are mapped in Figure 89, and although the higher concentrations associated with the Swansea and surrounding area are visible, the only other strong spatial controls which are apparent are those of systematically low concentrations around Snowdonia. When k-means cluster analysis is applied to these data, the breaks are identified at more evenly distributed concentrations than observed for the other contaminants explored in this project, and therefore encapsulate much of Wales (Figure 90). When only the NSI(XRFS) data are shown (to avoid the distortion from the Swansea higher concentrations), only the lowest 26% of the data are excluded (Figure 91). Since the objective here is to identify those much higher concentrations which require separating into different domains, in this instance only the upper clusters have been used – as shown in Figure 92. This shows the Swansea area is the only one to be systematically captured in the higher concentration thresholds. A survey of part of the Swansea valley found a median soil Ni concentration of 9 mg/kg (using a partial extraction method; 70 samples) (Davies, 1997).

A background survey of north west Pembrokeshire found a mean concentration of 29 mg/kg (260 samples) (Bradley *et al.*, 1978) and a study of soils in England and Wales reported a median of 24 mg/kg (1521 samples) (Archer and Hodgson, 1987), which are both reasonably close to the NSI(XRFS) median of 24 mg/kg.

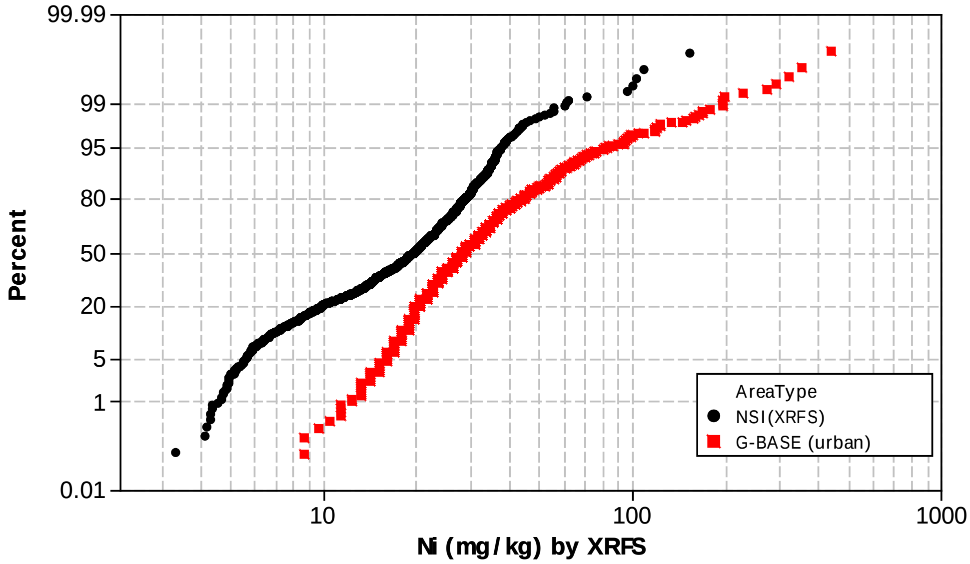
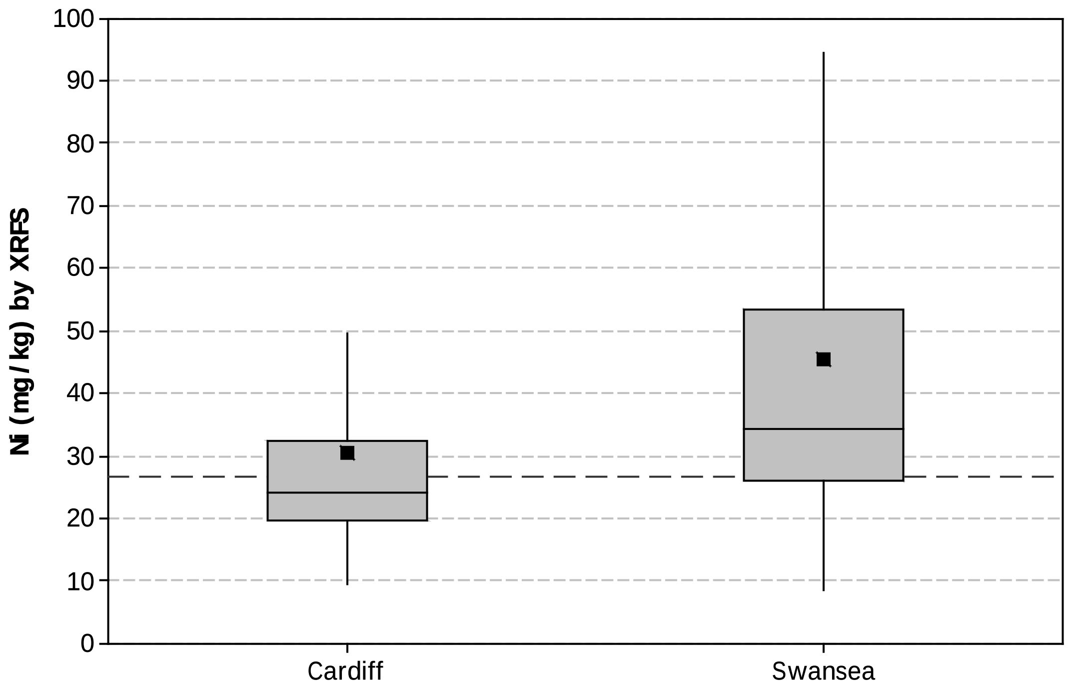


Figure 87: Cumulative probability plot of topsoil Ni concentrations for G-BASE urban and NSI(XRFS) results



The horizontal black dashed line is at the 75th percentile of the NSI(XRFS) data set

Figure 88: Boxplot of urban topsoil Ni data collected by the G-BASE project. Dashed lines extend the interquartile range of the national NSI(XRFS) data

(a) Ni (mg/kg)	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
All data	1675	29.1	3.31	17.6	24.1	33.2	437	6

(b) Ni (mg/kg) Area type	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
NSI(XRFS)	798	20.6	3.31	11.7	19.5	26.8	152	3
Urban	877	36.9	8.64	21.4	27.8	38.8	437	6

(c) Ni (mg/kg) Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Cardiff	507	30.6	9.55	19.6	24.2	32.4	437	9
Swansea	370	45.7	8.64	26.0	34.2	53.5	320	4

Table 35: Statistical summary of topsoil Ni in the main data sets (a) whole data set (b) by sampling density (c) by urban centres sampled by G-BASE

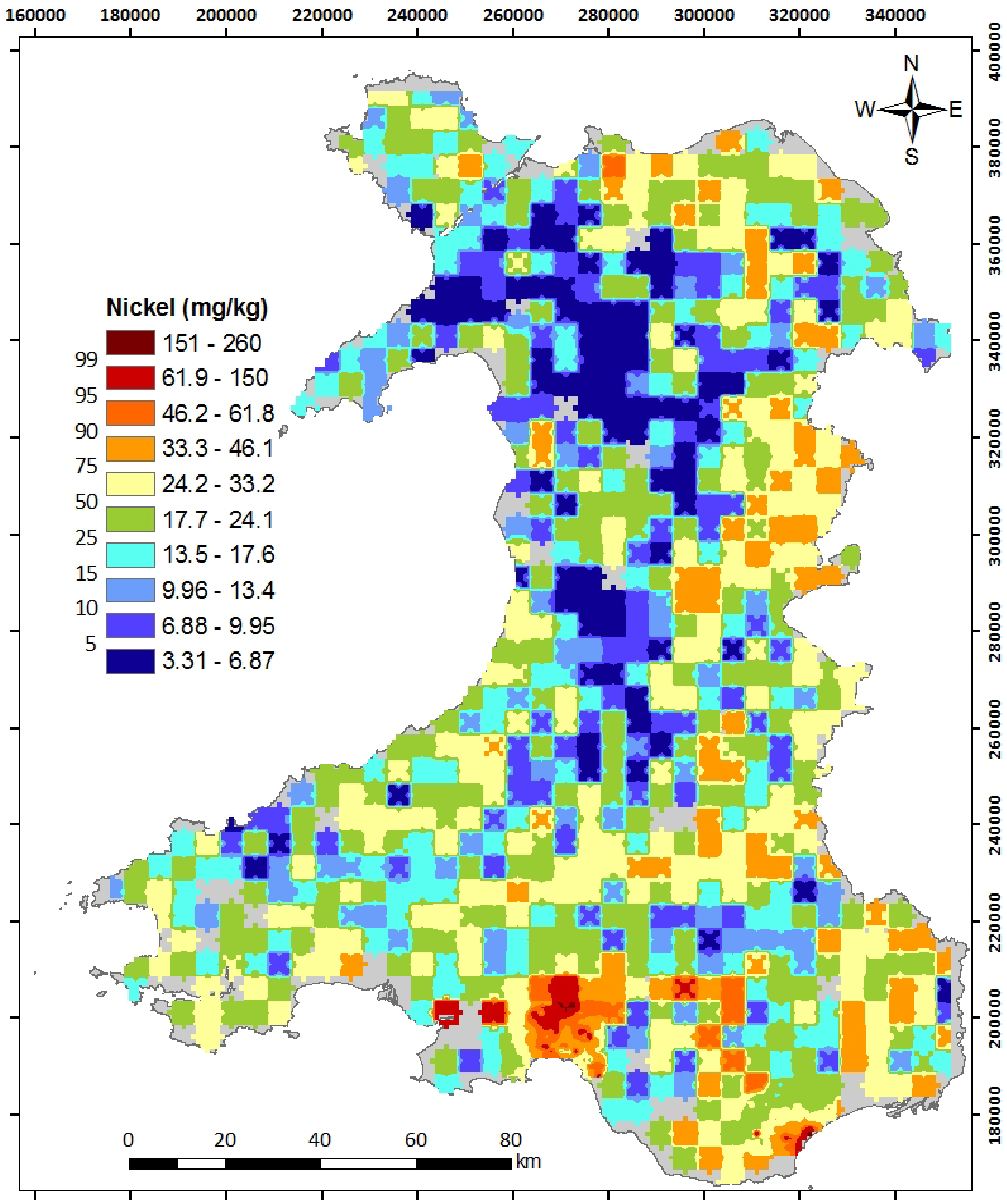


Figure 89: Interpolated map of topsoil Ni data. Colour thresholds are designed for highly skewed data

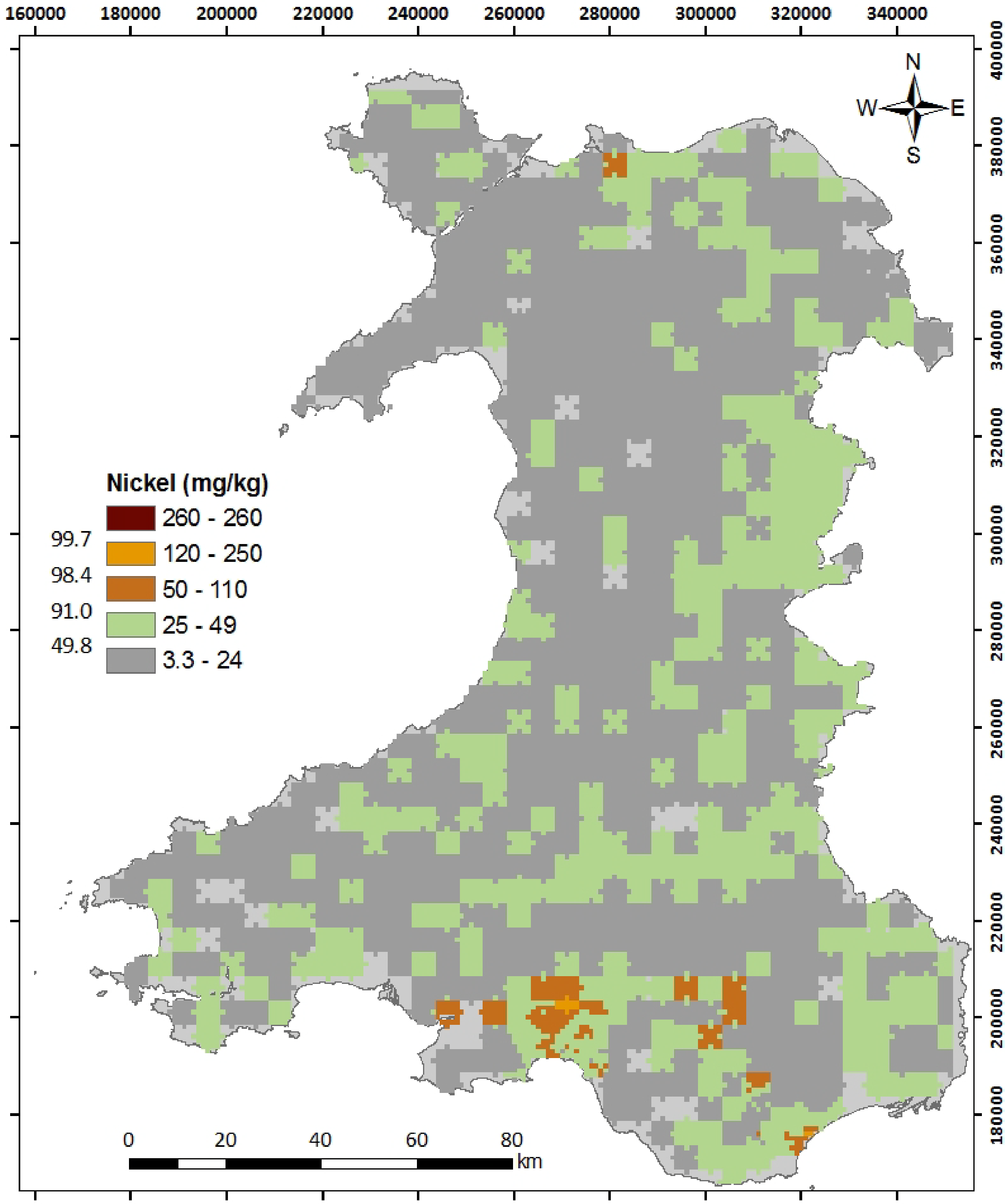


Figure 90: Interpolated map of topsoil Ni concentrations. Thresholds determined by using k-means cluster analysis

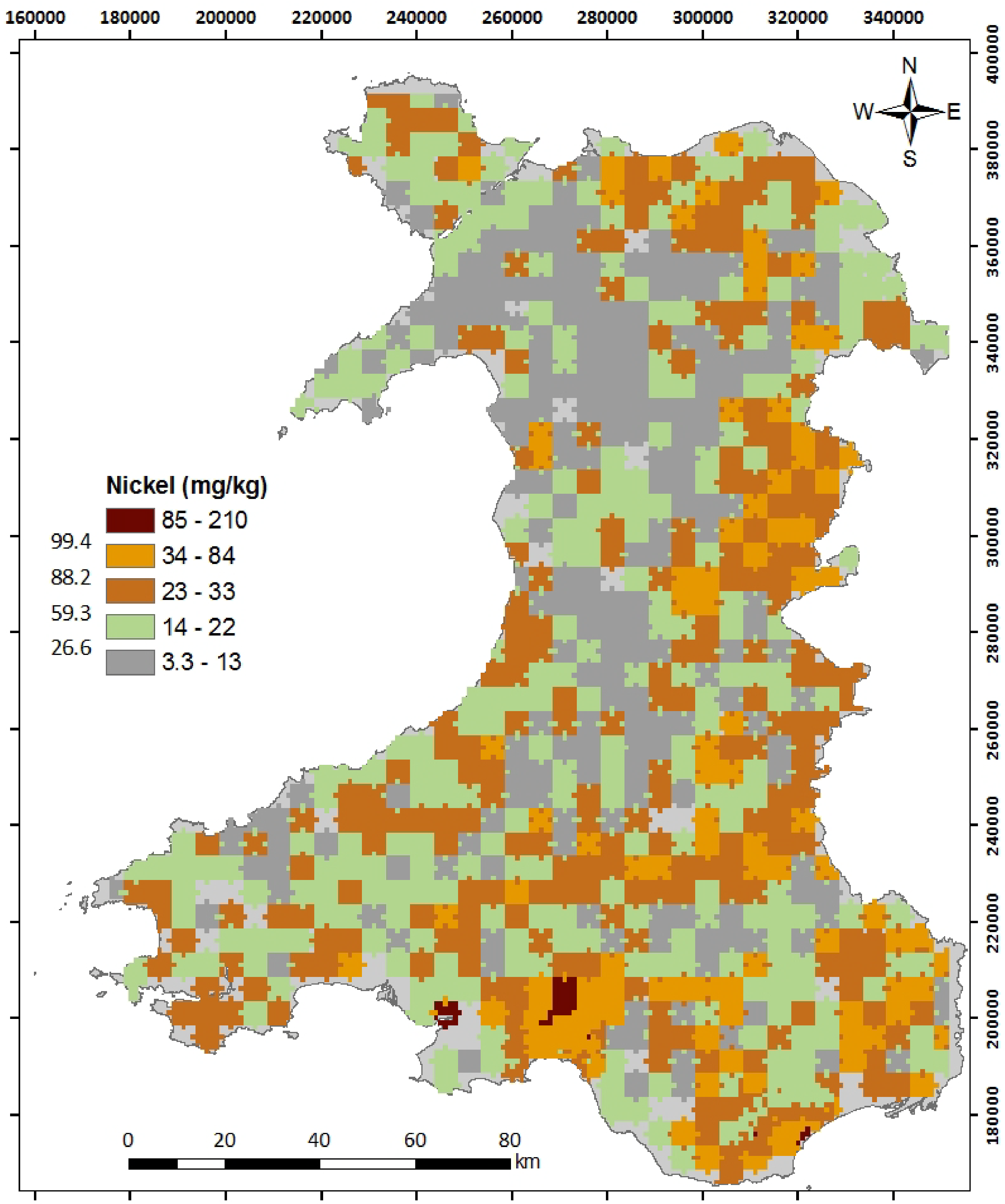
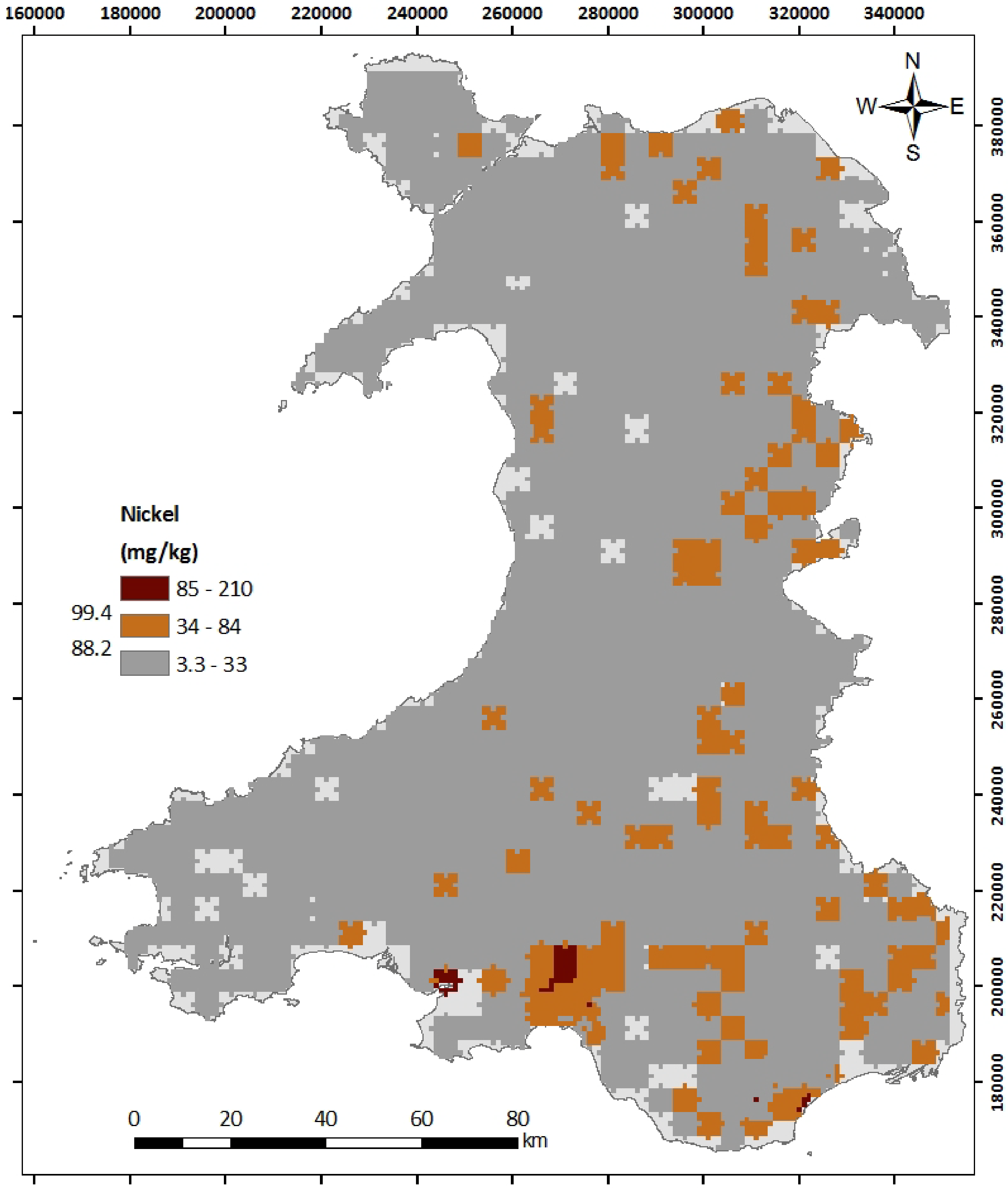


