Determination of stream sediment background concentrations in mineralised catchments impacted by mining using Tellus data from Northern Ireland: Final Project Report

Applied and Medical Geochemistry Team

COMMISSIONED REPORT CR/14/021

Tellus Border is a €5 million cross-border project to map the environment and natural resources in the border region of Ireland and continue the analysis of data in the border counties of Northern Ireland. It is a joint initiative between the Geological Survey of Northern Ireland, the Geological Survey of Ireland, Queen's University, Belfast and Dundalk Institute of Technology.

This research is supported by the EU INTERREG IVA-funded Tellus Border project managed by the Special EU Programmes Body (SEUPB) .

The views and opinions expressed in this research report do not necessarily reflect those of the European Commission or SEUPB.

BRITISH GEOLOGICAL SURVEY

Applied and Medical Geochemistry Team

COMMISSIONED REPORT CR/14/021

Determination of stream sediment background concentrations in mineralised catchments impacted by mining using Tellus data from Northern Ireland: Final Project Report

B Palumbo-Roe, EL Ander, MR Cave and CC Johnson

Bibliographical reference

PALUMBO-ROE, B, ANDER, EL, CAVE, MR & JOHNSON, CC. 2013. Determination of stream sediment background concentrations in mineralised catchments impacted by mining using Tellus data from Northern Ireland: Final Project Report. *British Geological Survey Commissioned Report,* CR/14/021, 56pp.

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 report is a required deliverable for a BGS research project “Determination of stream sediment background concentrations in mineralised catchments impacted by mining using Tellus data from Northern Ireland” commissioned by the Tellus Border project.

Acknowledgement

This research is supported by the European Union INTERREG IVA-funded Tellus Border project. Tellus Border is a €5 million cross-border project to map the environment and natural resources in the border region of Ireland and continue the analysis of data in the border counties of Northern Ireland. It is a joint initiative between the Geological Survey of Northern Ireland (GSNI), the Geological Survey of Ireland, Queen’s University, Belfast and Dundalk Institute of Technology.

GSNI is gratefully acknowledged for providing some of the spatial datasets.

Contents

FIGURES

TABLES

Summary

Background metal(loids) concentrations, intended as concentrations of naturally occurring substances rather than anthropogenic, are more often integrated in the assessment of water and sediment quality. This approach allows that ecosystems may be adapted or acclimatised to certain concentrations of metals in surface water and sediments as a result of their natural abundance. Background values of metal(loids) have long been recognised to be higher in mineralised catchments than those in unmineralised, and this is in fact the same as the central precept of geochemical exploration for economic ore deposits. From the environmental perspective, these mineralised zones should be considered as a separate baseline unit from that of the unmineralised formation.

Information on the baseline conditions of catchments prior to mining is needed to better understand what restoration goals are achievable in mining impacted catchments. The geochemical baseline data also provide a reference point against which changes can be measured and can be used both by industry and regulators in future mine applications.

In this project an approach for deriving pre-mining baseline sediment concentrations using systematically collected survey geochemical data is demonstrated using the mineralised area associated with the Ordovician-Silurian rocks in southern Co. Armagh in Northern Ireland as study area. The Tellus geochemical survey data for sediments were used for this scope.

International literature has usefully provided methodologies and examples of deriving 'background' concentrations in mineralised catchments. Statistical methods in use to distinguish between anomalous and background concentrations in geochemical exploration of mineral deposits all converge on various methods of discriminating outliers and making estimates of central tendency, spread and identification of upper thresholds of background.

The statistical method used in this project is the method of Sinclair (1976a) and applied using the 'PROBPLOT' code (Stanley, 1987), reproduced in an 'R' script environment. This method chooses threshold values between anomalous and background geochemical data, based on partitioning a cumulative probability plot of the data.

Data analysis has primarily focused on elements for which there are sediment quality standards derived in other jurisdictions, which may be adopted in the UK regulatory framework in future.

Probability distribution plots of stream sediment lead (Pb) zinc (Zn), arsenic (As), chromium (Cr) and nickel (Ni) concentrations have been partitioned in the respective contributing populations and population statistics derived (mean and standard deviation). Interpretation of the significance of the resulting groupings of data and understanding different background populations has then been achieved through analysis of the spatial distribution of the groups in a GIS framework.

Where data exceed environmental quality standards, these populations can assist in identifying where natural background concentrations (due to mineralogical variations in the catchment

geology) may contribute to the exceedance. This is designed to aid the decision-making process in relation to why quality standards may have failed, or if there is any merit in 'remediation' of a natural ecosystem. Separation of the more widespread, potentially natural, high concentrations from the data populations which reflect very high concentrations (more likely to arise from anthropogenic sources) could also help in targeting key sites for further investigation.

1 Introduction

This report is a required deliverable for a BGS commissioned research project to investigate the "Determination of stream sediment background concentrations in mineralised catchments impacted by mining using Tellus data from Northern Ireland" as detailed in the Project Plan (Tellus Border project reference 10761).