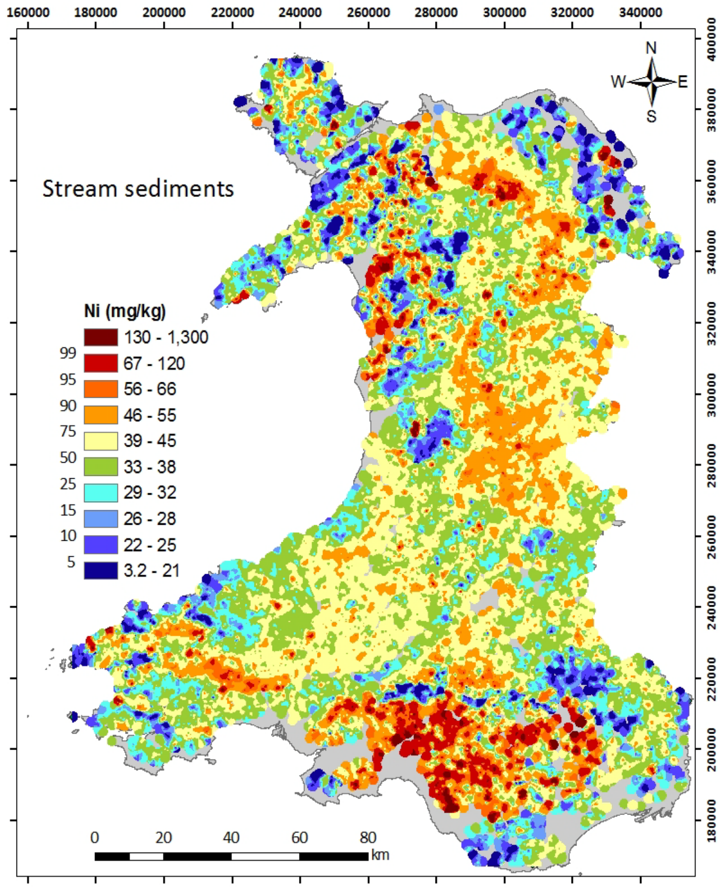
Figure 91: Interpolated map of topsoil Ni concentrations from the NSI(XRFS) data set only. Thresholds determined by using k-means cluster analysis



Thresholds determined by using k-means cluster analysis, with only the upper percentile breaks being applied

Figure 92: Interpolated map of topsoil Ni concentrations from the NSI(XRFS) data set only

(a)



(b)

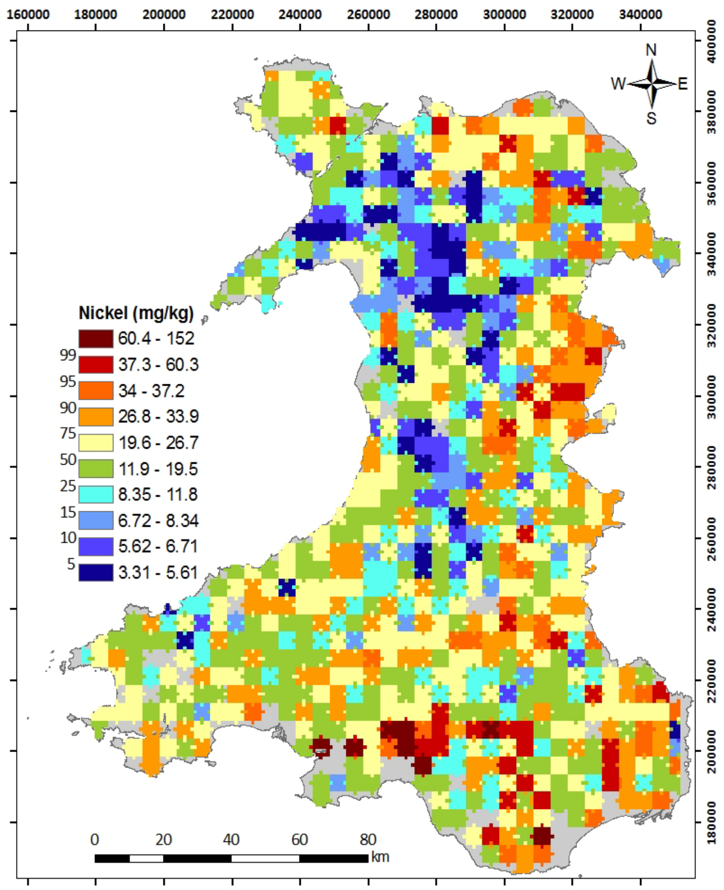


Figure 93: Nickel in (a) G-BASE stream sediments and (b) NSI(XRFS) soils

7.2 DOMAIN SELECTION

The information shown below was used to compile the data for domains, with higher topsoil Ni concentrations than typical, in:

- urban/industrial activities in the South Wales Coalfield; and
- basic/ultrabasic rocks.

Whilst in Figure 93 there is some evidence of other higher concentrations in central-east mid-Wales, and in NE Wales, these do not translate into concentrations >40 mg/kg, so have not been investigated further in this study. Any further data becoming available should be used to establish whether there is evidence for a higher topsoil Ni concentration domain in those areas.

7.2.1 Urban areas

The soils data from G-BASE appears to be generally enhanced in concentration when compared to the NSI(XRFS) data (Figure 87), particularly for the G-BASE Swansea data (Figure 88). When the data are intersected with the urbanisation spatial data set, there is a small positive relationship with the size of the urban centre (Figure 94).

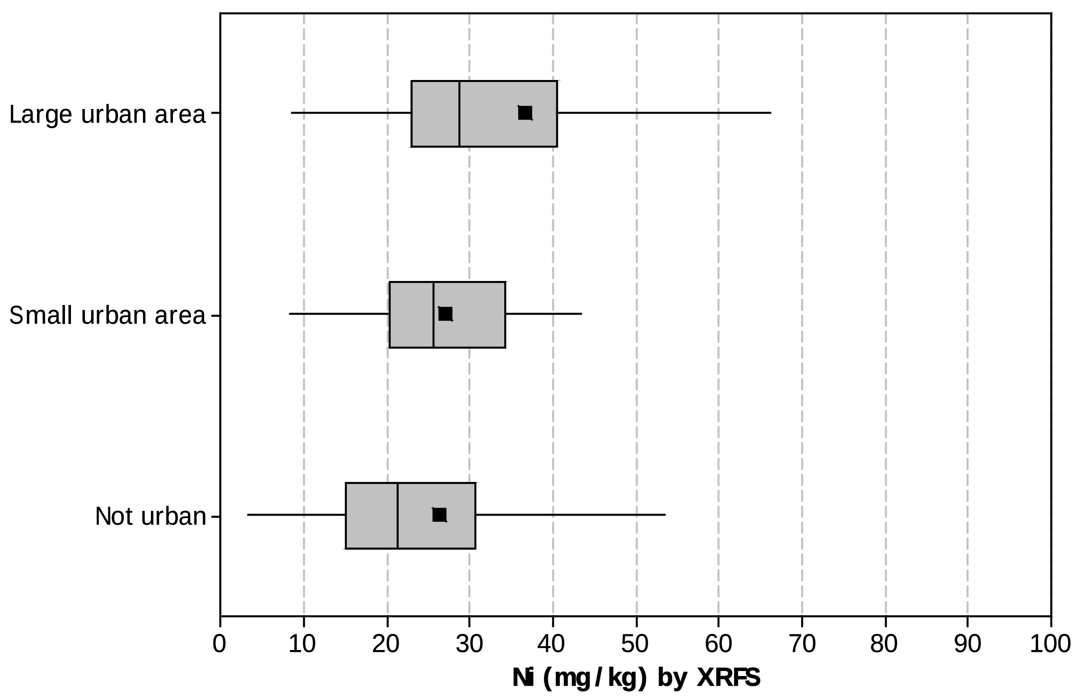
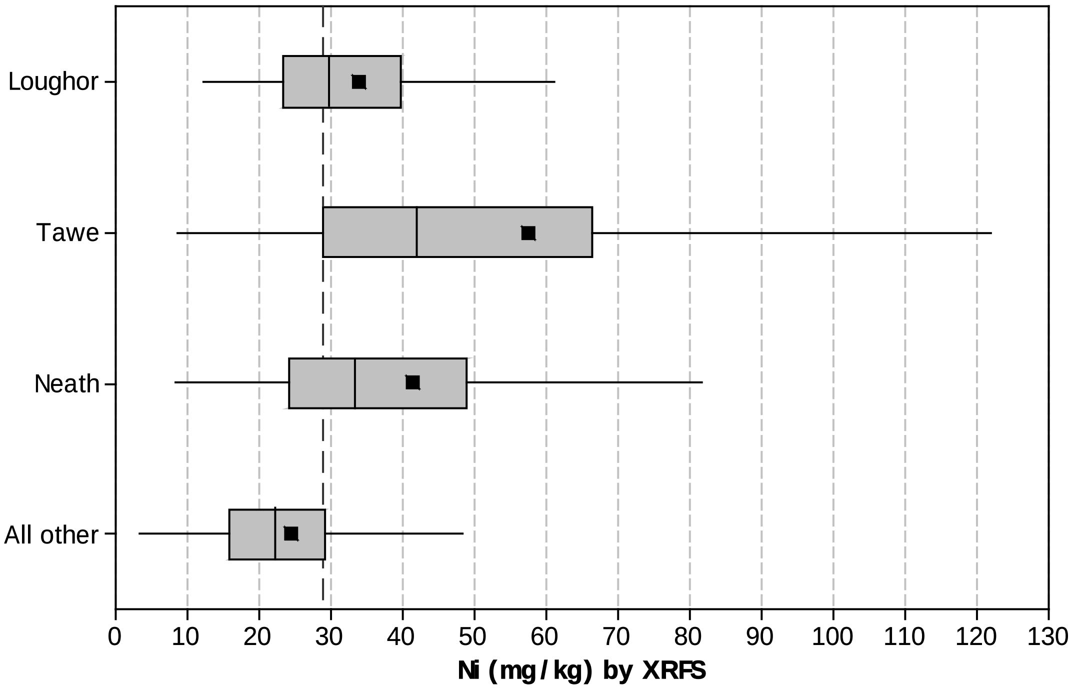


Figure 94: Boxplot of topsoil Ni data classified by the extent of urbanisation

The enhanced Ni concentrations in stream sediments clearly visible across the South Wales Coalfield (Figure 93a) were attributed to natural sources, probably augmented by some diffuse pollution (British Geological Survey, 2000). However, there is no evidence of such systematic elevation in Ni concentration in the soils from the NSI(XRFS) data set (Figure 93b). However, in the west of the coalfield area there do appear to be higher concentrations in the region of the Swansea valley (Tawe river) in Figure 93b, and generally higher concentrations in the G-BASE Swansea data set (Table 35). These observations are confirmed when the entire data set is intersected with the three valleys also inspected in the As data set (Section 3.2.1), as shown in Figure 95 and Table 36.

There are other isolated soil samples with Ni concentrations in the range 40 – 60 mg/kg in the South Wales Coalfield; any further data that later becomes available may help to establish whether the higher concentration domain that is suggested by the G-BASE stream sediment data should be extended further.



Vertical bold dashed line represents the 75th percentile of 'All other' data

Figure 95: Boxplot of topsoil Ni data classified by selected valleys in south Wales

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Loughor	40	33.8	12.1	23.5	29.7	39.7	71.5	1
Tawe	138	57.5	8.7	28.8	42.0	66.5	320	3
Neath	164	41.3	8.4	24.2	33.3	48.9	272	4
All other	1333	24.6	3.3	16.0	22.3	29.1	437	10

Table 36: Summary statistics of topsoil Ni (mg/kg) over selected valleys in south Wales

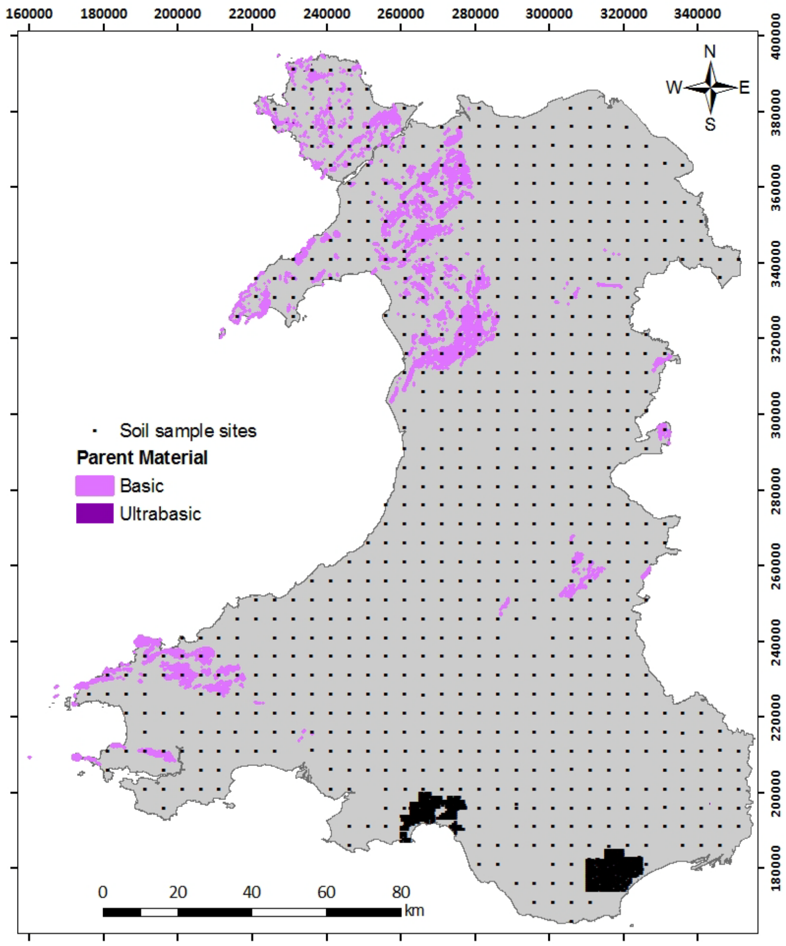
7.2.2 Basic and ultrabasic rocks

Basic (mafic) rocks have Ni-bearing minerals within their matrix – the proportion of this is further increased in ultrabasic (ultramafic) rocks. Elevated concentrations seen in other areas/studies of soils and sediments, including the Welsh stream sediment data (Figure 93a) (British Geological Survey, 2000) and the equivalent study of NBCs in England (Ander *et al.*, 2012), where other UK studies are also reviewed.

Notwithstanding this expectation of higher topsoil Ni concentrations over basic and ultrabasic rocks, it is not the case in these data. This is shown in Table 37 for the six samples which were collected over basic rocks. The locations of basic and ultrabasic rocks are shown in Figure 96. The ultrabasic rocks are very difficult to see on this map, having small scattered outcrops in Anglesey and a total outcrop area <1 km^2, whilst the basic rocks have an outcrop area of 157 km^2 (~1% of land area). Therefore, it is not clear to what extent the ultrabasic parent material should have a separate domain from that of basic soil parent material, and this has not been attempted here. It should be borne in mind, therefore, that soils formed over the ultrabasic rocks are likely to be considered a separate domain as more data becomes available, and it would be expected that the topsoil Ni concentration found over the basic rocks may tend to increase with as more data becomes available from further sampling and analysis.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Basic	6	17	4.9	6.2	21.5	23.3	26.0	-0.8
Ultrabasic	0	-	-	-	-	-	-	-
All other	1669	29	3.3	17.6	24.2	33.3	437	6

Table 37: Summary statistics of topsoil Ni (mg/kg) over basic/ultrabasic rocks



Geological data from the BGS SPMM (© NERC, all rights reserved)

Figure 96: Map of the distribution of basic and ultrabasic rocks in Wales, with soil sample sites. The small area of ultrabasic rocks occurs in Anglesey. The dark black areas in south Wales are the high density G-BASE sample sites in Swansea and Cardiff

7.3 DOMAIN DATA SUMMARY

The summary statistics for the domains are shown in Table 38, and it should be noted that there are <30 samples for the basic domain (see Cave *et al.*, 2012) and that there is no evidence in these data for higher Ni concentrations. However, there is strong evidence from studies elsewhere that basic rocks will generally yield higher Ni concentrations. The Urban 1 Domain reflects the three selected valleys in south Wales shown in Section 7.2.1, which gives a domain that appears to have a characteristically different data population to that of the main data sets, as shown in Figure 97 and Figure 98. The spatial occurrence of the domains is shown in Figure 99.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Basic	6	17	4.9	6.2	21.5	23.3	26.0	-1
Urban 1	342	47	8.4	26.0	35.2	54.4	320	4
Principal	1327	25	3.3	16.0	22.4	29.4	437	10

Table 38: Summary statistics of topsoil Ni (mg/kg) in the domains

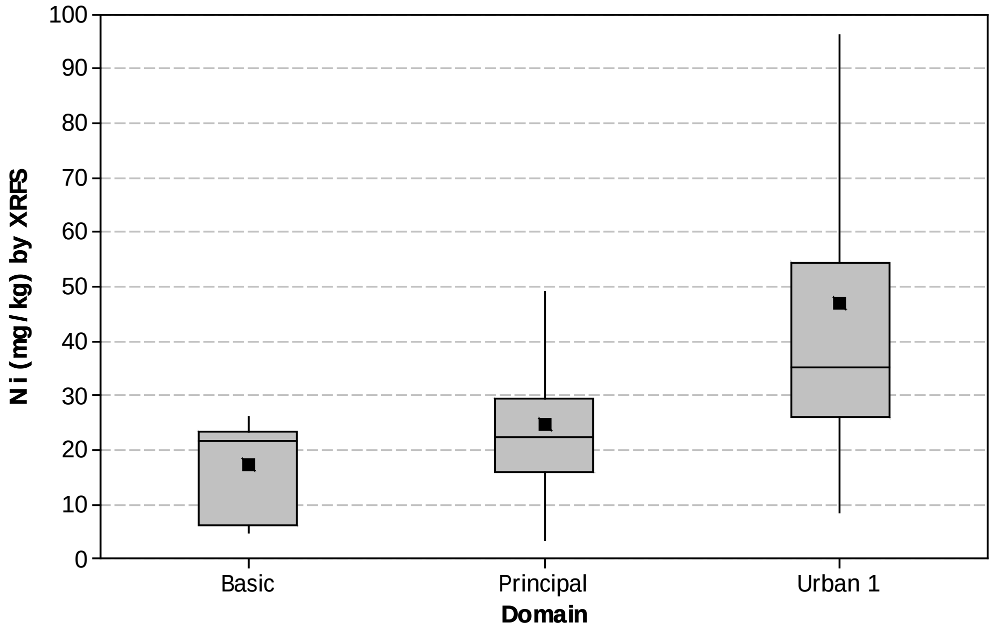


Figure 97: Boxplot of topsoil Ni data classified by domains

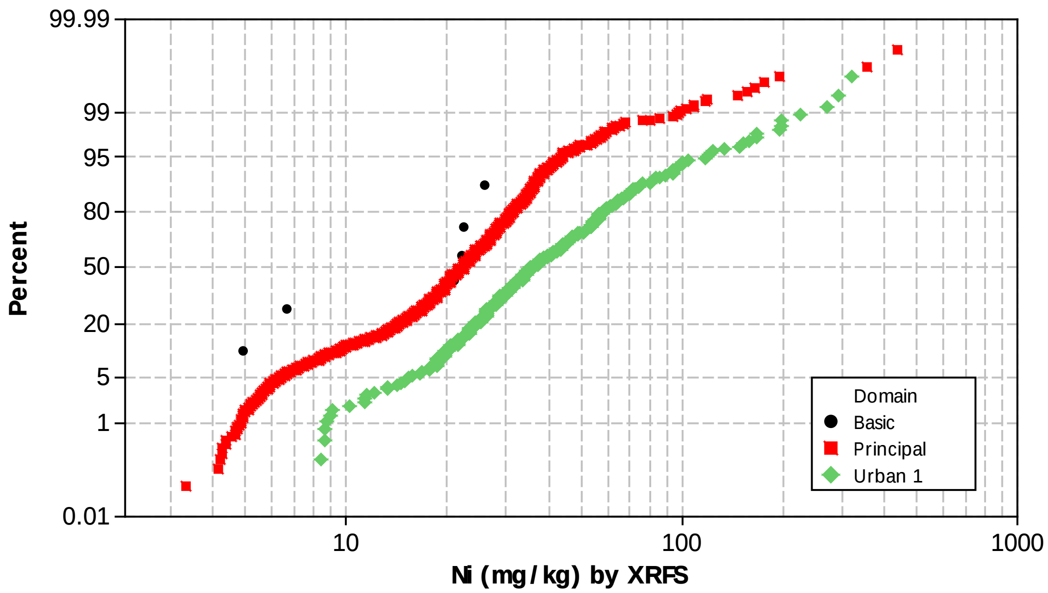
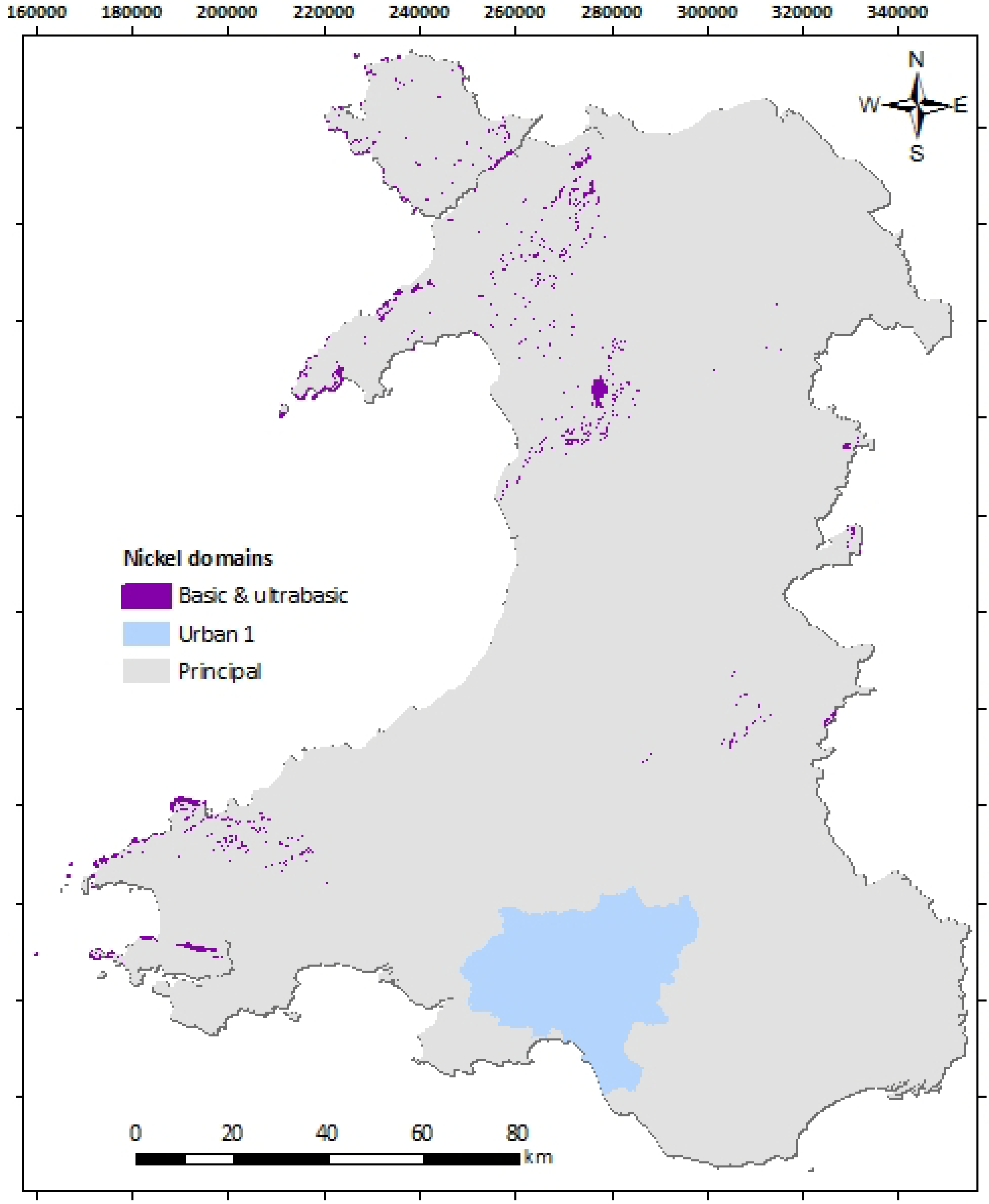


Figure 98: Probability plot of topsoil Ni data classified by domains



Contains Ordnance Survey data © Crown copyright and database rights (2012)
Geological data from the BGS SPMM (© NERC, all rights reserved)

Figure 99: Map showing the location of Ni domains

7.4 Ni NBC CALCULATION

The NBC calculations for Ni have been carried out according to the method described in Cave *et al.* (2012).

7.4.1 Principal Domain

Untransformed data have a skewness coefficient >1, but an octile skew <0.2 ,which indicates underlying Gaussian distribution with outliers (Figure 100). Therefore, untransformed data with robust percentiles were used to calculate the NBC (Figure 101, Figure 102 and Table 39).

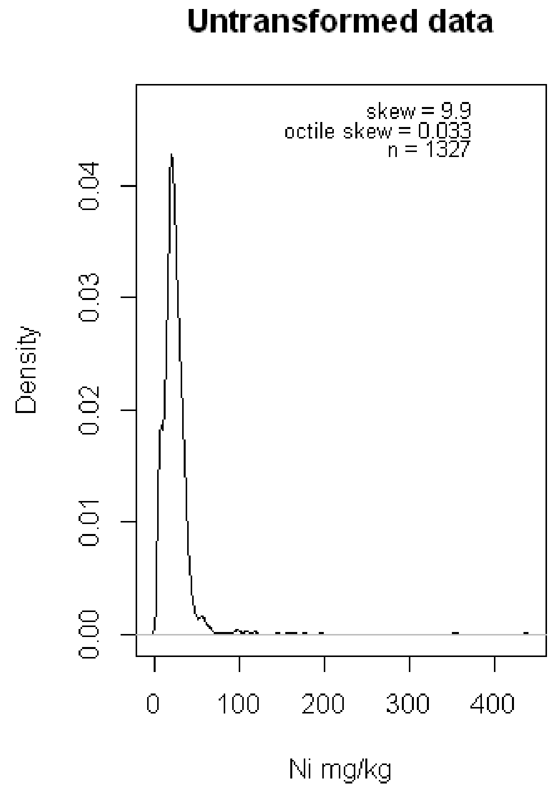


Figure 100: Density distributions for the untransformed data for Ni in the Principal Domain (n = number of samples)

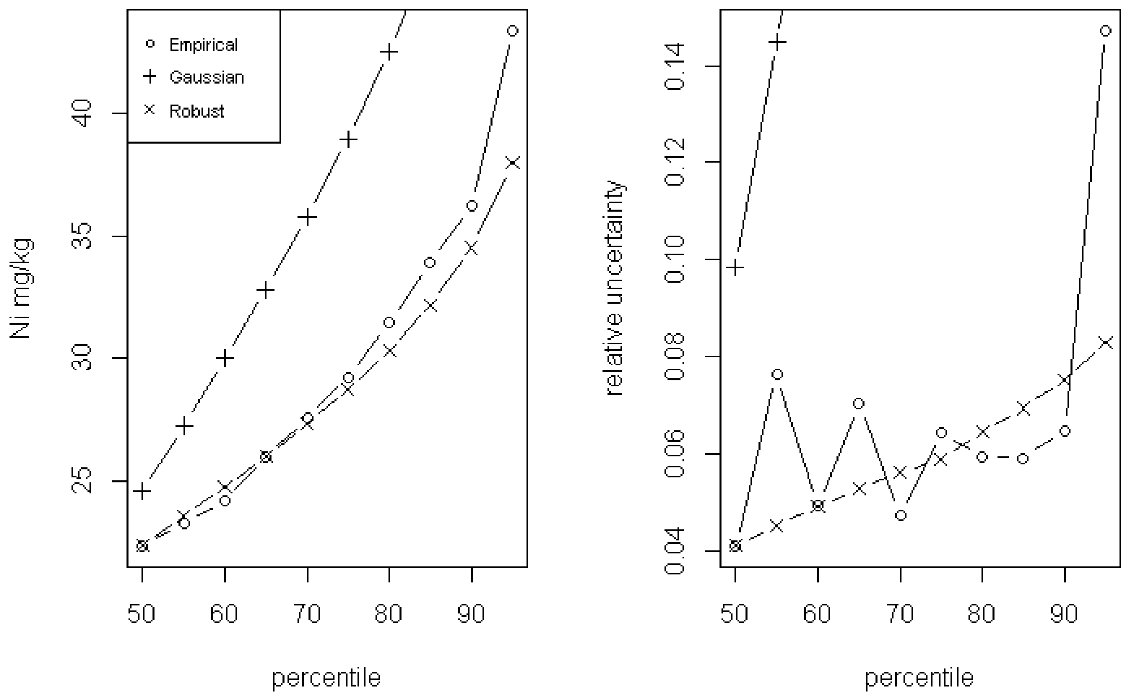


Figure 101: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Ni in the Principal Domain

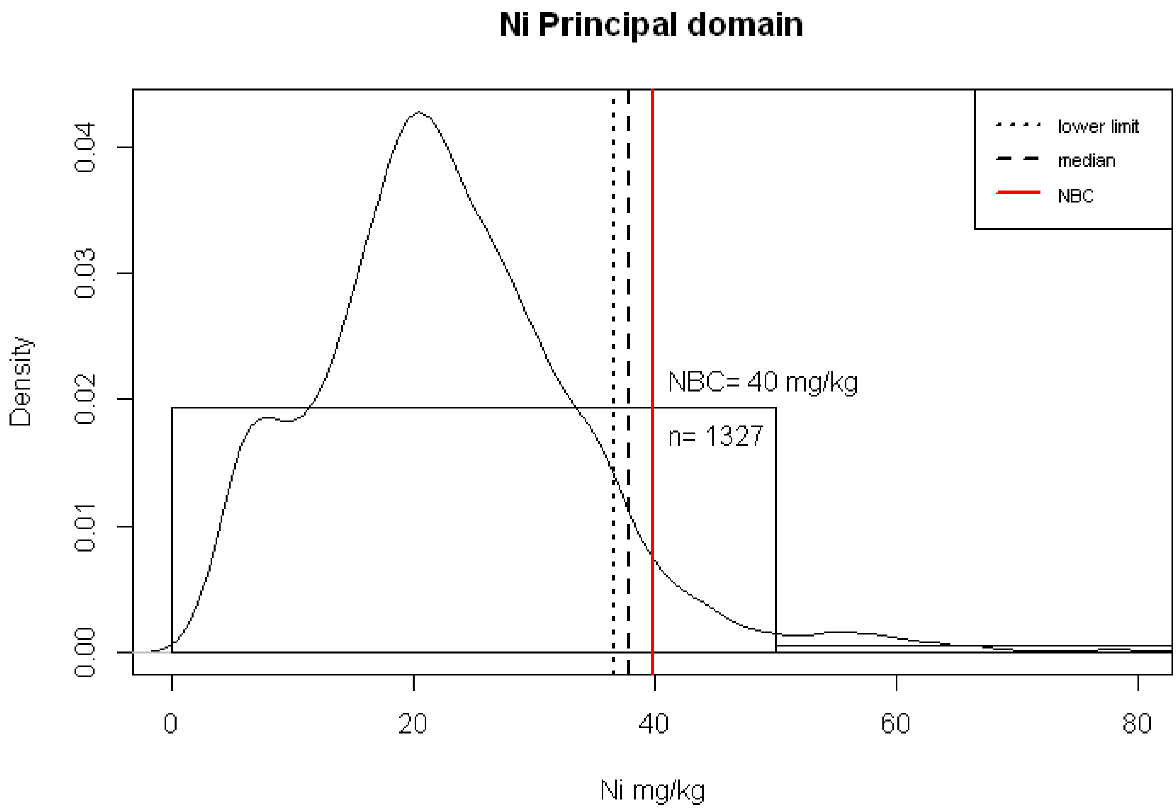


Figure 102: Summary density plot and histogram of the distribution for Ni in the Principal Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	21.4	22.4	22.4	23.5	24.6	25.9	**21.4**	**22.4**	**22.4**
55	22.4	23.3	24.0	25.3	27.3	29.3	**22.6**	**23.5**	**23.7**
60	23.8	24.2	25.1	27.1	30.0	32.9	**23.8**	**24.8**	**25.1**
65	25.1	26.0	26.9	29.0	32.8	36.6	**25.1**	**26.0**	**26.5**
70	26.9	27.6	28.3	30.9	35.8	40.5	**26.4**	**27.3**	**28.0**
75	28.4	29.3	30.3	33.0	39.0	44.8	**27.8**	**28.8**	**29.6**
80	30.5	31.5	32.4	35.3	42.5	49.5	**29.4**	**30.3**	**31.3**
85	32.6	33.9	34.9	38.0	46.7	55.0	**31.2**	**32.2**	**33.4**
90	35.6	36.2	37.9	41.5	51.9	61.9	**33.5**	**34.5**	**36.0**
95	40.6	43.4	46.8	46.5	59.7	72.1	**36.8**	**38.0**	**39.8**

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 39: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Ni (mg/kg) in the Principal Domain

7.4.2 Basic Domain

Because there are only six samples an NBC could not be calculated (Cave *et al.*, 2012). This may be calculated when more data becomes available.

7.4.3 Urban 1 Domain

Untransformed skewness coefficient and octile skew fail the test criteria for underlying Gaussian distribution. When $\log_e$ transformed was used, the skewness coefficient was <1 and octile skew <0.2, indicating an underlying Gaussian distribution (Figure 103 and Figure 104) for which parametric percentiles were used (Table 40) to calculate an example NBC as shown in Figure 105.

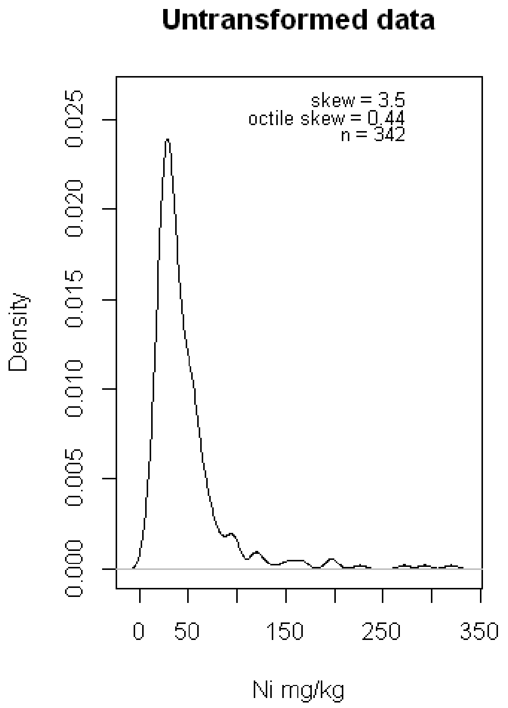


Figure 103: Density distributions for the raw data and the $\log_e$ transformed data for Ni in the Urban 1 Domain (n = number of samples)

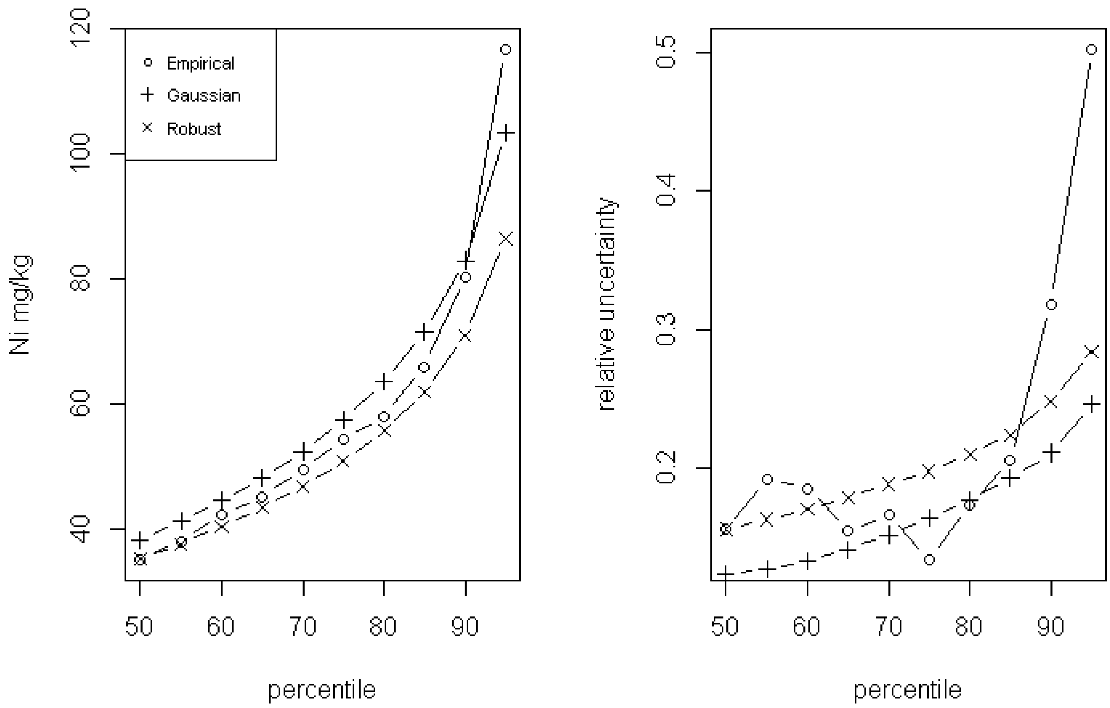


Figure 104: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Ni in the Urban 1 Domain

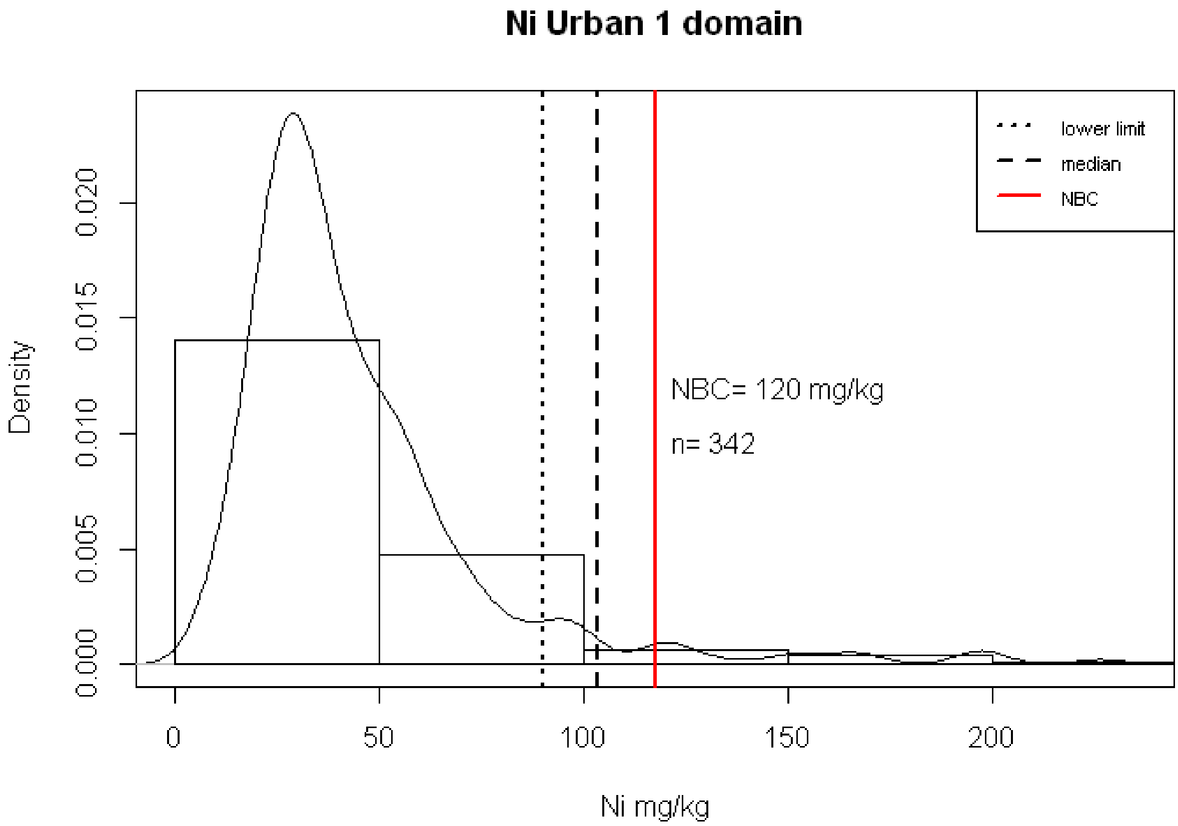


Figure 105: Summary density plot and histogram of the distribution for Ni in the Urban 1 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	33.3	35.2	38.2	35.5	38.2	40.9	33.3	35.2	38.8
55	35.2	38.1	42.5	38.2	41.2	44.3	35.4	37.7	41.5
60	37.9	42.3	45.2	41.3	44.5	47.9	37.7	40.4	44.4
65	42.2	45.2	49.5	44.6	48.2	52.0	40.1	43.4	47.6
70	45.2	49.5	54.4	48.3	52.4	56.8	42.9	46.8	51.4
75	48.9	54.4	57.8	52.5	57.4	62.5	46.1	50.8	55.9
80	54.4	58.0	64.2	57.8	63.6	69.6	50.0	55.7	61.3
85	58.9	66.0	71.7	64.3	71.5	78.9	54.9	62.0	68.5
90	69.7	80.4	94.6	73.6	83.0	92.5	61.8	70.9	78.4
95	93.7	117	152	90.0	103	**117**	73.0	86.5	97.6

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 40: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Ni (mg/kg) in the Urban 1 Domain

7.4.4 Nickel NBC summary

A summary of all Ni Domain NBCs and associated information is given in Table 41.

Percentile	Principal Domain n =1327 Area = 19,800 km^2 Transformation: none			Urban 1 Domain n = 342 Area = 1200 km^2 Transformation: $\log_e$			Basic Domain* n = 6 Area = 200 km^2 Transformation: n/a		
	lower	middle	upper	lower	middle	upper	lower	middle	upper
50	21.4	22.4	22.4	35.5	38.2	40.9	nc	nc	nc
55	22.6	23.5	23.7	38.2	41.2	44.3	nc	nc	nc
60	23.8	24.8	25.1	41.3	44.5	47.9	nc	nc	nc
65	25.1	26.0	26.5	44.6	48.2	52.0	nc	nc	nc
70	26.4	27.3	28.0	48.3	52.4	56.8	nc	nc	nc
75	27.8	28.8	29.6	52.5	57.4	62.5	nc	nc	nc
80	29.4	30.3	31.3	57.8	63.6	69.6	nc	nc	nc
85	31.2	32.2	33.4	64.3	71.5	78.9	nc	nc	nc
90	33.5	34.5	36.0	73.6	83.0	92.5	nc	nc	nc
95	36.8	38.0	**39.8**	90.0	103	**117**	nc	nc	nc

n = number of samples used in calculation. * - <30 samples. nc = not calculated.

Table 41: Nickel (mg/kg) NBC summary table

8 Lead (Pb)

8.1 EXPLORATORY DATA ANALYSIS

The NSI(XRFS) data can be seen to be systematically lower than that for G-BASE urban results (Figure 1, Table 42). However, the largest contribution to that difference can be seen to be the Swansea G-BASE data (Table 42c, Figure 107). When the data are mapped, the highest concentrations can be seen to be around the area of the Swansea valley, and in NE Wales around the Halkyn Mountain area (Figure 108). When k-means cluster analysis is applied to the data, it is predominantly the Swansea valley area which is shown up (Figure 109); however, when the NSI(XRFS) data alone is used, this also highlights areas of north and central Wales (Figure 110) where Pb has historically been exploited as an economic ore.