The naturally elevated background concentrations of metal(loids) in stream sediments and water from mineralised catchment can be retrospectively discriminated from the impacts of ore extraction and processing by using data from systematically collected samples.

The scientific objectives of this project are:

- To define sediment baselines in mineralised catchments impacted by mining and to provide a tool for better assessment of sediment quality in these catchment types;

- To test the applicability of a statistical methodology to split polymodal data by comparison with available spatial information, in order to derive baseline data;

- To produce a peer-reviewed scientific publication.

The Project Gantt chart is shown in Figure 1 and this forms the outline for this end of project report.

The purpose of this report is to provide the evidence for completion of tasks detailed in Figure 1. This is summarised in Table 1.

Week commencing	6th	13th	20th	27th	3rd	10th	17th	24th	1st	8th	15th	22nd	29Jul/ Aug	5th	12th	19th	26th	2nd	9th	16th	23th	30Sep /Oct	* #
Month	May				June				July				Aug					Sep					Oct/Nov
Task																							
1. Project start-up																							
2. Literature search																							
3. Data exploration																							
4. Mid project review																							
5. Statistical outputs																							
6. Preparation of publication																							
7. Delivery of outputs																							
Milestones														A									B
Outputs	1								2					2,3				2				2	4, 5 & 6

* 24th October Final Project Conference # "End of Project" report by 1 Nov 2013

Milestones	Outputs	
A Completion of exploration phase	1	Agreed Project Plan
B Completion of output preparation	2	Monthly report (5^{th} each month)
	3	Mid Project report
	4	End project report
	5	Draft publication ready for submission
	6	Contribution for Final Project conference

Figure 1: Project Gantt chart.

Table 1: Summary status of Project items

Item	Status	Comment/Date of completion
TASKS: 1. Project Start-up	COMPLETE	Videoconference 2nd May 2013
2. Literature search	COMPLETE	1st November 2013
3. Data exploration	COMPLETE	
4. Mid project review	COMPLETE	Teleconference 8th August 2013
5. Statistical Output	COMPLETE	
6/7. Preparation of publication and delivery of outputs	COMPLETE	1st November 2013
OUTPUTS: 1. Agreed Project Plan	COMPLETE	Approved 03/05/2013 – email from Marie Cowan
2. Monthly reports	COMPLETE	June report submitted 5th July 2013. July report submitted 5th August 2013.
3. Mid-project report	COMPLETE	8th August 2013
4. End-project report	COMPLETE	1th November 2013
5. Draft publication	ON GOING	Outline publication produced in July/August after literature search and initial statistical outputs prepared.
6. Contribution to Tellus Border Conference	COMPLETE	24th October 2013
MILESTONES: A. Completion of exploration phase	COMPLETE	Completion coincident with mid-project review
B. Delivery of all outputs	COMPLETE	1st November 2013

2 Literature Search

An important part of the initial phase of the project has been to gather together literature relevant to the sediment quality standards (sections 2.2 and 2.3), the approaches to derive baseline element concentrations, with a specific focus on mineralised areas (section 2.4), and mineralisation and mining in Northern Ireland (section 2.5). References have been entered into an EndNote library and most references are available as pdf files.

2.1 INFLUENCE OF CONTAMINATED SEDIMENTS FOR ATTAINMENT OF EUROPEAN WATER FRAMEWORK DIRECTIVE AND GUIDANCE ON SEDIMENT QUALITY STANDARDS

Over the last decades there has been a strong focus on tackling water pollution from point sources such as industrial discharge, treatment works, mine waters, while little attention has been paid on the potential ecological consequences of *in situ*-contaminated sediments for attainment of the European Water Framework Directive (WFD, 2000/60/EC) objectives of “good chemical and ecological status” of surface waters (EC, 2000). Yet, sediments can have an important role as a source of contaminants to water, or as a sink from the water column; they can be important in mediating the exposure of benthic organism to contaminants and governing ecological quality in mining impacted rivers. There is concern that they may become a secondary source of pollution where water quality is improving (SedNet, 2004; Butler, 2009).

The assessment of "pressures" in the form of point source and diffuse pollution to the status of surface water, as required by the WFD, incorporates compliance with environment quality standards (EQS), set out in the Dangerous Substances Directive (76/464/EEC and Daughter Directives) (EC, 1976).

The methodology to develop these guidelines is described in a technical Guidance document (TGD) for Deriving Environmental Quality Standards (EC, 2011). Unlike chemical thresholds to assess water quality, the complex processes affecting availability of contaminants in sediments requires the use of multiple lines of evidence to generate sediment standards relevant for hydrophobic substances and some metals to protect benthic (sediment welling) species (EC, 2011). Data used for the derivation of EQS for sediment may include ecotoxicity data from experiments with benthic organisms, aquatic toxicity tests in conjunction with equilibrium partitioning and field/mesocosm studies. The above mentioned EU guidance (EC, 2011) provides further suggestions to policy makers, Member States or Basin Authorities on the use of a tiered assessment framework, in which the use of sediment standards is only one of a number of lines of evidence to decide where management measures are warranted.