The regional NSI(XRFS) data set median concentration of 55 mg/kg (Table 42) shows that concentrations of soil Pb are generally elevated in comparison to the European GEMAS measured median of 16 mg/kg (Reimann *et al*., 2012). There is extensive metalliferous mineral occurrence in Wales (Figure 3), and much of this is Pb-bearing either at grades that were of economic value, or as accessory mineralisation, and so is likely to form a primary cause of regionally elevated soil Pb concentrations. The importance of this link between mineralisation and mining activities leading to elevated soil concentrations means that it has been the subject of a long history of research in Wales; studies such as that of Griffith (1918) demonstrated that soil and livestock productivity were severely compromised by Pb in the valley soils of north Cardiganshire and documented over a century of awareness of this problem. Griffith (1918) estimated 3000 acres were affected by mining in north Cardiganshire.

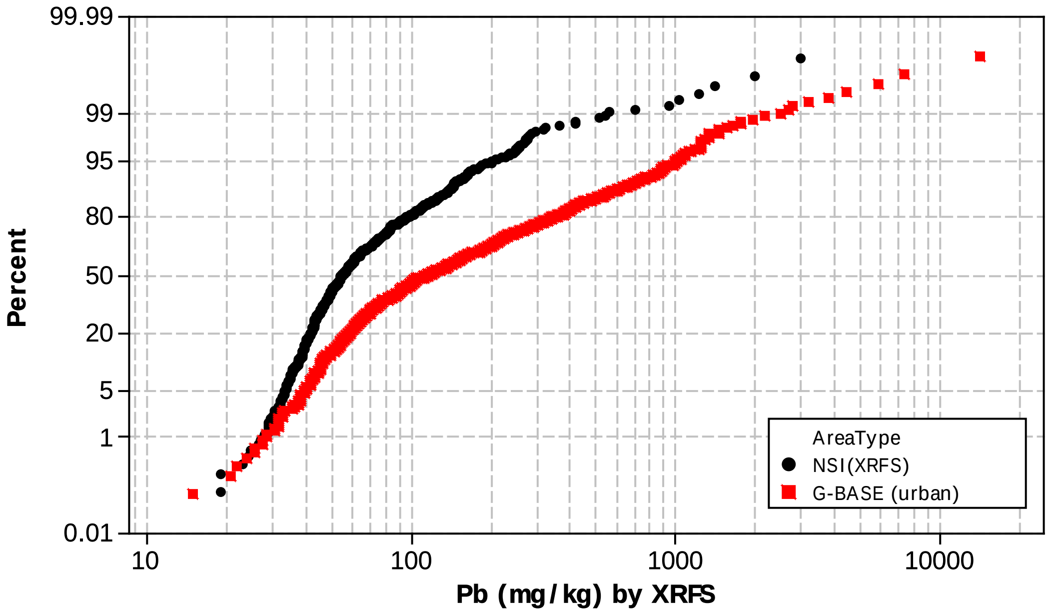
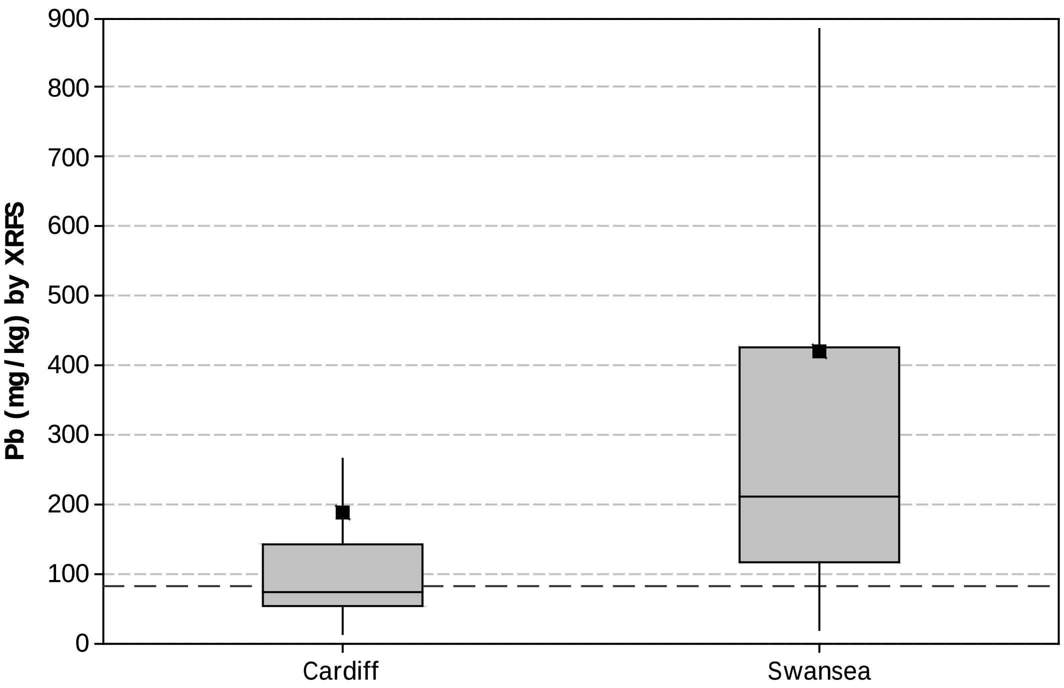


Figure 106: Cumulative probability plot of topsoil Pb concentrations for G-BASE urban and NSI(XRFS) results



Dashed line extends the interquartile range of the national NSI(XRFS) data

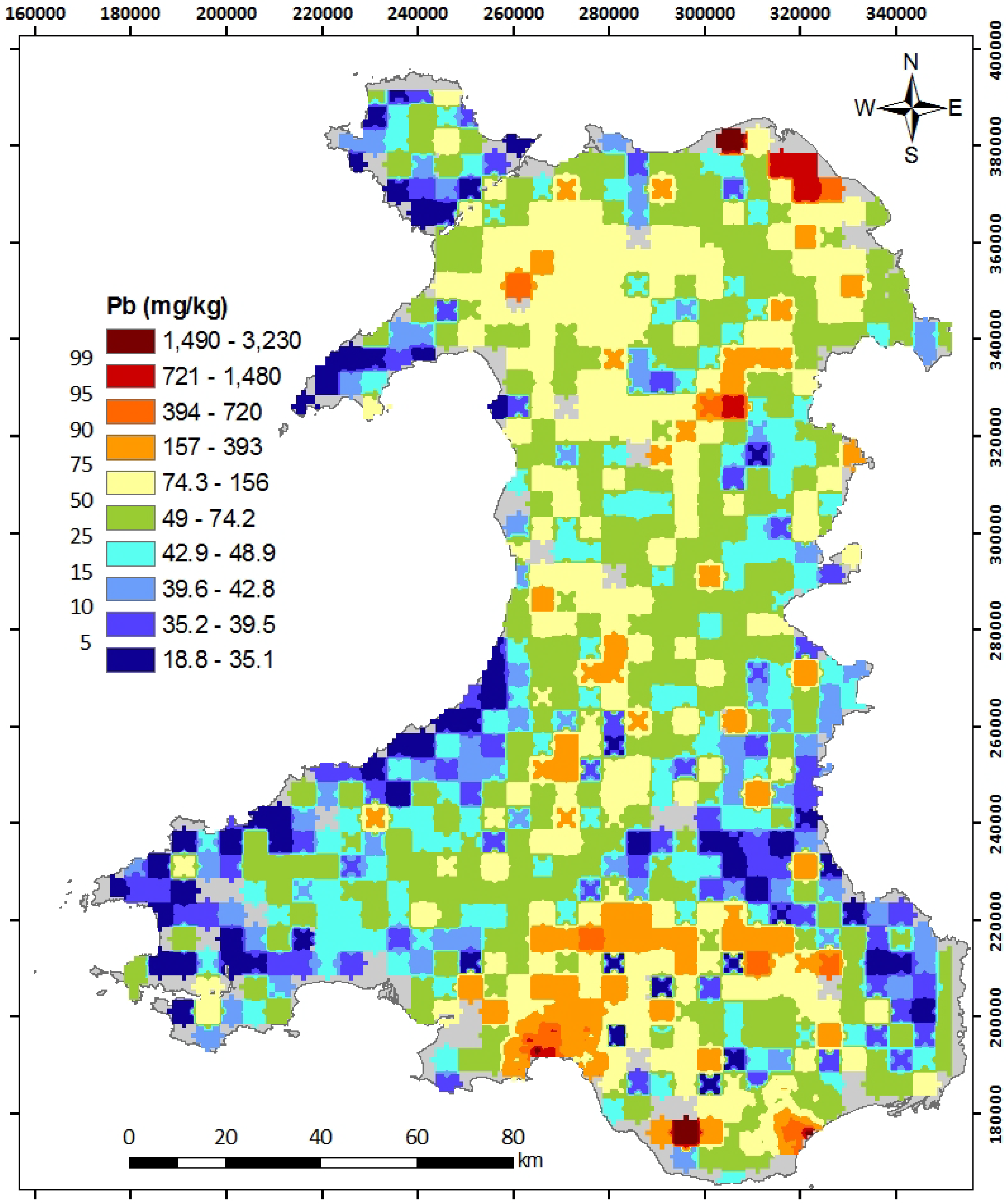
Figure 107: Boxplot of urban topsoil Pb data collected by the G-BASE project

(a) Pb (mg/kg)	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
All data	1675	192	14.8	48.9	74.2	156	14328	16

(b) Pb (mg/kg) Area type	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
NSI(XRFS)	798	87.8	18.8	43.1	54.9	83.2	2966	12
Urban	877	287	14.8	63.0	115	273	14328	12

(c) Pb (mg/kg) Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Cardiff	507	188	14.8	53.7	75.1	144	7377	11
Swansea	370	422	20.6	117	213	426	14328	11

Table 42: Statistical summary of topsoil Pb in the main data sets (a) whole data set (b) by sampling density (c) by urban centres sampled by G-BASE



Colour thresholds are designed for highly skewed data

Figure 108: Interpolated map of topsoil Pb data

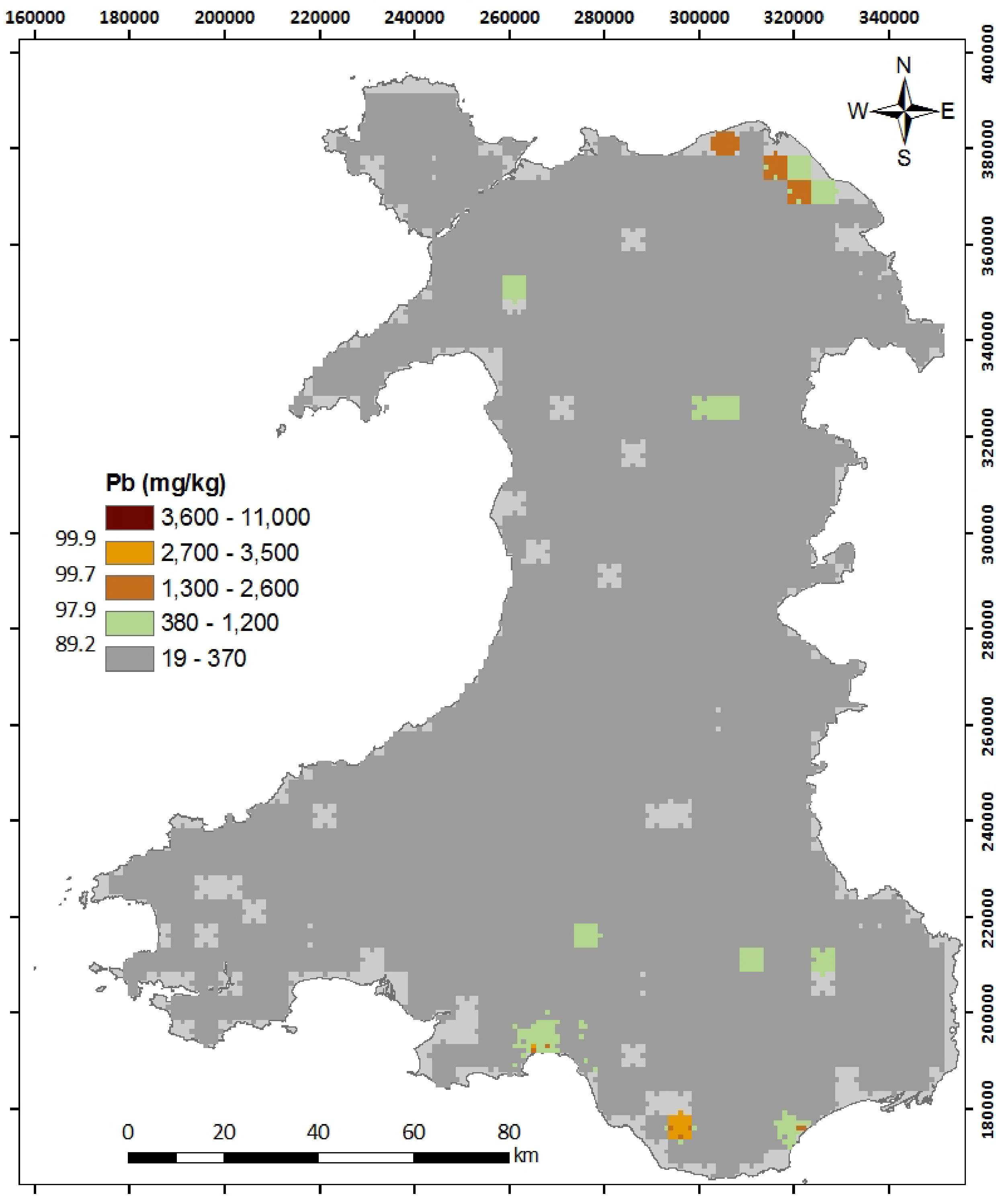


Figure 109: Interpolated map of topsoil Pb concentrations. Thresholds determined by using k-means cluster analysis

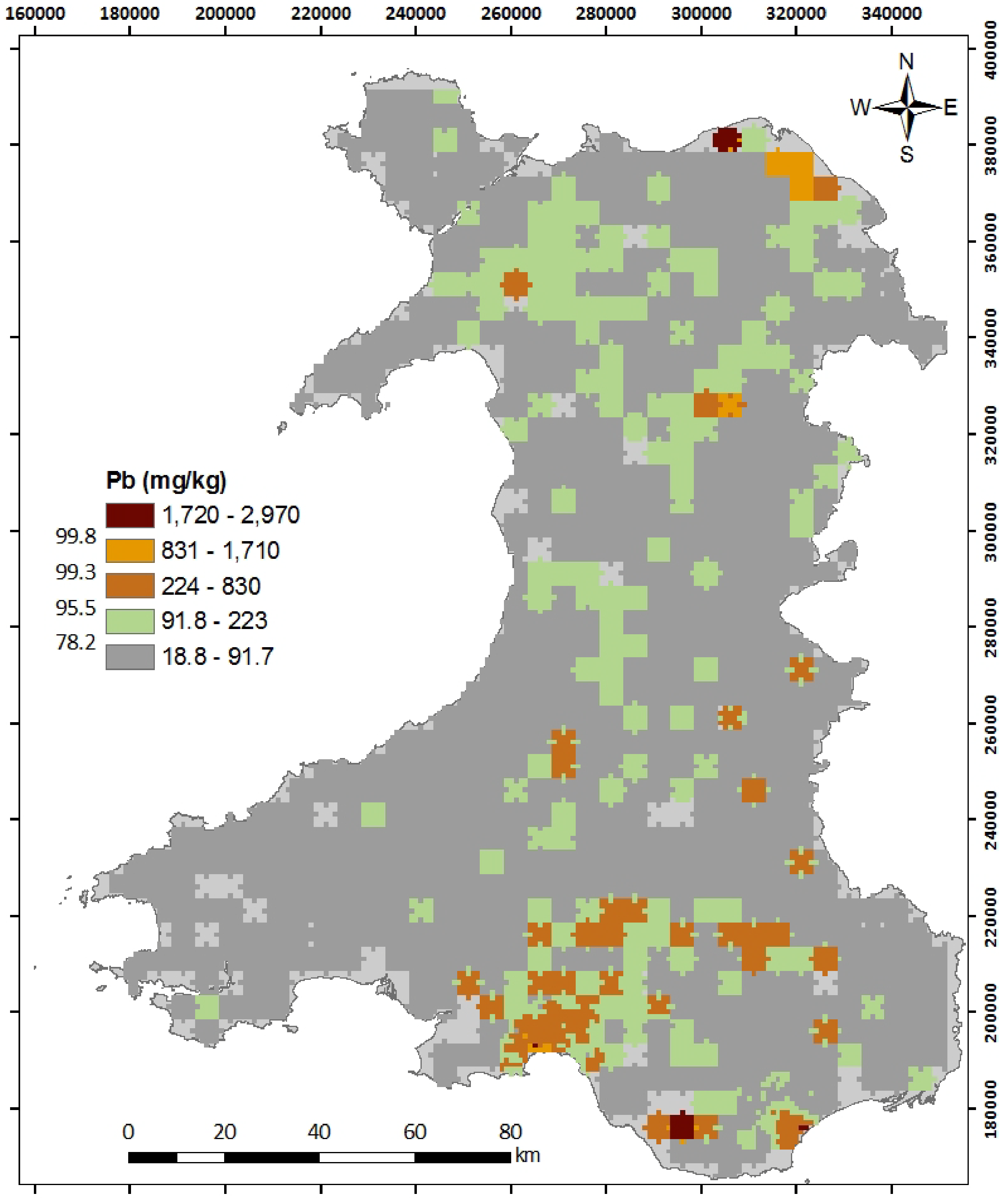


Figure 110: Interpolated map of topsoil Pb concentrations from the NSI(XRFS) data set only. Thresholds determined by using k-means cluster analysis

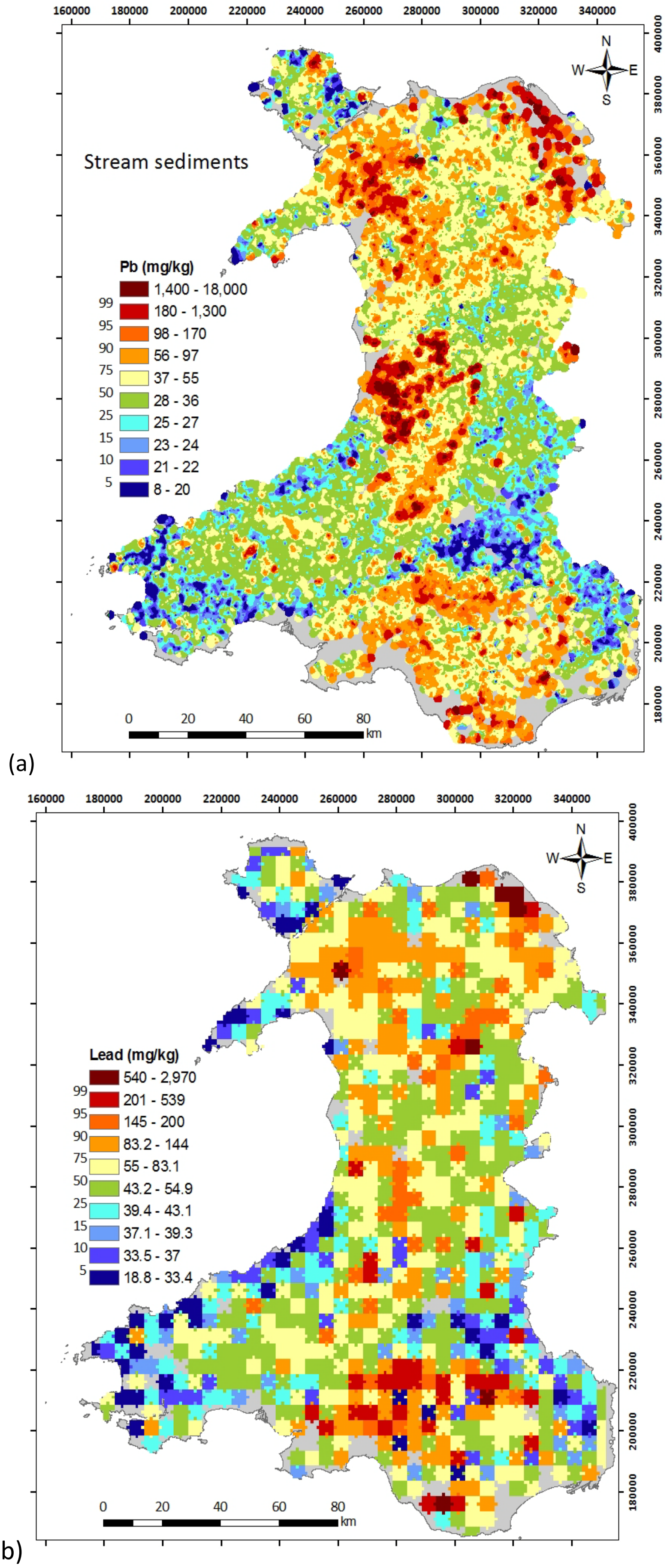


Figure 111: Lead in (a) G-BASE stream sediments and (b) NSI(XRFS) soils

8.2 DOMAIN SELECTION

The information shown below was used to compile the data for domains related to:

- non-ferrous metalliferous mineralisation and mining areas; and
- urban areas.

8.2.1 Urban areas

The impact of urbanisation is an almost ubiquitous feature of Pb concentrations in baseline mapping. This is seen in the Swansea data above and to a lesser extent in the Cardiff data. However, in the stream sediment data, which is compared to the NSI(XRFS) data in Figure 111, this likely control is seen to be of lower impact than that thought to arise from metalliferous mineralisation and mining, and mineral processing. High concentrations in the Swansea valley, the consequence of the extensive smelting and general industrial activity, were investigated by Davies (1997) in an area with a nickel smelter and historical base metal smelters. The median soil Pb concentration was 100 mg/kg, using a partial extraction and the maximum concentration 800 mg/kg (Davies, 1997).