Currently, in the UK there are no statutory Environmental Quality Standards (EQS) for sediment quality. The UK Technical Advisory Group on the Water Framework Directive (UKTAG) does not recommend setting mandatory standards in sediments (UKTAC, 2012). This is because the high uncertainty in deriving sediment predicted no-effects concentrations (PNECs) due to lack of sediment toxicity data for many substances and concerns on the suitability of the equilibrium partitioning approach to supplement the lack of sediment toxicity data (UKTAC, 2012). Furthermore, it is recognised there are difficulties in using measurements on sediments to provide the basis for environmental control regimes, given the high spatial variability of monitoring data. Whilst statutory sediment quality standards are not foreseen, there may be scope to develop guideline values for sediments as opposed to mandatory or statutory standards. Where a PNEC for sediments has been developed, the UKTAG recommends that it can be used as a guideline. These guideline values might be part of a wider process of assessment, for example as a trigger for further evidence gathering to support a case for investigation and regulatory action.

An assessment of metal mining-contaminated river sediments in England and Wales by Hudson-Edwards et al. (2008), commissioned by the Environment Agency, reports on the development within the EA of interim sediment guideline values that could be used to trigger further investigation. The guidelines are based on the approach of Environment Canada, which considers a Toxic Effect Level (TEL, the concentration below which sediment associated contaminants are not considered to represent significant hazards to aquatic organisms) and a Predicted Effect Level (PEL, the concentration representing the lower limit of the range of concentrations associated with adverse biological effects) (Table 2).

Re-suspension of contaminated sediments during floods can also contaminate floodplain soils used for agriculture and may represent a risk to grazing livestock. It is in this context that standards set to protect livestock have been used to score contaminated sediment hazard in the Historic Mine Sites risk-based inventory carried out in Ireland (EPA & GSI, 2009). Reference guidelines were provided by the Central Veterinary Research laboratory (CVRL) (Ireland), assuming sediment representing 10% of diet with no consumable herbage growth and 100% metal bioavailability (Table 3).

Further detail on the derivation and application of environmental quality standards is outside the scope of this report.

Table 2: Draft sediment quality criteria (TEL, PEL) for England and Wales (Hudson-Edwards et al., 2008)

Element	TEL mg/kg	PEL mg/kg
As	5.9	17
Cd	0.596	3.53
Cr	37.3	90
Cu	36.7	197
Pb	35	91.3
Ni	18	35.9
Zn	123	315

Table 3: Stream sediment guidelines used for the project "Historic Mine Sites – Inventory and Risk Classification" (EPA& GSI, 2009)

Element	Guideline Value mg/kg (dry matter)	Source
Ag	1000	CVRL
As	300	CVRL
Ba	1000	CVRL
Cd	100	CVRL
Cr III	1000	CVRL
Cu	100	CVRL
Fe	10000	CVRL
Hg	5	CVRL
Mn	5000	CVRL
Ni	1000	CVRL
Pb	1000	CVRL
Sb	1000	CVRL
Se	12	CVRL
Sn	1000	CVRL
V	500	CVRL
Zn	5000	CVRL

2.2 METAL BACKGROUND REFERENCE CONCENTRATIONS

The Technical Guidance document (TGD) for Deriving Environmental Quality Standards (EQSs) (EC, 2011) considers "background concentrations" of the substance in question to which biota may be acclimatised, the so called "added risk approach", as part of the process to derive EQSs for water and sediments. The 'added risk' approach allows that ecosystems may be adapted or acclimatised to certain concentrations of metals in surface waters and sediments as a result of their natural abundance, and that these background concentrations can therefore be taken into account when assessing risk against water and sediment quality standards (e.g. Ander and Casper, 2008). If sites fail the environmental quality standards, consideration of the natural background concentration may be undertaken to further assess compliance and prior to any expensive or time-consuming remediation (UKTAG, 2012).

2.3 METAL BACKGROUND CONCENTRATIONS IN MINERALISED AREAS

Background values of metal(loids) have long been recognised to be higher in mineralised catchments than those in unmineralised areas (e.g. Figure 2) (Rose et al., 1979), and this is in fact the central precept of geochemical exploration for economic ore deposits, where this area of science was developed. From the environmental perspective, these mineralised zones should be considered as a separate baseline unit from that of the unmineralised formation, for those elements which are elevated within the economic or gangue mineralisation. Indeed, to try and restore any part of the catchment ecosystem to metal(loid) concentrations associated with the unmineralised zone could be considered not to be 'technically feasible', 'scientifically reasonable' or 'economically achievable' (Runnels et al., 1992). Runnells et al. (1992) draw attention to sites studied in the USA which have not been affected by mining, and have stream water pH <4, and high total dissolved Fe (up to 17 mg/L) and Zn (up to 0.94 mg/L), as well as studies on unmined mineralised sites from Canada with up to 16 mg/L dissolved Zn. It might be reasonably anticipated that associated stream sediments would also be markedly enhanced in concentration of some trace elements.