Data relating the sample sites to the extent of urbanisation is shown in Figure 112 and Table 43. However, when the samples collected by G-BASE in Cardiff and Swansea are also included, it can be seen that the marked effect is in the Swansea data, and this is stronger than that from the extent of urbanisation (Figure 113 and Table 47). For this reason, the three valleys associated with higher typical Pb topsoil concentrations (Loughor, Tawe and Neath) have also been used to establish if there is a higher concentration domain (Figure 114) than of the general urbanisation impact.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Large Urban Area polygon	452	359	20.6	75.1	152	400	14300	12
Small Urban Area polygon	40	111	27.8	49.0	81.1	113	761	4
Non-urban	1183	131	14.8	45.5	61.5	111	7380	14

Table 43: Summary statistics for topsoil Pb (mg/kg), over urban and non-urban areas

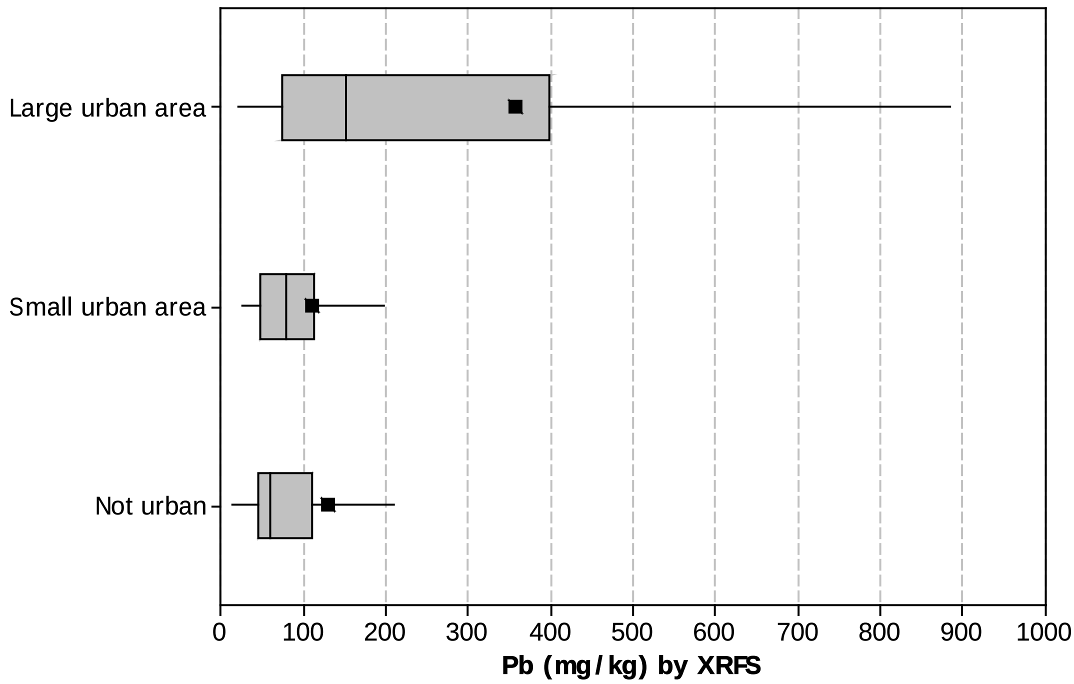


Figure 112: Boxplot of topsoil Pb concentrations over urban and non-urban areas

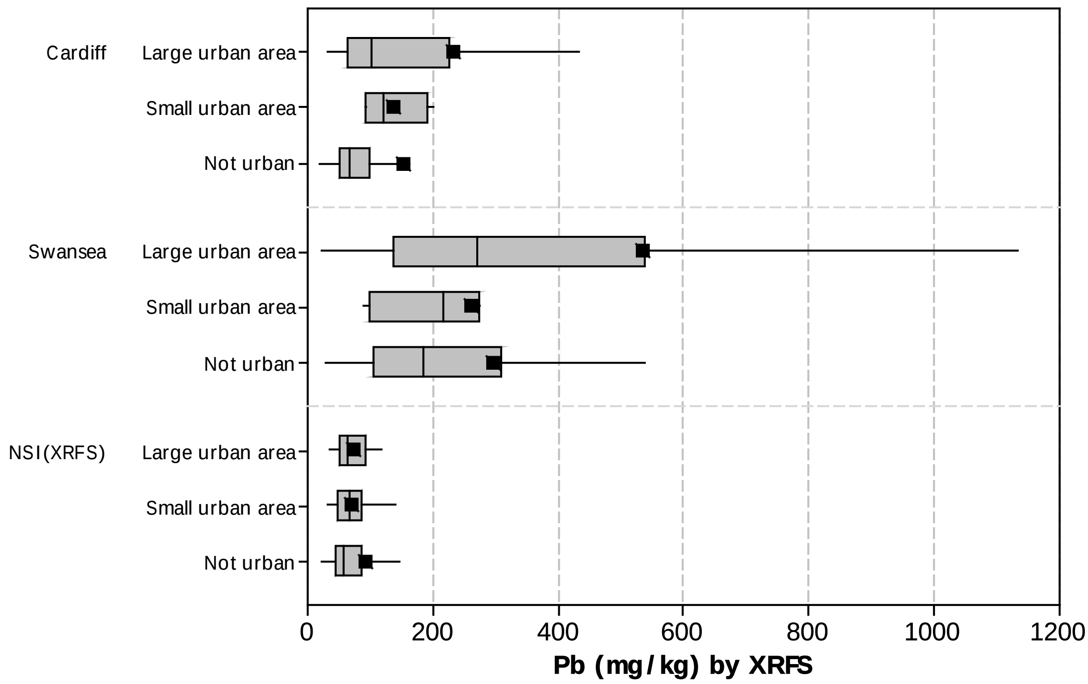


Figure 113: Boxplot of topsoil Pb concentrations categorised by urbanisation extent and areas

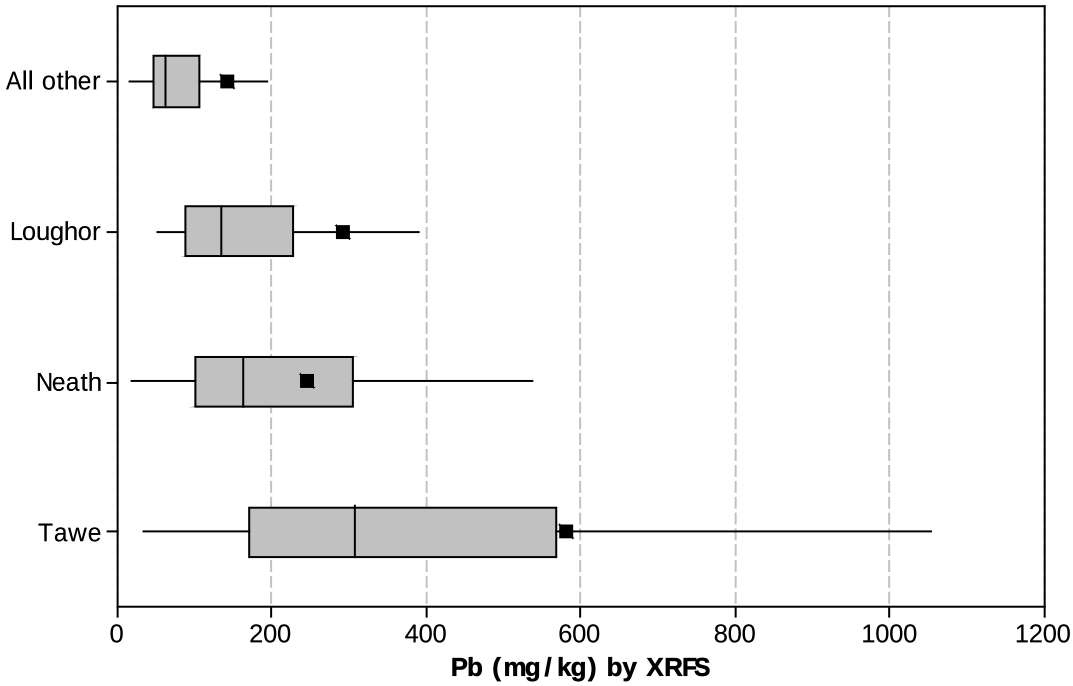


Figure 114: Comparison of Pb concentrations in soils of the Loughor, Tawe and Neath valleys

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Large urban areas								
Cardiff	239	231	28.4	61.5	100	224	1764	3
Swansea	199	534	20.6	136	269	539	14300	10
NSI(XRFS)	14	68.9	33.5	49.3	62.2	88.3	116	0.6
Small urban areas								
Cardiff	6	133	89.8	89.8	118	189	198	0.7
Swansea	7	258	87.8	96.6	213	271	761	2
NSI(XRFS)	27	67.2	27.8	44.8	63.1	83.1	169	2
Not urban								
Cardiff	262	150	14.8	47.9	64.4	95.1	7380	11
Swansea	164	294	25.5	104	182	308	5920	9
NSI(XRFS)	757	88.9	18.8	43.0	54.5	83.0	2970	11

Table 44: Summary statistics for G-BASE urban and NSI(XRFS) Pb (mg/kg) data categorised by urbanisation extent

8.2.2 Non-ferrous metalliferous mineralisation and mining

There are major and minor ore-fields across Wales which have abundant Pb mineralisation, or include accessory Pb minerals. Where the Pb was economic, these comprised some of the largest Pb mines in Britain at the time they were exploited.

Davies and Paveley (1988) report systematic surface soil Pb data for Wales, but the sample numbers and description of the sampling teams are sufficiently similar to those for Wales published within the original NSI data (McGrath and Loveland, 1992) that it is not clear if these are an independent data set, or not, and during our current study we have been unable to clarify this. Davies and Paveley (1988) reported data deciles, including a median of 29 mg/kg.

They also show substantially elevated concentrations over the Halkyn Mountain area, but less so over the ore-field of mid-Wales, which they attribute to sampling missing the sinuous valley soils, which are the most heavily contaminated by Pb. They also note the historical Pb mines in the Carboniferous Limestone of the Vale of Glamorgan giving rise to locally elevated concentrations. Halkyn Mountain, a major source of British Pb ore (1845-1938), has been found to have a mean Pb concentration of 3300 mg/kg in a study of 60 garden and field soils (Davies and Roberts, 1975).

A baseline survey of northwest Pembrokeshire (Bradley *et al*., 1978), an area without significant metalliferous mineralisation, found a mean soil Pb concentration of 40 mg/kg (261 samples), identical to the mean concentration for England and Wales of Archer and Hodgson (1521 samples; median 37 mg/kg) (Archer and Hodgson, 1987). Alloway and Davies (1971) studied soils in the Aeron catchment to calculate background concentrations in comparison to mineralised catchments of Ceredigion. They estimated a 'Normal range' of Pb 12-72 mg/kg, compared to alluvium in the Rheidol (Pb of 90 – 2900 mg/kg), which is greatly affected by mineralisation and mining wastes. These data are similar to those of Griffith (1918) who found 2300-3500 mg/kg in the Rheidol valley soils, although higher than the study of Davies and Lewin (1974) - mean 900 mg/kg in the same valley, and a mean of 450 mg/kg (median 128 mg/kg) in the Tanat valley (Fuge *et al*., 1989). Alloway and Davies (1971) also note the likely impact of smelter fumes, where this was done close to the point of mineral extraction, on the redistribution of Pb from ore in the local environment.

The stream sediment map (Figure 111a) shows a strong influence of both the natural dispersion of Pb from the underlying ores, and the impact from mining and mineral processing. This was widespread in the mid-Wales, NE (including Halkyn) and Snowdonia areas. However, it is harder to discern this impact in the comparable soil map in Figure 111b, or in the boxplot comparing data overlying these areas with the rest of the topsoil data (Figure 115). This is mostly likely to be due to the relatively low sample density, which results in low sample numbers per ore-field (Table 45), with only the Halkyn area topsoils having substantially elevated Pb concentrations.

Due to the well-understood abundance of Pb in these ore-fields, and the enhanced concentrations observed in the stream sediment data, it is likely that, as more systematic study data becomes available, so the typical domain concentrations may increase.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Anglesey	3	73.0	27	27.4	50.3	141	141	1.5
north	22	146	38	53.9	80.1	102	955	3.1
Halkyn	15	388	41	62.0	127	183	2018	1.9
central	31	90.1	41	53.0	73.6	136	219	1.1
south	7	502	53	56.0	75.0	186	2966	2.6
none	1597	192	15	48.7	73.5	159	14328	16

Table 45: Summary statistics for topsoil Pb (mg/kg), over mineralised areas

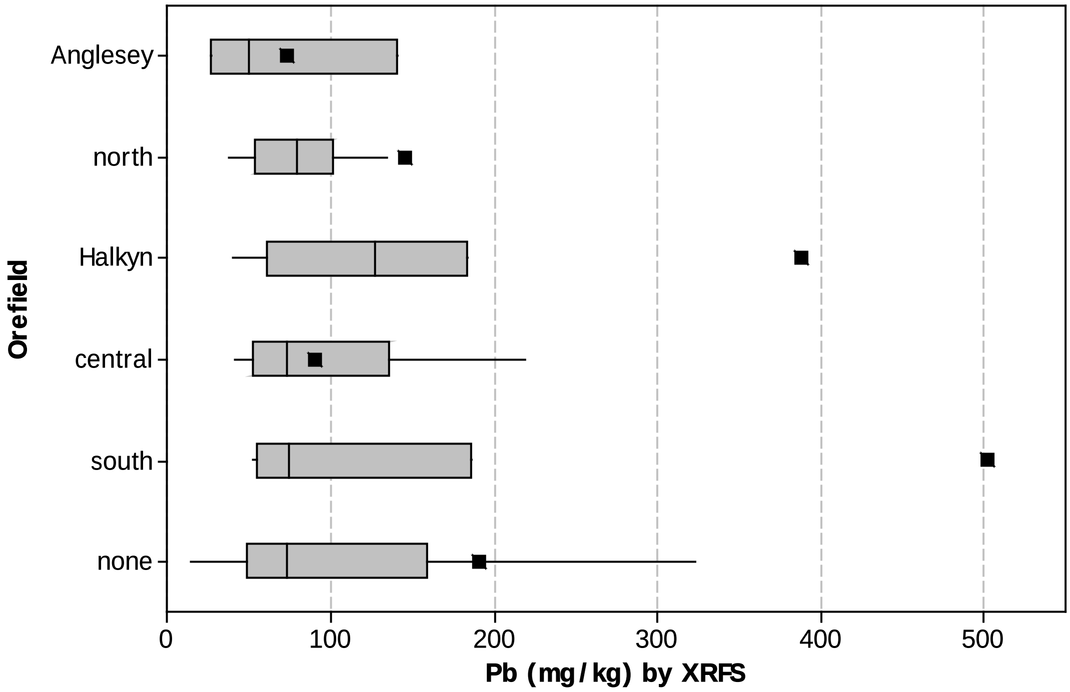


Figure 115: Boxplot of topsoil Pb over metalliferous mineralisation and mining areas

8.3 DOMAIN DATA SUMMARY

As a result of the exploratory data analysis above, the samples associated with the Loughor, Tawe and Neath valleys have been classified as Urban 1 Domain, with all other samples from 'Large urban areas' as 'Urban'. The 'Urban' samples that fall outside the Urban 1 Domain are classified as the Urban 2 Domain. The samples from the Halkyn area have been classified as the Mineralisation 1 Domain on the basis of the concentrations (although there are only fifteen samples). The other mineralised areas have been classified as the Mineralisation 2 Domain, although it may be expected that typical concentrations increase when more data becomes available. The remaining data are used to define the Principal Domain. These data are summarised in Figure 116 and Figure 117, with summary statistics in Table 46. The domains are mapped in Figure 118.

Area name	Number	Mean	Minimum	25th percentile	Median	75th percentile	Maximum	Skewness
Mineralisation 1	15	388	41	62.0	127	183	2018	1.9
Mineralisation 2	61	157	27	53.0	75.1	128	2966	7
Urban 1	342	386	19	112	208	408	14328	13
Urban 2	291	249	21	61.5	102	232	2696	3.1
Principal	966	105	15	43.4	56.4	83.5	7377	16

Table 46: Summary statistics of topsoil Pb (mg/kg) in the domains

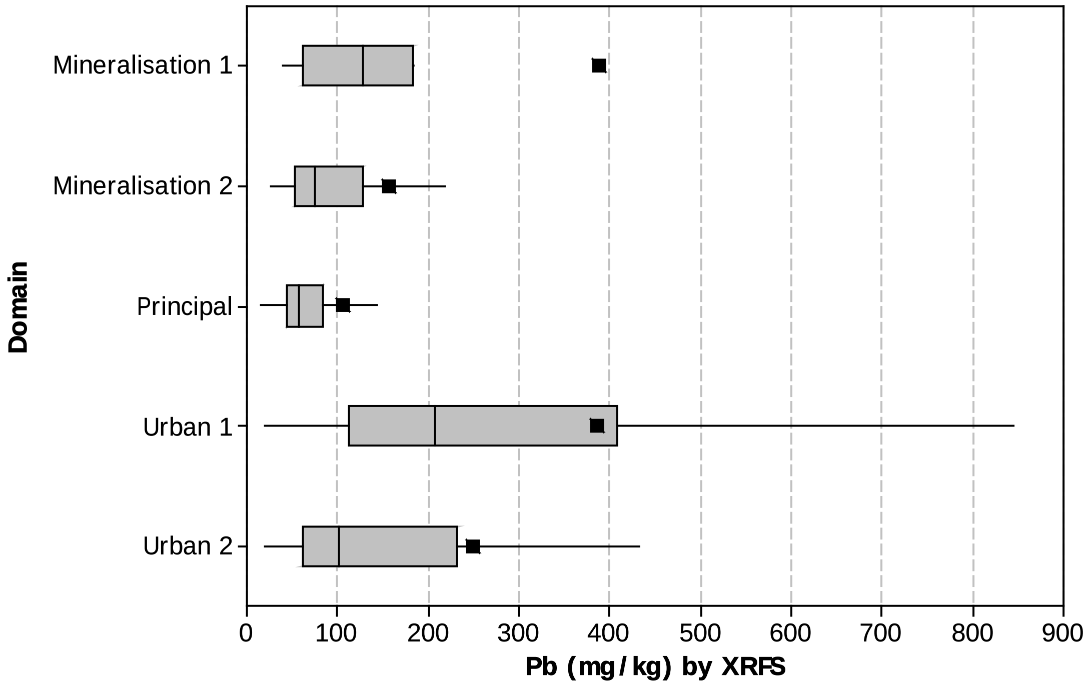


Figure 116: Boxplot of Pb data classified by domains

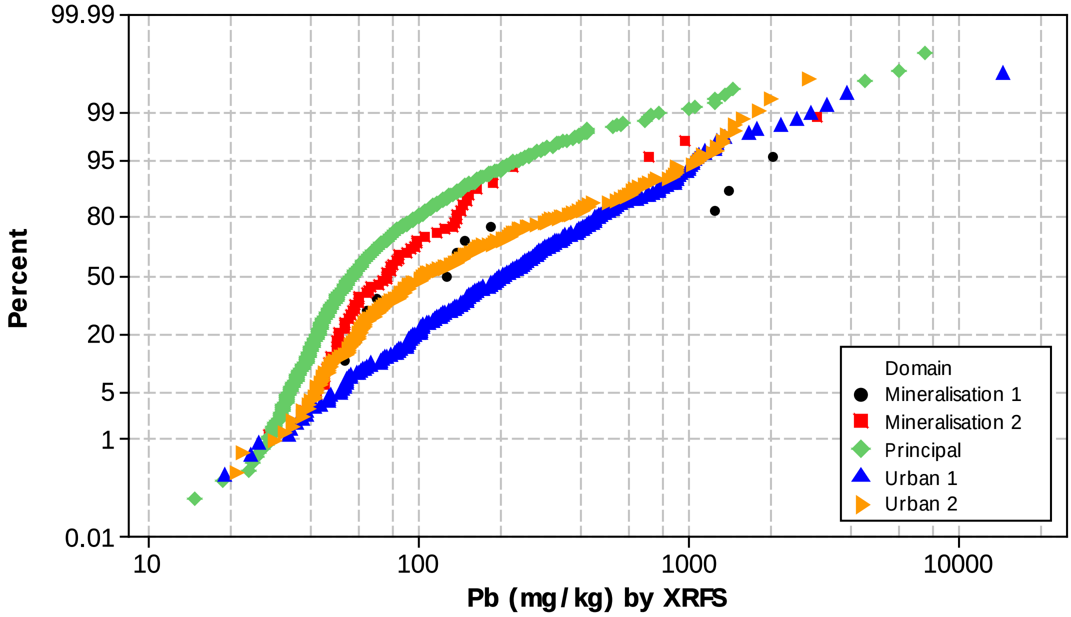
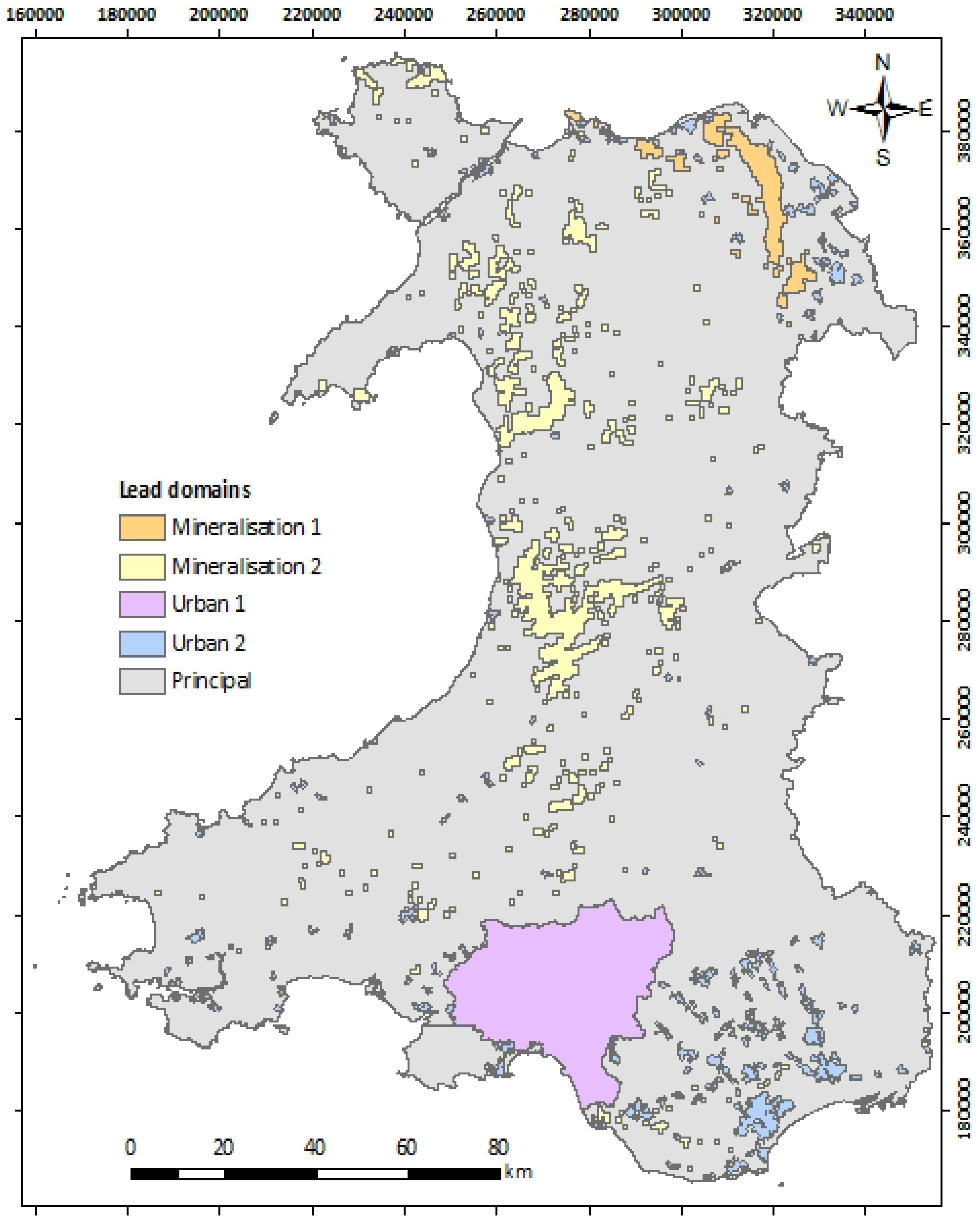


Figure 117: Probability plot of Pb topsoil data classified by domains



Contains Ordnance Survey data © Crown copyright and database rights (2012)
Urban areas defined from OS Strategi® used under OS OpenData™ terms
Mineralisation after the Ove Arup data

Figure 118: Map showing the location of Pb domains

8.4 Pb NBC CALCULATION

The NBC calculations for Pb have been carried out according to the method described in Cave *et al.* (2012).