If mineralisation is sufficient to be deemed an economic ore deposit, then subsequent working of the ore body can lead to wider dispersion of metals to river systems. This can be through direct discharge of effluents from mine adits, of rock wastes and mine tailings and from associated smelting activities (Byrne et al., 2013). Ore processing may also have taken place using diverted water culverts for 'washing' the ore and disposal and consequent dispersal of the mine tailings into the nearby watercourse as was common practise in the past. All of these processes will be expected to lead to further enrichment in metal(loid) concentrations within the stream network, and along the floodplain of these streams and the main rivers (Hudson-Edwards et al., 2008; Bird et al., 2010) and can result in a stream sediment anomaly just as an *in-situ* ore vein can (Figure 2).

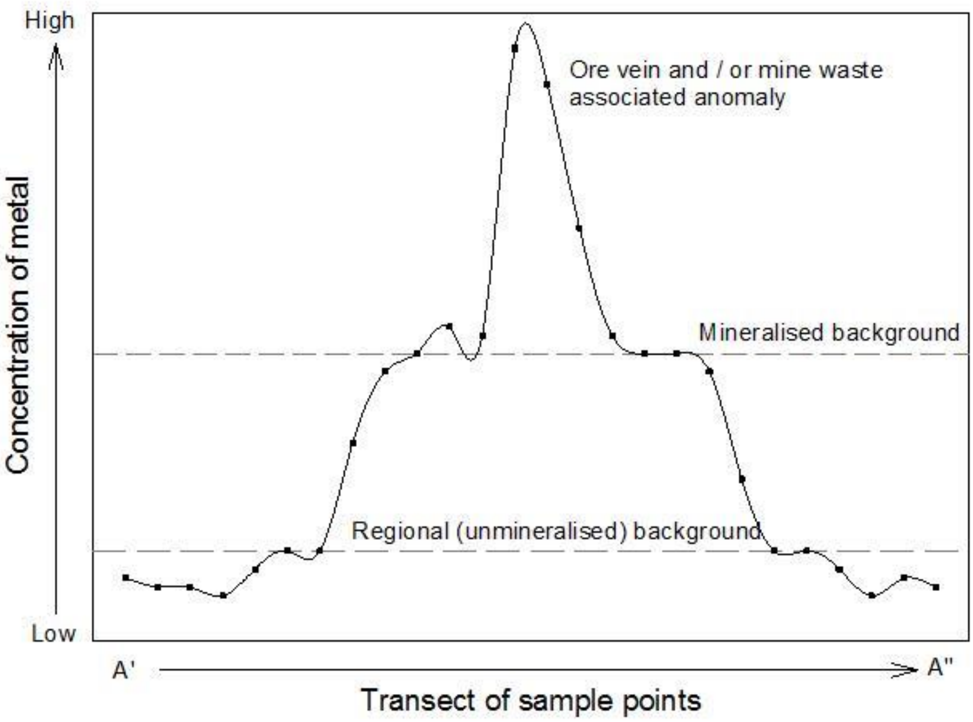


Figure 2: Idealised illustration of the naturally elevated baseline occurring over an area of mineralisation, and an anomaly associated with a mineral vein or mine wastes.

Anomalies can occur not only due to the inherent chemical/mineralogical properties of the clastic material weathered into the stream sediments of a catchment, or as a result of secondary processes such as high concentrations of Mn or Fe precipitation in a stream course. Thus, in addition to examining total concentrations of a potentially toxic element such as Pb or Zn, the relative concentrations of Mn and/or Fe can be used to assess the predominant process, as demonstrated by Butt and Nichol (1979).

2.4 METHODOLOGY TO DERIVE BASELINE ELEMENT CONCENTRATIONS IN MINERALISED AREAS

The most obvious and unequivocal approach to comparing current metal(loid) loadings in catchments with that of the pre-mining geochemical landscape is to use pre-mining data. This, however, is not always possible due to the duration over which mining has taken place in some of the orefields in UK and worldwide, and therefore the lack of such data in any environmental compartment (waters, sediments, soils, biota).

One advantage of geochemical baseline mapping data, is that unmined catchments will have been sampled at the same density as mined catchments in such a study. This means that although all data are from the same (or similar) points in time, they may spatially reflect a variety of mined, mineralised and unmineralised environments over the same bedrock and open up the possibility of using a spatial, rather than temporal, comparison. This can also be thought of as a more extensive version of using an ‘upstream’ sample in a single catchment study; the greater number of data available from geochemical mapping give more information on univariate and multivariate data populations.

Various approaches have been used to establish the baseline values in mining impacted catchments, including the use of historical records, geochemical analogues, and geochemical modelling (Runnels et al., 1992; Alpers, 2000; Runkel et al., 2007).

Statistical methods have long been in use to distinguish between anomalous and background concentrations in geochemical exploration of economic metalliferous mineral deposits. Some of the statistical approaches have been described by Matschullat et al. (2000), and all converge on various methods of discriminating outliers and making estimates of central tendency, spread and identification of upper thresholds of background.

The simplest statistical approaches are to use values that are multiples of the [mean ($\overline{x}$) + standard deviation (σ)], that will represent specific upper percentiles of the dataset, if conditions of Gaussian distributions are met.

Stream sediment data published in the Provisional Geochemical Atlas for Northern Ireland (Applied Geochemistry Research Group, 1973) was used by Butt and Nichol (1979) to calculate geometric mean, "threshold" ($\overline{x} + 2\sigma$) and "probably anomalous" ($\overline{x} + 3\sigma$) values for several parent materials. The 'Silurian-Ordovician (Keady)' data is shown below (Table 4), along with the values for the entire dataset. This work noted the presence of 'clastic' Pb-Zn, probably from local comminution and physical transport by glacial and subsequent stream flow processes. They postulate that where the Pb/Zn ratio lowers, this is as a result of the greater solubility of Zn in relation to Pb, and reflects zones of hydromorphic dispersion/accumulation on Mn oxides or in organic-rich sediments (Butt and Nichol, 1979).

Table 4: Background stream sediment statistics for the Ordovician-Silurian of the Keady area (n = 42) from Butt and Nichol (1979)

	Geometric mean	Threshold	Probably anomalous
	mg/kg		
Cu	20	35	45
Pb	60	140	180
Zn	135	260	320

n = 42

2.5 MINERALISATION AND MINING IN CO. ARMAGH, NORTHERN IRELAND

The mineralisation associated with the Ordovician-Silurian turbidites in southern Co. Armagh has been chosen as the focus for this study. This is because the known occurrence of mineralisation that has previously been mined in this region increases the likelihood of finding catchment reaches which are affected and unaffected by mining in the mineralised zone and also gives a large area of unmineralised 'background' for potential comparison.

The South Armagh–Monaghan mining district was a relatively major area of mineral extraction. In S Armagh 57 shafts and adits are recorded by GSNI, with significant workings. Historical production of lead in Co. Armagh is in the South Armagh-Monaghan Mining District, centred on the town of Keady in south Armagh. Historical mines described by Cole (1922) are Derrynoose Mine, College Mine, Clay Mine, Carrickgallogy Mine, and Creggan Mine.

Mineralisation is principally galena, sphalerite, pyrite, chalcopyrite with calcite and barytes gangue (Mitchell, 2004). The abandoned mine workings are sufficiently old to pre-date any waste control legislation, and the most modern mineral extraction methods, with Derrynoose Lead Mine [H796 316] having been abandoned in 1842 (Mitchell, 2004). Outside of the S Armagh area, but still within the same succession, there was lead extraction at Conlig-Whitespots. This produced an estimated 13,500 t of Pb and operated intermittently in the period 1780-1899 (Cole, 1922; Mitchell, 2004); the ore from this site was smelted elsewhere

(Wales) (Moles et al., 2004) so local dispersion is not increased from that mechanism, as at other sites. This mineralisation is similar in composition to that of the S Armagh district. Tailings are described as 'typically' 10 wt% Pb (Moles et al., 2004), and although occupying a restricted area have been reported as being heavily eroded. A large proportion of the Pb now occurs as cerrusite (PbCO3) as a result of secondary reactions of the primary ore minerals (Moles et al., 2004), which has also been previously observed in the terrane in Leadhills.

The Historic Mine Sites – Inventory and Risk Classification Project by EPA & GSI (2009) lists Tassan mine (lead/silver) as the largest and most productive of the Monaghan District mines; concentrations of Pb in stream sediment downstream of the mine are reported to exceed the 1000 mg/kg guideline limit for livestock. In the same mining district Hope mine (Cornalough) was a small mine that produced a limited quantity of Pb ore with a reported limited impact on the environment (EPA & GSI, 2009).

Mesothermal quartz vein hosted Au is found at Clontibret, Co Monaghan, mineralisation includes stibnite (Sb_2S_3). These veins are the same intrusion that hosts Pb mineralisation in S Armagh-Monaghan mining district. Clontibret mine is reported by EPA & GSI (2009) as of concern for the elevated concentrations of Sb, As and Au in stream sediments. Au is recognised in two areas of S Armagh (Mitchell, 2004). Alluvial Au is also common throughout Down and Armagh, and was widely recognised in panned concentrates collected by the Tellus project field teams (Field Database).

2.5.1 Geology: the Longford-Down-Southern Uplands terrane

The majority of the Ordovician-Silurian strata of the Down and Armagh section are turbidite sequences, in a structurally complex setting (Mitchell, 2004), which represent an extention of the of the Southern Uplands-Down-Longford terrane of Scotland into Ireland (Breward et al. 2011). The sediments predominantly comprise arenaceous to argillaceous facies in fining up depositional sequences. Of particular note geochemically are: the substantially higher carbonate content (up to 20%) of the Hawick Group, arising from replacement of the clay matrix, along with concretions/nodules; and, the higher organic matter in the Moffatt Shale Group with beds of black mudstone (Mitchell, 2004). Stream sediment data from the Tellus dataset have been used to ascertain lithogeochemical signatures in this terrane, and these have confirmed that there is a general along-strike conformity within the lithological groups, other than for base metal mineralisation (Breward et al., 2011).