8.4.1 Principal Domain

Untransformed and $\log_e$ transformed skewness coefficient and octile skew fail the test criteria for an underlying Gaussian distribution (Figure 119). Therefore a Box-Cox transformation was applied (Figure 120 and Figure 121). Percentiles are summarised in Table 47 and an example NBC is given in Figure 122.

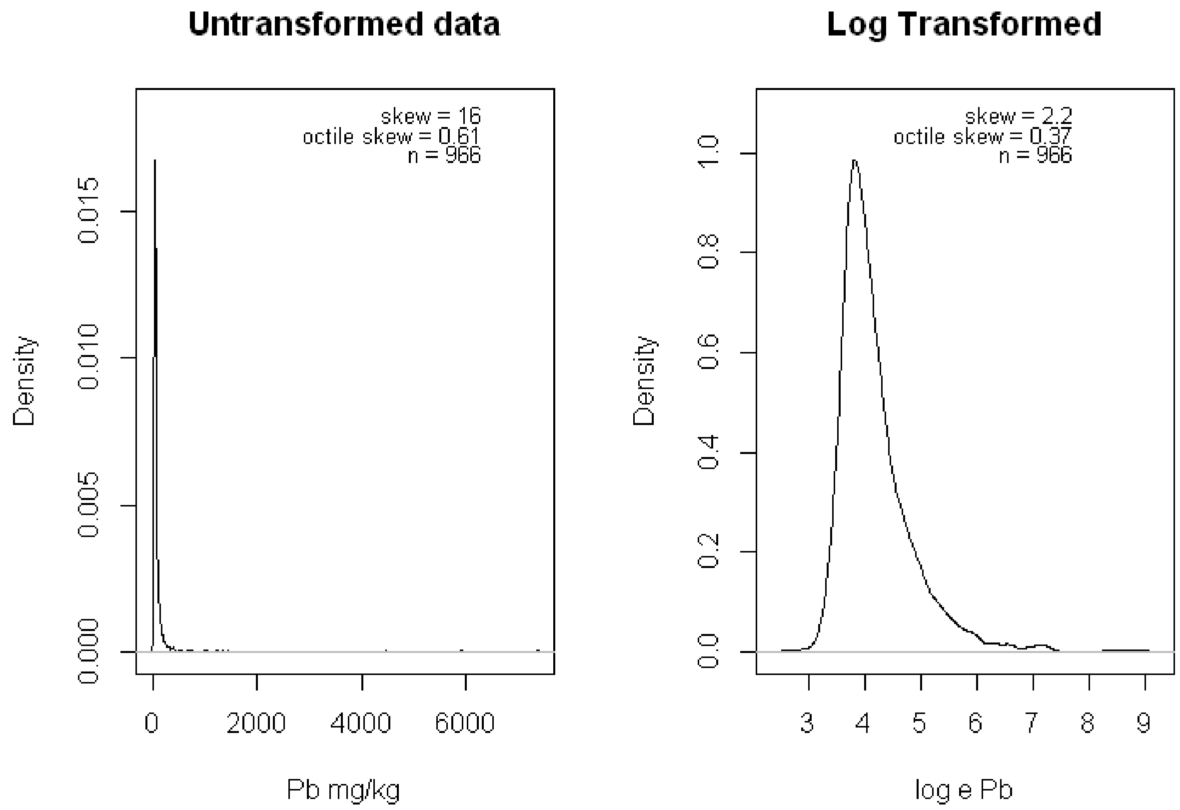


Figure 119: Density distributions for the raw data and the $\log_e$ transformed data for Pb in the Principal Domain (n = number of samples)

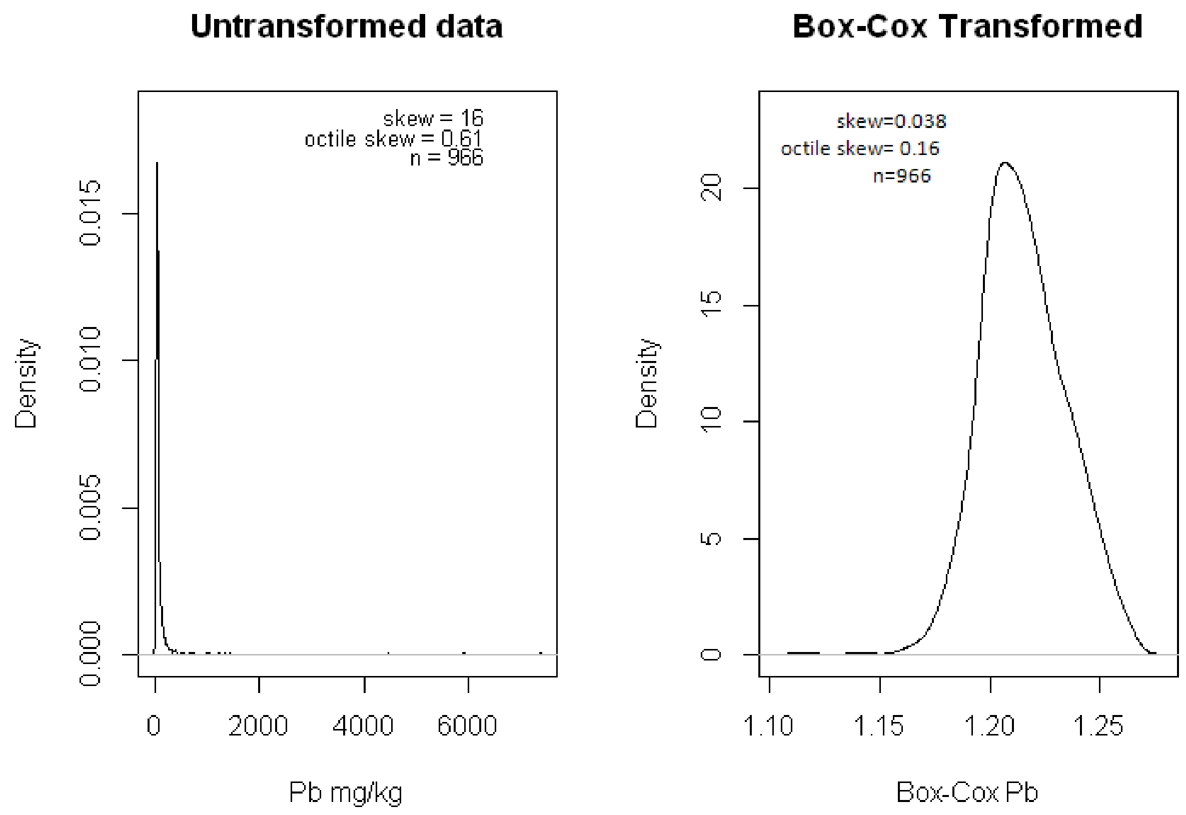


Figure 120: Density distributions for the raw data and the Box-Cox transformed data for Pb in the Principal Domain (n = number of samples)

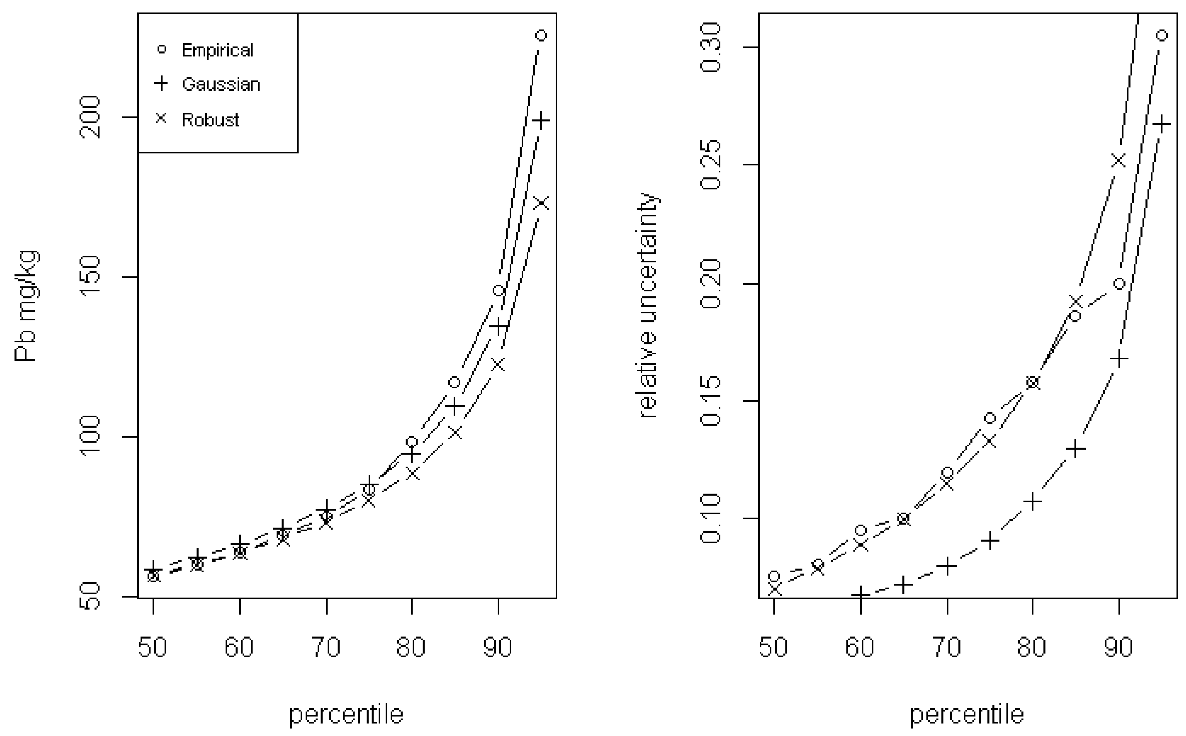


Figure 121: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Pb in the Principal Domain

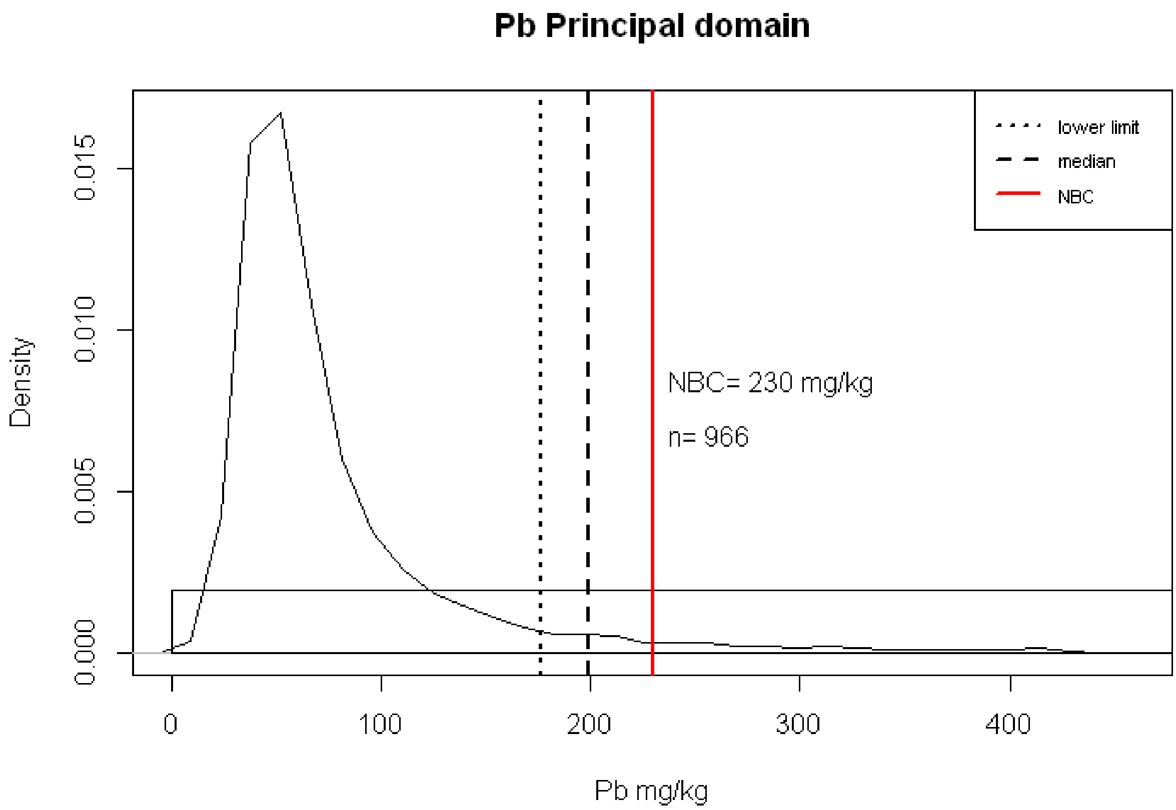


Figure 122: Summary density plot and histogram of the distribution for Pb in the Principal Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	54.3	56.4	58.1	**57.0**	**58.6**	**60.3**	54.1	56.4	58.2
55	57.6	60.2	61.9	**60.5**	**62.3**	**64.2**	57.1	59.8	61.9
60	61.1	63.7	66.9	**64.4**	**66.5**	**68.8**	60.4	63.6	66.2
65	66.4	69.3	72.9	**68.9**	**71.4**	**74.1**	64.3	68.1	71.3
70	71.6	75.1	80.1	**74.3**	**77.4**	**80.6**	68.8	73.4	77.4
75	79.0	83.4	90.6	**81.1**	**84.9**	**88.8**	74.3	80.0	85.1
80	89.8	98.5	105	**90.0**	**94.9**	**99.9**	81.5	88.8	95.7
85	106	117	128	**103**	**110**	**117**	91.6	101	111
90	135	146	165	**124**	**135**	**147**	108	123	139
95	197	226	269	**176**	**199**	**229**	143	173	213

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 47: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Pb (mg/kg) in the Principal Domain

8.4.2 Mineralisation 1 Domain

With only fifteen samples in this domain it is not possible to calculate an NBC (Cave *et al.*, 2012).

8.4.3 Mineralisation 2 Domain

Untransformed and $\log_e$ transformed skewness coefficient and octile skew fail the test criteria for underlying Gaussian distribution (Figure 123). Therefore a Box-Cox transformation was applied giving skewness coefficient and octile skew that meet the test criteria for underlying Gaussian distribution. However, the Box-Cox transform results in a very high and uncertain NBC, with the outliers (Figure 124) having disproportionate influence on upper end of the 95 percentile confidence interval. The suggested NBC has been derived from the $\log_e$ transformation (Figure 125, Figure 126 and Table 48). There is a clear need for more data for this domain.

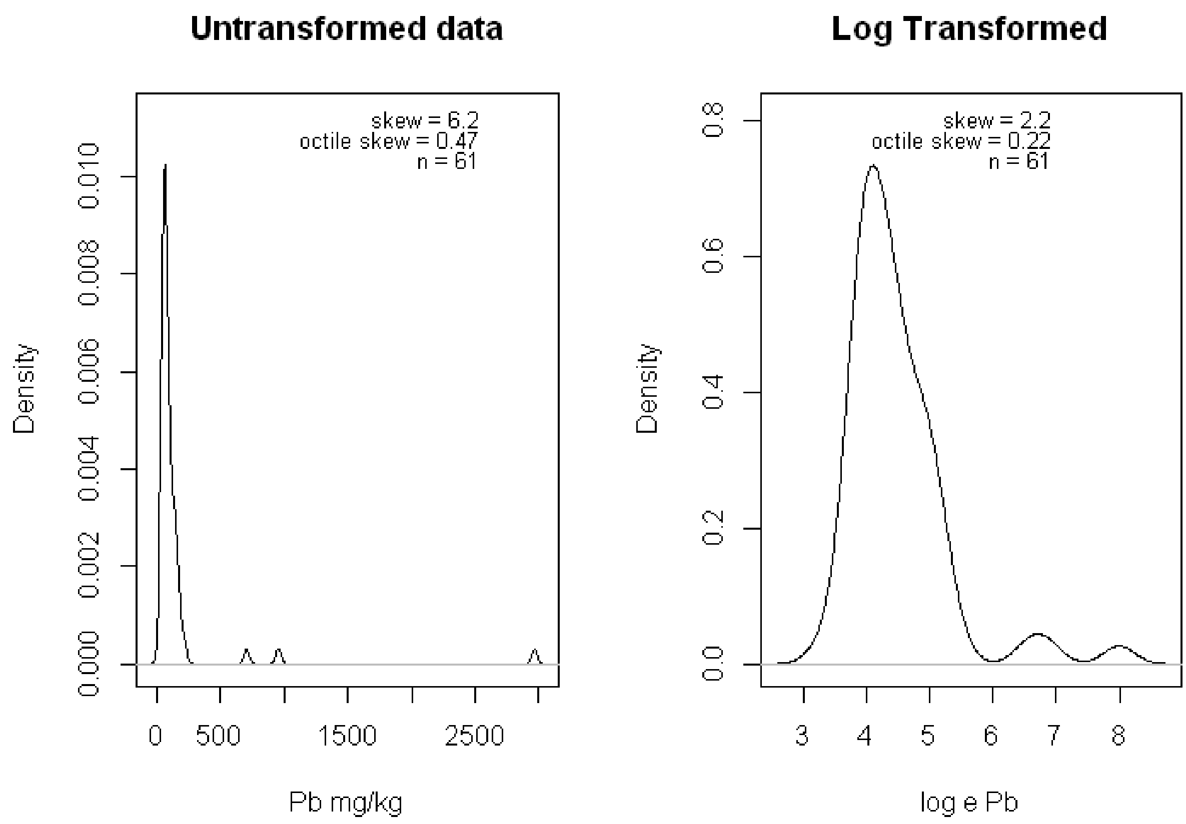


Figure 123: Density distributions for the raw data and the $\log_e$ transformed data for Pb in the Mineralisation 2 Domain (n = number of samples)

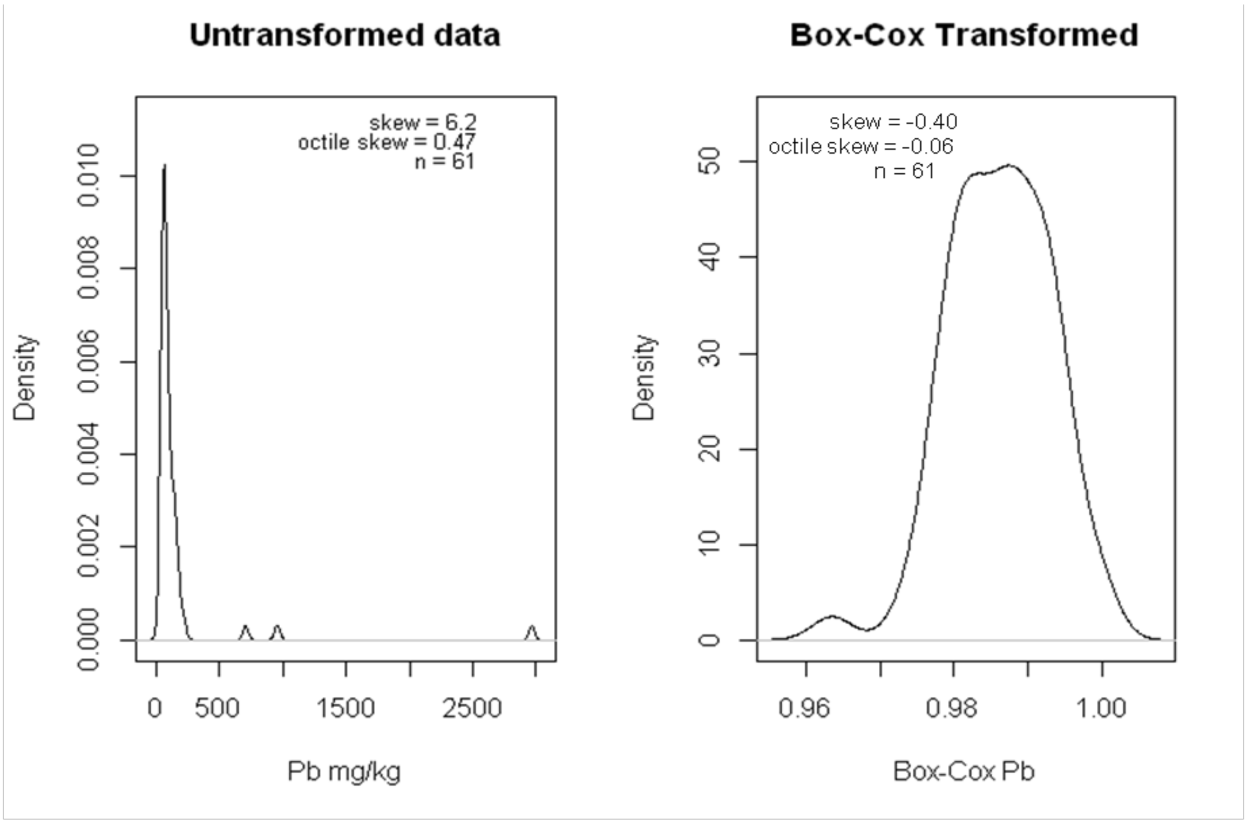


Figure 124: Density distributions for the raw data and the Box-Cox transformed data for Pb in the Mineralisation 2 Domain (n = number of samples)