3 Exploratory Data Analysis (EDA)

3.1 REVIEW OF EXISTING DATA

The first stage of the EDA has been to set up the relevant datasets in both a database and GIS environment to allow EDA and later statistical analysis and preparation of tables and figures for publication and dissemination.

Spatial datasets supplied by GSNI have been compiled to allow their integration into the EDA process.

The spatial datasets used are:

- TELLUS Geochemistry sediments
- historic mines (GSNI Abandoned mines project)
- geology (solid) at 1:250,000
- geology (drift) at 1:250,000

- drainage (stream and river network polylines)
- river basins (polygons)
- topography (OSNI)
- counties

Attributes of the "solid geology", "river basins" and "counties" layers have been joined to the TELLUS geochemistry data by using spatial location in W:\Teams\GeochemB\TellusScientificServices\Data\Mininglegacy\DATA\GIS\GIS DATA\TellusGeochemistry\Sediments\exports outputs from GIS\Sediment_XRF_joint_to_wmu_live_DBO_Join_counties_Join_Geology_Join_Output.xlsx

The following paragraphs give a brief description of the principal datasets.

3.1.1 Tellus geochemical dataset

Stream sediment were collected in 1994-96 (2,908 sites in the west) and in 2004-06 (2,966 sites in the east) as part of the Tellus Project. The project, which comprised an integrated airborne geophysical survey and ground geochemical survey of Northern Ireland, was implemented to provide high resolution regional baseline datasets to underpin government and private body policy decisions concerning sustainable economic development, social infrastructure, environment and human health.

Sediment samples were collected from predominantly 1st and 2nd order streams at a density of approximately one site per 2.4 km^2. Sediments were wet sieved at site to yield a <150µm fraction for analysis. Sample preparation and analysis were undertaken at the laboratories of the British Geological Survey, Keyworth, Nottingham. Samples were freeze dried and a sub-sample of this material was pulverised and homogenised in an agate ball-mill for 30 minutes prior to preparation of a 12g pressed powder pellet. Sediments were analysed by X-ray fluorescence spectroscopy (XRF). Full details of all sampling, analytical and quality control methods are given in Smyth (2007).

3.1.2 Historic mines

The GSNI database has 595 entries of mineral locations and has identified 1999 mine shaft locations in Northern Ireland. The 595 mineral occurrences include both metallic and non-metallic minerals with 50 classes. Of the 1999 "mine shaft" entries, 860 are related to mines in the database and a total of 118 mines are named (see Appendix).

Table 5 illustrates the mineral types and number of occurrence present in the study area: Co. Armagh, Northern Ireland (GSNI Database). Within the metalliferous minerals the main metal occurrences in Co. Armagh are associated to lead mineralisation, hosted in the sedimentary rocks of Southern Uplands-Down-Longford Terrane, composed of an Ordovician and Silurian turbidite sequence comprising greywacke sandstone, siltstone and mudstone (Cole, 1922) (see section 2.5).

Table 5: Mineral type and number of occurrence in Co. Armagh, Northern Ireland (GSNI Database)

Mineral Name	Occurrence
Baryte	2
Calcite	1
Dolomite	1
Fluorite	1
Galena	21
Hematite	1
Iron	1
Malachite	1
Manganese Oxide	1
Pyrite	10
Sphalerite	3

3.1.3 River Basins

The river basins districts (RBD) are the primary reporting units of the Water Framework Directive (WFD). The North Eastern RBD is entirely in Northern Ireland and three are cross-border International River Basin Districts with the Republic of Ireland (Neagh Bann RBD, Shannon RBD, North Western RBD). Within the RBDs, the 35 river basin bodies comprised in Northern Ireland are listed in Appendix.

3.2 EXPLORATION OF THE DATA

Following on from the initial data gathering phase, exploratory data analysis (EDA) has been undertaken. Component parts undertaken are as follows:

- Comparison of sediment concentrations in the Northern Ireland dataset with available UK adopted sediment quality indicators.
- Exploration of the Ordovician-Silurian geochemical dataset.

3.2.1 Comparison of sediment concentrations in the Northern Ireland dataset with available UK adopted sediment quality indicators

Table 6 quantifies the number of sites in Northern Ireland which would fail to meet the sediment quality guidelines available for As, Cd, Cr, Cu, Pb, Ni, and Zn. Figure 3 shows the distribution of sediment quality failures for Pb by counties.