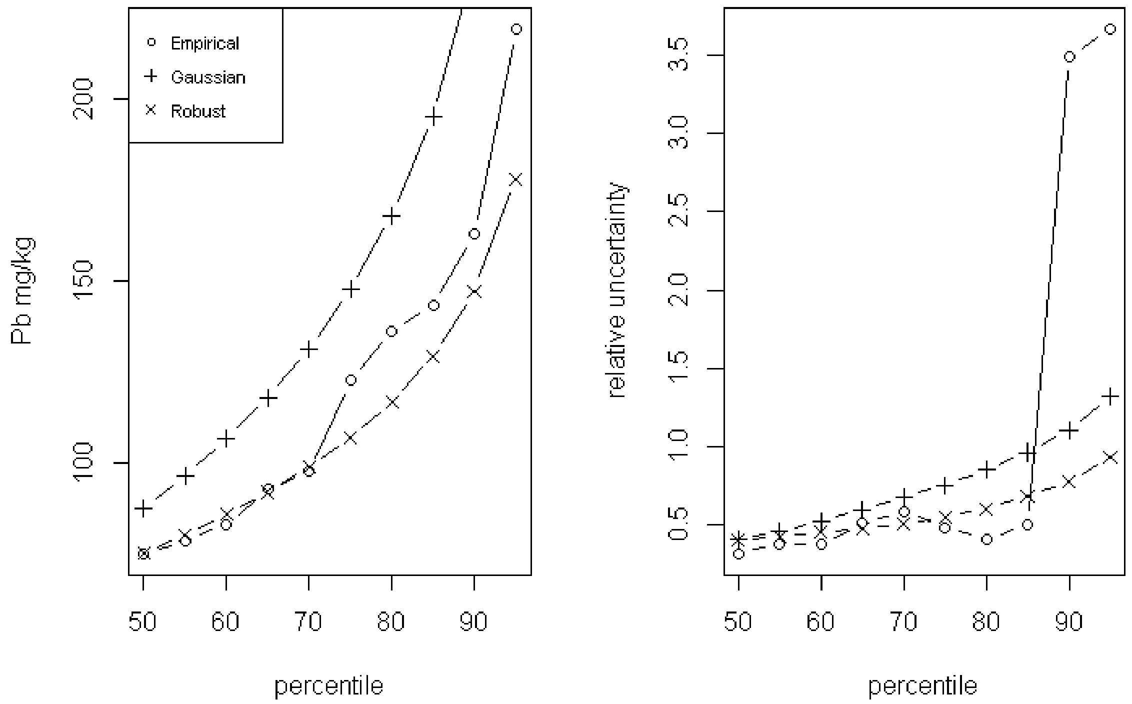


Figure 125: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Pb in the Mineralisation 2 Domain

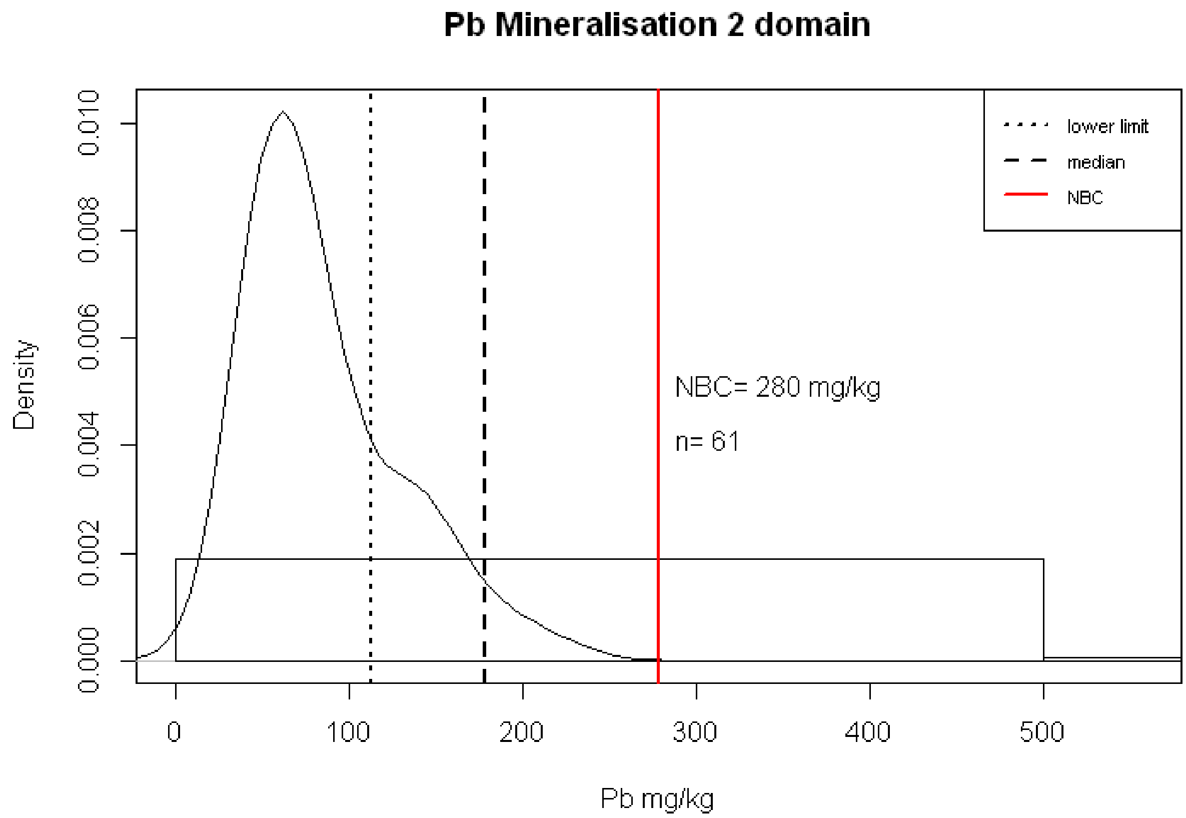


Figure 126: Summary density plot and histogram of the distribution for Pb in the Mineralisation 2 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	60	75	84	73	88	110	**60**	**75**	**90**
55	65	79	95	78	97	120	**63**	**80**	**97**
60	74	83	100	83	110	140	**66**	**86**	**100**
65	76	93	120	88	120	160	**70**	**92**	**110**
70	80	98	140	94	130	180	**74**	**99**	**120**
75	84	120	140	100	150	210	**79**	**110**	**140**
80	97	140	150	110	170	250	**84**	**120**	**150**
85	110	140	190	120	200	310	**90**	**130**	**180**
90	140	160	710	130	240	400	**99**	**150**	**210**
95	150	220	950	160	310	570	**110**	**180**	**280**

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 48: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Pb (mg/kg) in the Mineralisation 2 Domain

8.4.4 Urban 1 Domain

Untransformed skewness coefficient and octile skew fail the test criteria for underlying Gaussian distribution. When $\log_e$ transformed was used, the skewness coefficient was <1 and octile skew <0.2, indicating an underlying Gaussian distribution (Figure 127 and Figure 128) for which parametric percentiles were used (Table 49) to calculate an example NBC as shown in Figure 129.

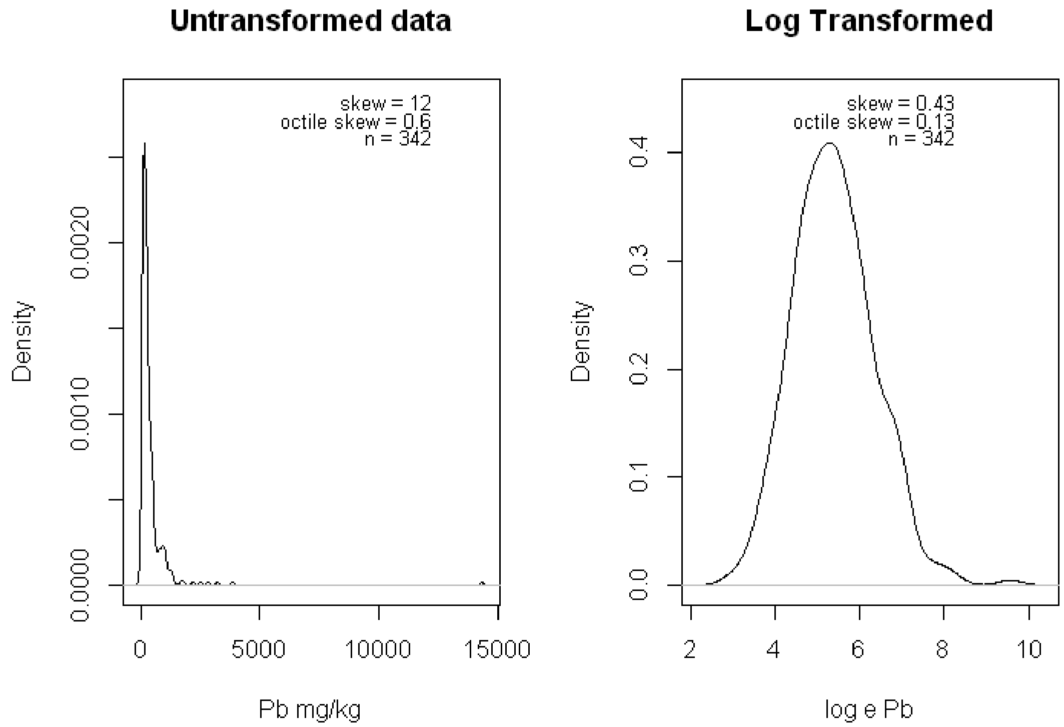


Figure 127: Density distributions for the raw data and the $\log_e$ transformed data for Pb in the Urban 1 Domain (n = number of samples)

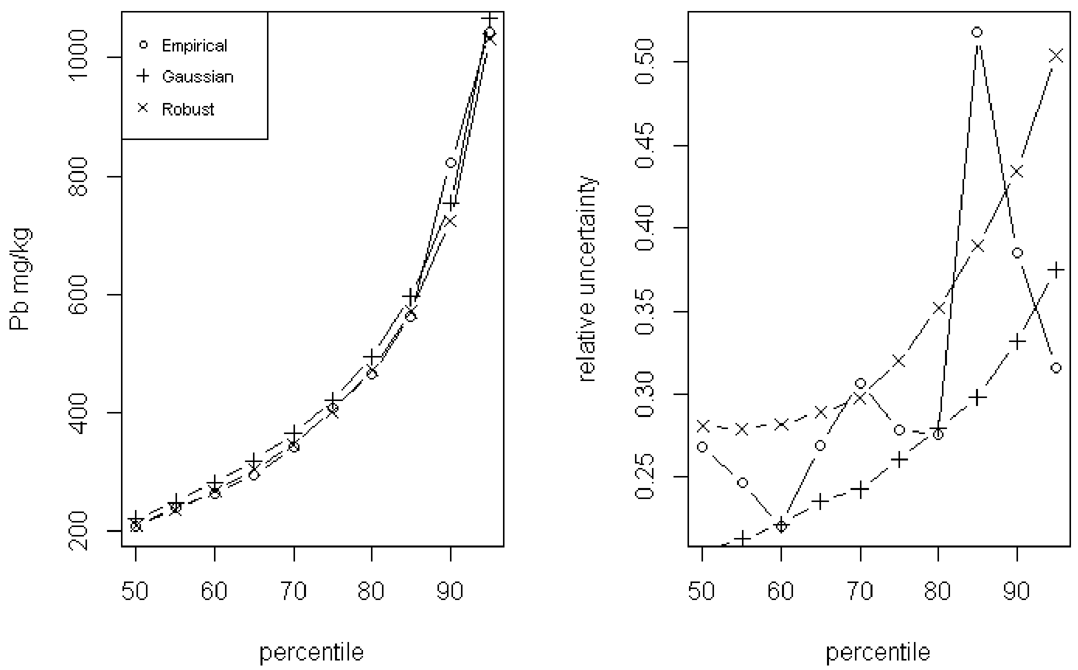


Figure 128: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Pb in the Urban 1 Domain

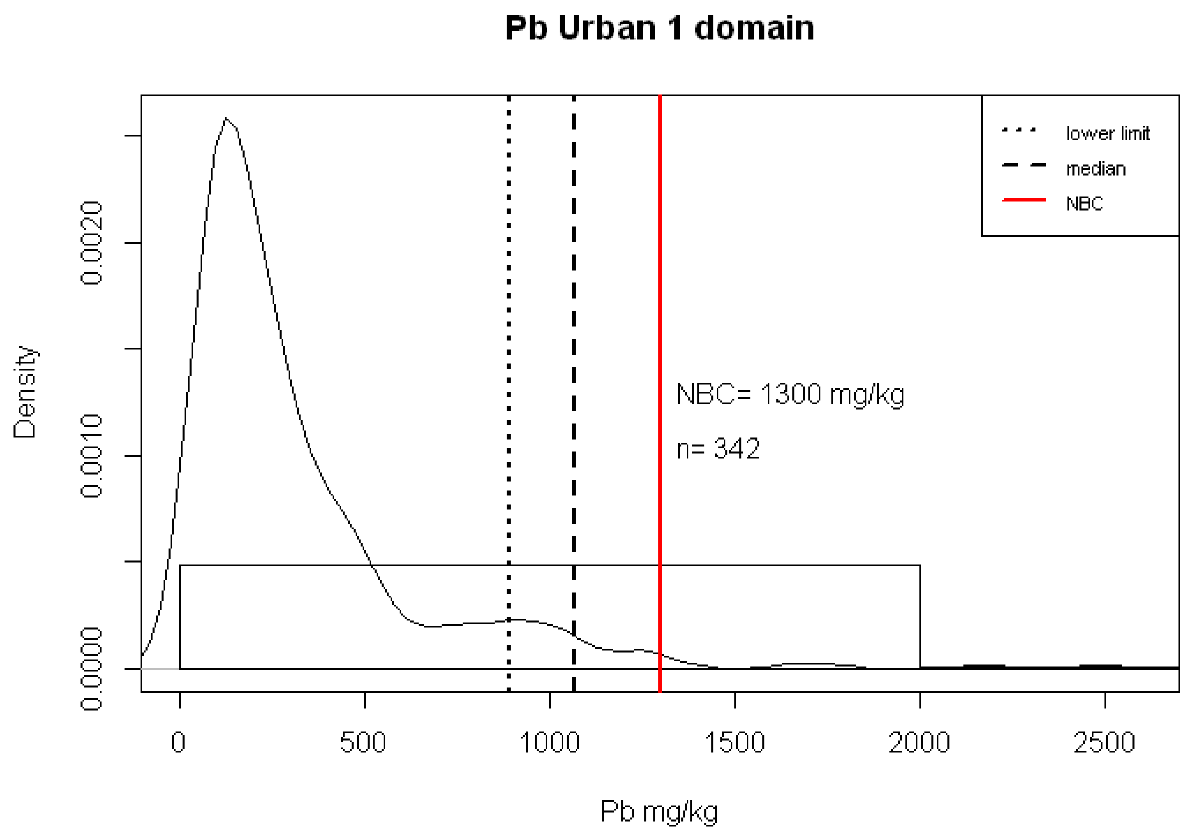


Figure 129: Summary density plot and histogram of the distribution for Pb in the Urban 1 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	188	208	242	**197**	**220**	**245**	185	208	242
55	208	239	267	**222**	**248**	**278**	208	235	273
60	242	264	301	**251**	**281**	**315**	235	266	307
65	264	294	343	**283**	**319**	**359**	264	303	349
70	295	341	406	**322**	**364**	**413**	299	347	398
75	343	406	457	**370**	**421**	**481**	340	401	463
80	415	466	538	**429**	**494**	**568**	393	472	550
85	478	564	761	**511**	**595**	**695**	464	571	673
90	658	822	971	**633**	**753**	**891**	571	724	864
95	915	1043	1243	**875**	**1068**	**1299**	776	1031	1255

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 49: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Pb (mg/kg) in the Urban 1 Domain

8.4.5 Urban 2 Domain

Untransformed skewness coefficient and octile skew fail the test criteria for underlying Gaussian distribution. Although Log_e transform gave an octile skew >0.2, the skewness coefficient was <1 (Figure 130), which indicated an underlying Gaussian distribution, for which parametric percentiles were used to calculate an example NBC (Figure 131). Percentiles are given in Table 50 and an example NBC shown in Figure 132.

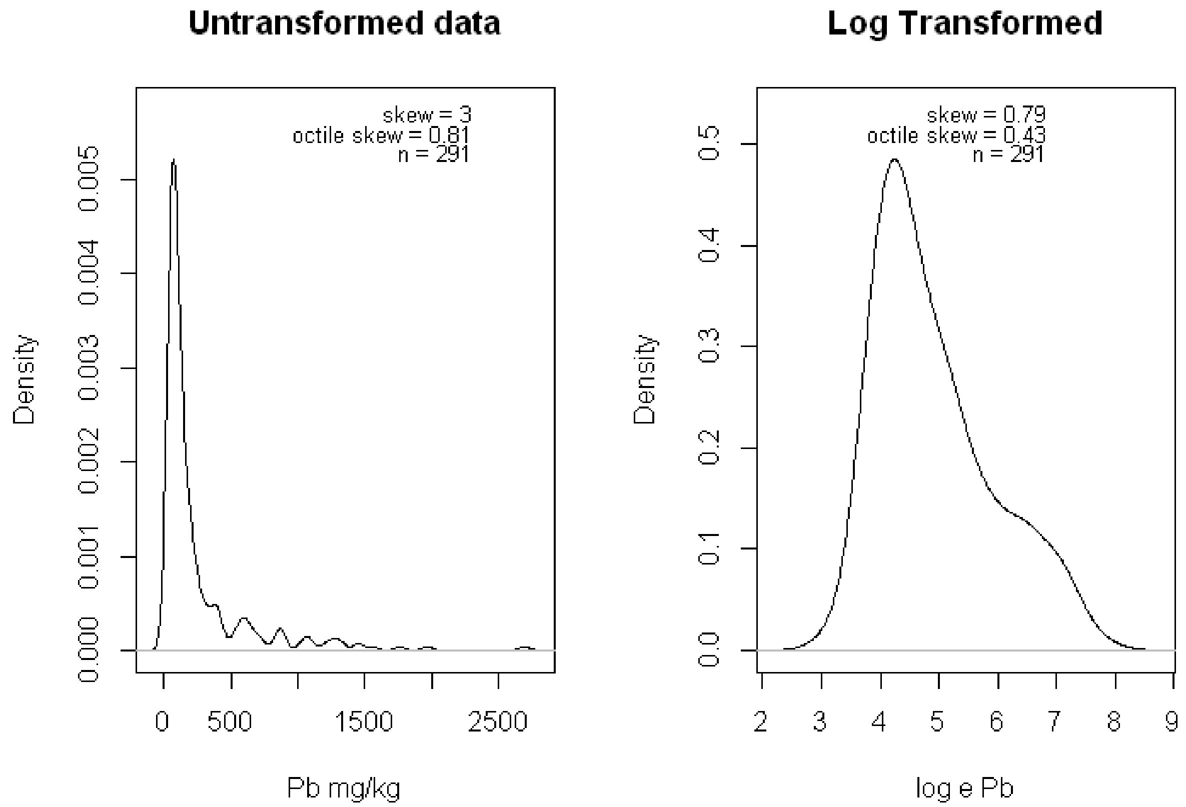


Figure 130: Density distributions for the raw data and the log$_e$ transformed data for Pb in the Urban 2 Domain (n = number of samples)