Table 6: Summary of the sediment quality failures of the Northern Ireland dataset based on the UK adopted sediment quality indicators

	As	Cd	Cr	Cu	Pb	Ni	Zn
	Number Occurrence (% total, N tot=5874)						
Below PEL	4693 (79.89)	5749 (97.87)	1366 (23.26)	5870 (99.93)	5682 (96.73)	1429 (24.33)	5489 (93.45)
Above PEL	1181 (20.11)	125 (2.13)	4508 (76.74)	4 (0.068)	192 (3.27)	4445 (75.67)	385 (6.55)

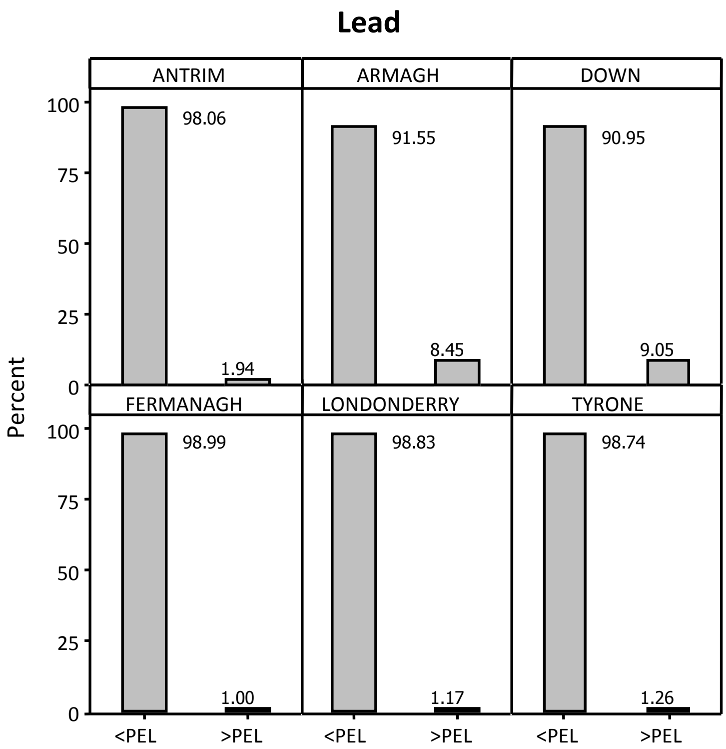


Figure 3: Bar charts showing the distribution by Northern Ireland Counties of sediment quality failure for Pb based on the UK adopted sediment quality indicators.

3.2.2 Exploration of the stream sediment data for the Ordovician-Silurian terrane

The data which form the basis of our exploration to define background metal values have been selected where samples lie over the Ordovician or Silurian succession of Down and Armagh, as shown in Figure 4. The regional geochemical dataset for Down-Armagh reflects the bedrock features, despite the thick cover of glacigenic deposits which are rarely cut through by streams Breward et al. 2011).

These data are statistically summarised in Table 7. When compared with available TEL and PEL it can be seen that the median is always greater than or equal to the TEL, and elements such as Cr and Ni almost ubiquitously exceed this value. This highlights the need to develop a parent material specific quantification of background data populations, since there is no suggestion that these ‘high’ Ni and Cr stream sediment concentrations are not predominantly of natural origin. Example EDA probability plot is shown in Figure 5 for Pb, demonstrating variation between these parent materials. The example map of samples sites compared to the PEL and TEL concentrations (Figure 6) shows that there is also spatial variation within these units. This thus demonstrates how these will form the basis of calculation of concentration populations using methods described in section 3.3. These data have also been explored by River Catchment (as described above) in relation to concentration of elements of interest in stream sediments – this also enables those which are of greatest interest in assessing the likely sediment sources and background concentrations to be assessed. An example of this shown for the surface water catchments over the Ordovician-Silurian of Armagh in Figure 7, demonstrating that some catchments have samples which are almost ubiquitously in exceedance of the TEL value shown in Table 2.

A preliminary examination, prior to the formal statistical outputs, has been made of Mn and Fe in relation to the trace elements, in order to try to use the approach of Butt and Nichol (section 2.4) in comparing potential physical versus chemical concentration of the trace elements. An example of this (Zn and Mn) is shown in Figure 8; there is some suggestion in these data (over a much larger area and with a different dataset to the original study) that it may be possible in the final statistical output to compare these processes as part of the tool for determining background concentrations.