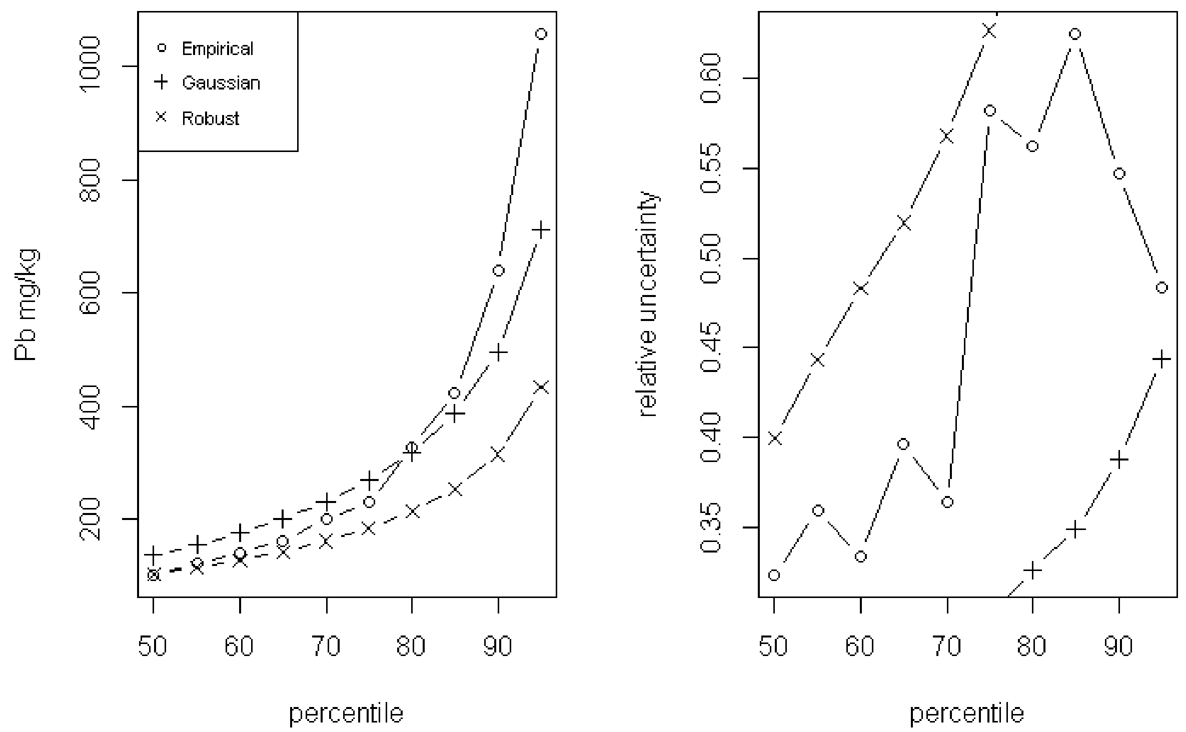


Figure 131: Comparison of empirical, Gaussian and Robust percentiles and relative uncertainty for Pb in the Urban 2 Domain

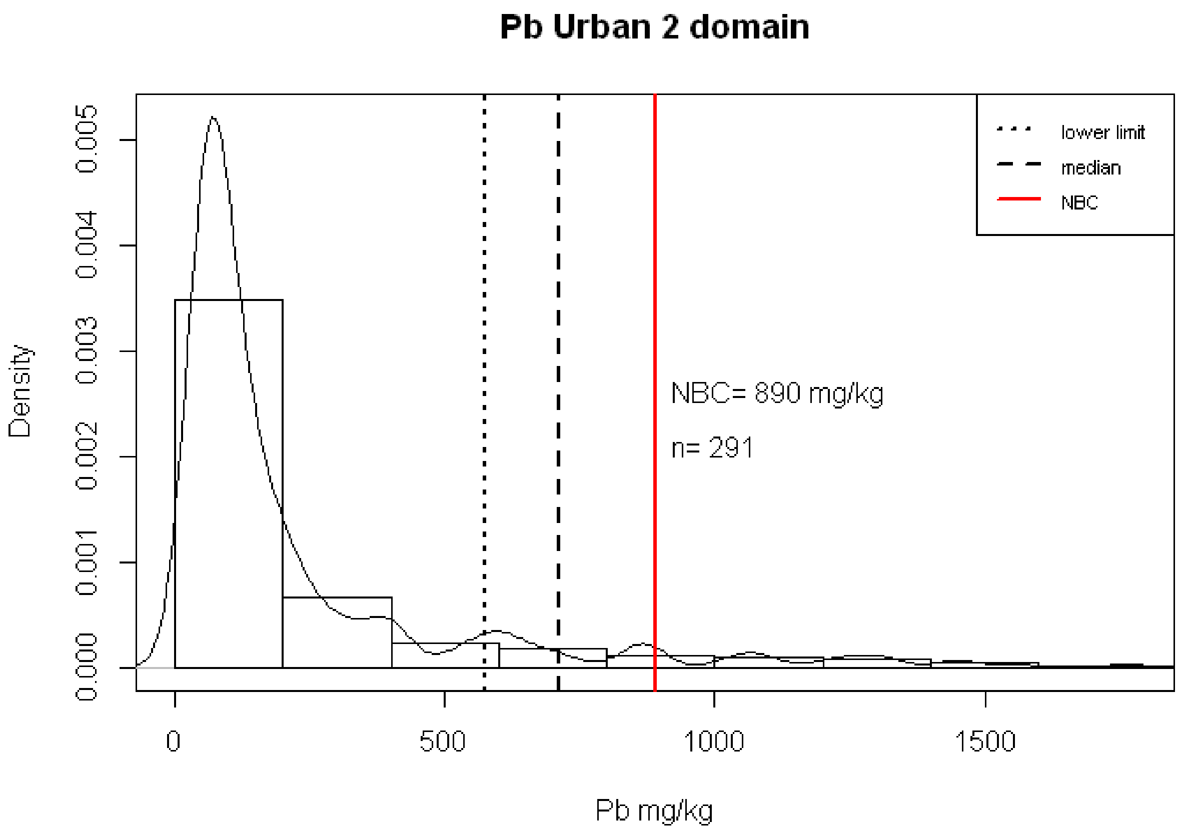


Figure 132: Summary density plot and histogram of the distribution for Pb in the Urban 2 Domain showing an example NBC (n = number of samples)

Percentile	Emp L	Empirical	Emp H	P L	Parametric	P H	R L	Robust	R H
50	90	102	123	**122**	**137**	**153**	90	102	131
55	101	122	145	**138**	**155**	**175**	99	114	149
60	116	142	164	**156**	**176**	**199**	109	128	171
65	137	162	202	**177**	**201**	**229**	120	144	195
70	159	201	232	**201**	**231**	**266**	134	162	226
75	193	231	327	**231**	**269**	**312**	150	185	266
80	232	327	416	**270**	**318**	**373**	170	214	317
85	344	425	609	**325**	**387**	**460**	197	254	389
90	523	639	873	**409**	**495**	**601**	236	315	501
95	781	1058	1293	**573**	**713**	**890**	306	434	730

Low (L) and High (H) values represent the 95% confidence intervals around the median. Shaded/bold values indicate data used to calculate NBC

Table 50: Empirical (Emp), parametric Gaussian (P) and Robust Gaussian (R) percentile values for Pb (mg/kg) in the Urban 2 Domain

8.4.6 Lead NBC summary

Summary of all Pb domain NBCs and associated information is given in Table 51.

Percentile	Principal Domain n =966 Area = 18200 km^2 Transformation: Box-Cox			Mineralisation 1 Domain n = 15 Area = 200 km^2 Transformation: n/a			Mineralisation 2 Domain n = 61 Area = 1100 km^2 Transformation: log_e		
	lower	middle	upper	lower	middle	upper	lower	middle	upper
50	57.0	58.6	60.3	nc	nc	nc	60	75	90
55	60.5	62.3	64.2	nc	nc	nc	63	80	97
60	64.4	66.5	68.8	nc	nc	nc	66	86	100
65	68.9	71.4	74.1	nc	nc	nc	70	92	110
70	74.3	77.4	80.6	nc	nc	nc	74	99	120
75	81.1	84.9	88.8	nc	nc	nc	79	110	140
80	90.0	94.9	99.9	nc	nc	nc	84	120	150
85	103	110	117	nc	nc	nc	90	130	180
90	124	135	147	nc	nc	nc	99	150	210
95	176	199	229	nc	nc	nc	110	180	280

Percentile	Urban 1 Domain n = 342 Area = 1200 km^2 Transformation: log_e			Urban 2 Domain n = 291 Area = 500 km^2 Transformation: log_e		
	lower	middle	upper	lower	middle	upper
50	200	220	245	120	140	150
55	225	248	278	140	150	170
60	252	281	316	160	180	200
65	285	319	361	180	200	230
70	324	364	416	200	230	270
75	371	421	485	230	270	310
80	432	494	574	270	320	370
85	516	595	700	330	390	460
90	644	753	899	410	490	600
95	890	1068	1299	570	710	890

n = number of samples used in calculation. nc = not calculated (n = <30)

Table 51: Lead (mg/kg) NBC summary table

9 Concluding remarks

1. Normal Background Concentrations have been calculated for Welsh soil using the methodology developed and applied for English soil (Ander *et al*. 2011, 2012, 2013; Cave *et al*. 2011). The results are summarised in Table 52.
2. High density G-BASE systematically collected topsoil data are not available for Wales, with the exception of the Swansea and Cardiff urban areas. This means the Welsh soil NBCs are based on a much smaller data set principally derived from the low density sampling of the NSI. Therefore, the Welsh NBCs will be associated with much greater uncertainty and there are knowledge gaps in the domains identified as being characteristically higher in many of the contaminants explored (*e.g.* urban and mineralisation domains).
3. The general controls on the concentrations and spatial distribution of contaminants in soils are similar to those seen in England. However, different underlying parent materials and styles of mineralisation (*e.g.* the absence of significant areas of As associated with ironstones as seen in England) give rise to differently defined domains.
4. Urban and industrial development is dominated in Wales by the South Wales Coalfield and the associated industries and urban areas. Although the G-BASE project has high density soil sampling for the Swansea area, results show that this area is associated with one of the most significant legacies of industrial contamination and so the Swansea results cannot be used as being representative of other urban areas in Wales.

Contaminant	Domain	NBC (mg/kg)	Samples (n)	Area (km^2)	Area (%)
Arsenic	Principal	**36**	1270	18900	89
	Mineralisation	**67**	63	1100	5
	Urban 1	**250**	342	1200	6
BaP[#]	Principal	**0.5**	71	20600	97
	Urban	**3.6**	32	600	3
Cadmium	Principal	**1.4**	681	18700	88
	Mineralisation 1	*n.d.**	15	200	1
	Mineralisation 2	**2.2**	57	1100	5
	Urban 1	**6.2**	45	1200	6
Copper	Principal	**43**	966	18400	87
	Mineralisation	**96**	76	1100	5
	Urban 1	**550**	342	1200	6
	Urban 2	**170**	291	500	2
Mercury	Principal	**0.25**	104	20600	97
	Urban	*n.d.**	7	600	3
Nickel	Principal	**40**	1327	19800	94
	Basic	*n.d.**	6	200	<1
	Urban 1	**120**	342	1200	6
Lead	Principal	**230**	966	18200	86
	Mineralisation 1	*n.d.**	15	200	1
	Mineralisation 2	**280**	61	1100	5
	Urban 1	**1300**	342	1200	6
	Urban 2	**890**	291	500	2

[#]Benzo[*a*]pyrene (BaP) is not explored in this report. Results are based on British soil and are as determined in the English soil data exploration. Welsh BaP domains based on those identified for Hg.
*fewer than 30 samples, no NBC determined. Area (%) reported to an integer.

Table 52: Summary of NBCs for each contaminant by domain

References

Alloway, B J, and Davies, B E. 1971. Trace element content of soils affected by base metal mining in Wales. *Geoderma*, 5, 197-208.

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 (unpublished)*, CR/11/145. Available on-line at: http://nora.nerc.ac.uk/19958 last accessed 27 February 2013.

Ander, E L, Cave, M R, Johnson, C C, and Palumbo-Roe, B. 2012. Normal background concentrations of contaminants in the soils of England. Results of the data exploration for Cd, Cu, Hg and Ni. *British Geological Survey Commissioned Report (unpublished)*, CR/12/041. Available on-line at: http://nora.nerc.ac.uk/19952 last accessed 27 February 2013.

Ander, E L, Johnson, C C, Cave, M R, Palumbo-Roe, B, Nathanail, C P and Lark, R M. 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. *Science of the Total Environment*, 604-618. doi 10.1016/j.scitotenv.2013.03.005.

Archer, F C, and Hodgson, I H. 1987. Total and extractable trace element contents of soils in England and Wales. *Journal of Soil Science*, 38, 421-432.

Avery, B W. 1990. *Soils of the British Isles*. Wallingford, Oxfordshire, CAB International.

Barraclough, D. 2007. UK Soil and Herbage Pollutant Survey. Introduction and Summary. Environment Agency (EA), UK.

Bradley, R I, Rudeforth, C C, and Wilkins, C. 1978. Distribution of some chemical elements in soils of north-west Pembrokeshire. *Journal of Soil Science*, 29, 258-270.

British Geological Survey. 2000. *Regional geochemistry of Wales and part of west-central England: stream sediment and soil*. (Keyworth, Nottingham: British Geological Survey.)

Brown, S E. 2001. Geochemical baseline data for the urban area of Cardiff. *British Geological Survey Open Report (unpublished)*, IR/01/102. Available on-line at: http://nora.nerc.ac.uk/7014/ last accessed 27 February 2013.

Cave, M R, Johnson, C C, Johnson, C C, and Palumbo-Roe, B. 2012. Methodology for the determination of normal background contaminant concentrations in English soils. *British Geological Survey Commissioned Report (unpublished)*, CR/12/003. Available on-line at: http://nora.nerc.ac.uk/19959 last accessed 27 February 2013.

Davies, B E. 1976. Mercury content of soils in western Britain with special reference to contamination from base metal mining. *Geoderma*, Vol. 16, 183-192.

Davies, B E. 1985. *Baseline survey of metals in Welsh soils*. Proceedings of the First International Symposium on Geochemistry and Health. (Science Reviews Ltd.) ISBN 0-905927-57-5

Davies, B E. 1997. Heavy metal contaminated soils in an old industrial area of Wales, Great Britain: Source identification through statistical data interpretation. *Water Air and Soil Pollution*, 94, 85-98.

Davies, B E, and Lewin, J. 1974. Chronosequences in alluvial soils with special reference to historic lead pollution in Cardiganshire, Wales. *Environmental Pollution*, 6, 49-57.

Davies, B E, and Paveley, C F. 1988. *Baseline trace metal survey of Welsh soils with special reference to lead*. Geochemistry and Health. (Science Reviews Ltd.) ISBN 0-905927-57-5

Davies, B E, and Roberts, L J. 1975. Heavy metals in soils and radish in a mineralised limestone area of Wales, Great Britain. *Science of The Total Environment*, 4, 249-261.

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. 186-206, Chapter 13. In: *Mapping the Chemical Environment of Urban Areas* C C Johnson, A Demetriades, J Locutura and R T Ottesen (editors). John Wiley & Sons Ltd., Chichester, UK.

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*, Vol. 5(4), 325-336.

Fuge, R, Paveley, C F, and Holdham, M T. 1989. Heavy-metal contamination in the Tanat Valley, North-Wales. *Environmental Geochemistry and Health*, 11, 127-135.

Griffith, J J. 1918. Influence of mines upon land and livestock in Cardiganshire. *Journal of Agricultural Science*, 9, 365-395.

Hughes, S and Reynolds, P. 1992. A guide to the industrial archaeology of the Swansea region. Royal Commission on the Ancient and Historical Monuments of Wales, Aberystwyth. January 1992. 52pp. ISBN: 9781871184013/

Johnson, C C, Ander, E L, Cave, M R, and Palumbo-Roe, B. 2012. Normal background concentrations (NBCs) of contaminants in English soils: Final project report. *British Geological Survey Commissioned Report (unpublished)*, CR/12/035. Available on-line at: http://nora.nerc.ac.uk/19946 last accessed 27 February 2013.

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 last accessed 27 February 2013.

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

Morley, S E, and Ferguson, A J. 2001. Geochemical baseline data for the urban area of Swansea. *British Geological Survey Open Report (unpublished)*, IR/01/036. Available on-line at : http://nora.nerc.ac.uk/7019/ last accessed 27 February 2013.

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. Defra (formerly MAFF) Soil Programme Technical Report Project SP0124.

Office for National Statistics, 2006. National Land Use Database: Land Use and Land Cover Classification. Available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/11493/144275.pdf last accessed 27 February 2013.

Ordnance Survey, 2011. Ordnance Survey Strategi. User guide and technical specification. Version 8.3. 12/2011. Available online at: https://www.ordnancesurvey.co.uk/oswebsite/docs/user-guides/strategi-user-guide.pdf last accessed 27 February 2013.

Ottesen, R T, Birke, M, Finne, T E, Gosar, M, Locutura, J, Reimann, C, and Tarvainen, T. 2013. Mercury in European agricultural and grazing land soils. *Applied Geochemistry*, Published online http://dx.doi.org/10.1016/j.apgeochem.2012.12.013 last accessed 27 February 2013.

Ove Arup. 1990. Mining Instability in Britain. Department of Environment contract. Hardcopy maps digitised and adapted by the British Geological Survey.

Rawlins, B G, McGrath, S P, Scheib, A J, Cave, M R, Breward, N, Lister, T R, Ingham, M, Gowing, C J B and Carter, S. 2012. The Advanced Soil Geochemical Atlas of England and Wales. British Geological Survey, Keyworth, Nottingham. Published as an electronic book. Available on-line at: www.bgs.ac.uk/gbase/advsoilatlasEW.html last accessed 27 February 2013.

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

Reimann, C, Flem, B, Fabian, K, Birke, M, Ladenberger, A, Négrel, P, Demetriades, A, and Hoogewerff, J. 2012. Lead and lead isotopes in agricultural soils of Europe – The continental perspective. *Applied Geochemistry*, 27, 532-542.

Rice, R and Sharp, G J. 1976. Copper mineralisation in the forest of Coed-y-Brenin, North Wales. *Transactions of the Institution of Mining and Metallurgy, (Section B: Applied Earth Science*), 85, B1−B13.

Rodrigues, S, Pereira, M E, Duarte, A C, Ajmone-Marsan, F, Davidson, C M, Grcman, H, Hossack, I, Hursthouse, A S, Ljung, K, Martini, C, Otabbong, E, Reinoso, R, Ruiz-Cortes, E, Urquhart, G J, and Vrscaj, B. 2006. Mercury in urban soils: A comparison of local spatial variability in six European cities. *Science of The Total Environment*, 368, 926-936.

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, 526 pp.

Thornton, I. 1991. Metal contamination of soils in urban areas. 47-75 In: *Soils in the urban environment*. Bullock, P, and Gregory, P J (editors). Oxford: Blackwell Scientific Publications.

Waters, C N, Price, S J, Davies, J, Tye, A M, Brown, S E and Schofield, D I. 2005. Urban geology of Swansea-Neath-Port Talbot. Chapter 1, 7-22. In: *Urban Geology in Wales: 2*. Bassett, M G, Deisler, V K and Nichol, D. (editors) National Museum of Wales Geological Series No 24, November 2005.

Welsh Government. 2012. Contaminated Land Statutory Guidance -2012. Welsh Government, Document Number WG15450. 67pp. Available on line at: http://wales.gov.uk/docs/desh/publications/120417contaminatedlandguideen.pdf last accessed 27 February 2013.

Appendix 1 : Listing of Project Reports from Defra Project SP1008

The following is a list of reports produced by Project SP1008.

English Soil (Reports available as pdf files linked from the Defra Project SP1008 and BGS project web sites)

Reports:

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

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

Cave, M R, Johnson, C C, Ander, E L, and Palumbo-Roe, B. 2012. Methodology for the determination of normal background contaminant concentrations in English soils. *British Geological Survey Commissioned Report*, CR/12/003. 42pp.

Johnson, C C, Ander, E L, Cave, M R, 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.

Technical Guidance:

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

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

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Lead. Technical Guidance Sheet No. TGS02, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Lead – supplementary information. Technical Guidance Sheet No. TGS02s, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Copper. Technical Guidance Sheet No. TGS03, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Copper – supplementary information. Technical Guidance Sheet No. TGS03s, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Benzo(a)pyrene. Technical Guidance Sheet No. TGS04, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Benzo(a)pyrene – supplementary information. Technical Guidance Sheet No. TGS04s, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Nickel. Technical Guidance Sheet No. TGS05, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Nickel – supplementary information. Technical Guidance Sheet No. TGS05s, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Cadmium. Technical Guidance Sheet No. TGS06, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Cadmium – supplementary information. Technical Guidance Sheet No. TGS06s, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Mercury. Technical Guidance Sheet No. TGS07, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Defra. 2012. Technical Guidance Sheet on normal levels of contaminants in English soils: Mercury – supplementary information. Technical Guidance Sheet No. TGS07s, July 2012. *Department for Environment, Food and Rural Affairs (Defra)*, Soils R&D Project SP1008.

Welsh Soil (Reports will be made available as pdf files linked from the Defra Project SP1008 and BGS project web sites)

Reports:

Ander, E L, Cave, M R and Johnson, C C. 2013. Normal background concentrations of contaminants in the soils of Wales. Exploratory data analysis and statistical methods. *British Geological Survey Commissioned Report*, CR/12/107. 144pp.

Technical Guidance:

Defra. 2013. Technical Guidance on normal levels of contaminants in Welsh soil: Arsenic. *British Geological Survey (Keyworth, Nottingham) and Defra (London)*, Soils R&D Project SP1008, January 2013.

Defra. 2013. Technical Guidance on normal levels of contaminants in Welsh soil: BaP. *British Geological Survey (Keyworth, Nottingham) and Defra (London)*, Soils R&D Project SP1008, January 2013.

Defra. 2013. Technical Guidance on normal levels of contaminants in Welsh soil: Cadmium. *British Geological Survey (Keyworth, Nottingham) and Defra (London)*, Soils R&D Project SP1008, January 2013.

Defra. 2013. Technical Guidance on normal levels of contaminants in Welsh soil: Copper. *British Geological Survey (Keyworth, Nottingham) and Defra (London)*, Soils R&D Project SP1008, January 2013.

Defra. 2013. Technical Guidance on normal levels of contaminants in Welsh soil: Lead. *British Geological Survey (Keyworth, Nottingham) and Defra (London)*, Soils R&D Project SP1008, January 2013.

Defra. 2013. Technical Guidance on normal levels of contaminants in Welsh soil: Mercury. *British Geological Survey (Keyworth, Nottingham) and Defra (London)*, Soils R&D Project SP1008, January 2013.

Defra. 2013. Technical Guidance on normal levels of contaminants in Welsh soil: Nickel. *British Geological Survey (Keyworth, Nottingham) and Defra (London)*, Soils R&D Project SP1008, January 2013.