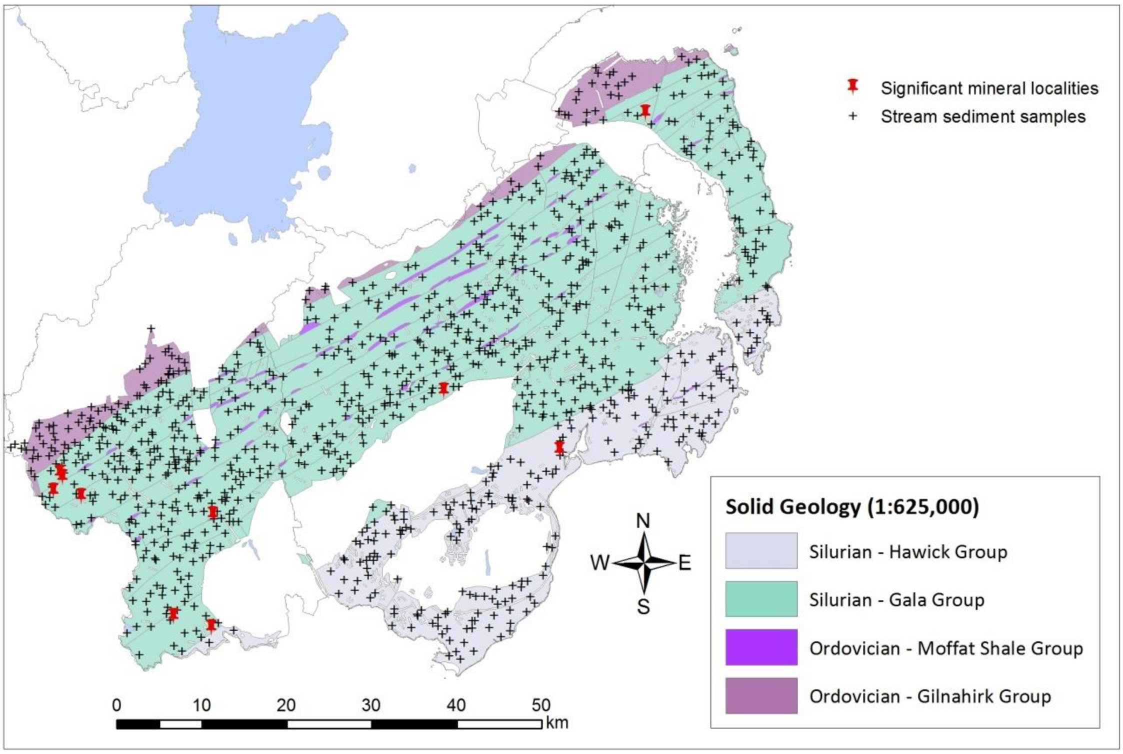


Figure 4: Map of the Ordovician-Silurian outcrop area and stream sediment sites.

Table 7: Summary statistics of selected stream sediment data

Variable	n	minimum	Q1	median	Q3	maximum	mean	skewness
		(mg/kg)						
Pb	1014	8	25	35	54	**1245**	50	11
Zn	1014	32	124	172	257	**3162**	237	6
Cd	1014	0.3	0.3	0.6	1.2	**56**	1.4	9
Sb*	1014	0.3	0.6	0.8	1.2	7.4	0.9	3
As	1014	0.9	7.3	11	**17**	**357**	16	9
Cu	1014	7	30	37	48	**260**	41	3
Cr	1014	16	**127**	**146**	**185**	**407**	**161**	1
Ni	1014	4	**52**	**64**	**83**	**250**	**69**	2

Shaded data ≥ TEL; Bold data ≥ PEL (cf. Table 2). * No TEL or PEL.

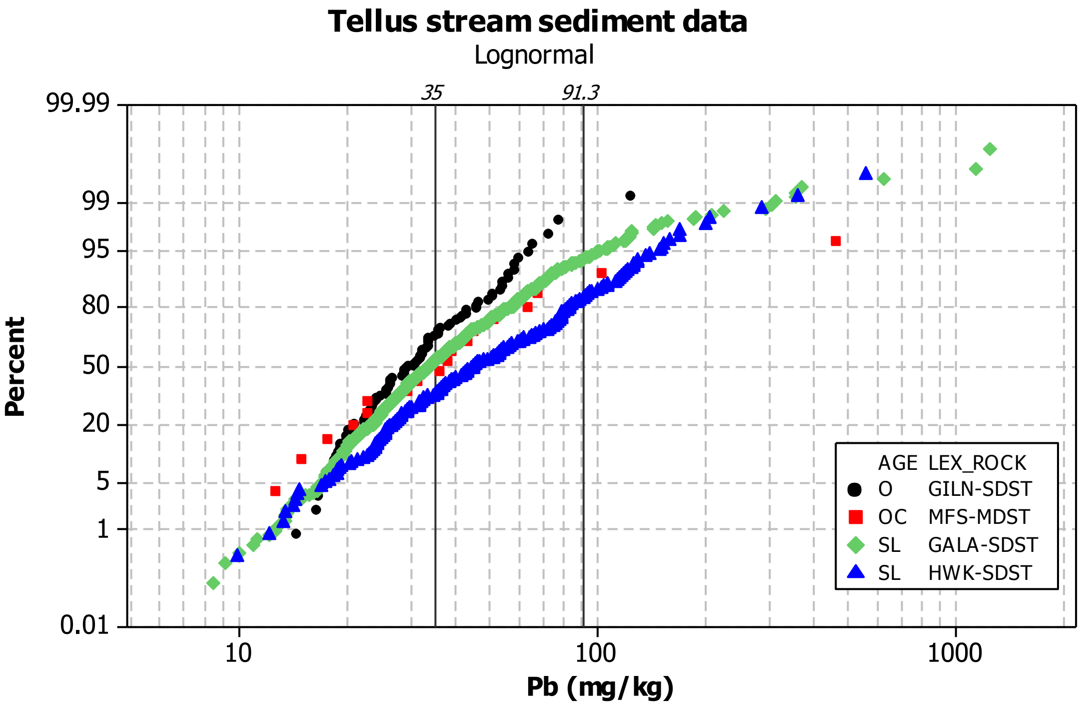


Figure 5: Probability plot of Pb data by parent material.

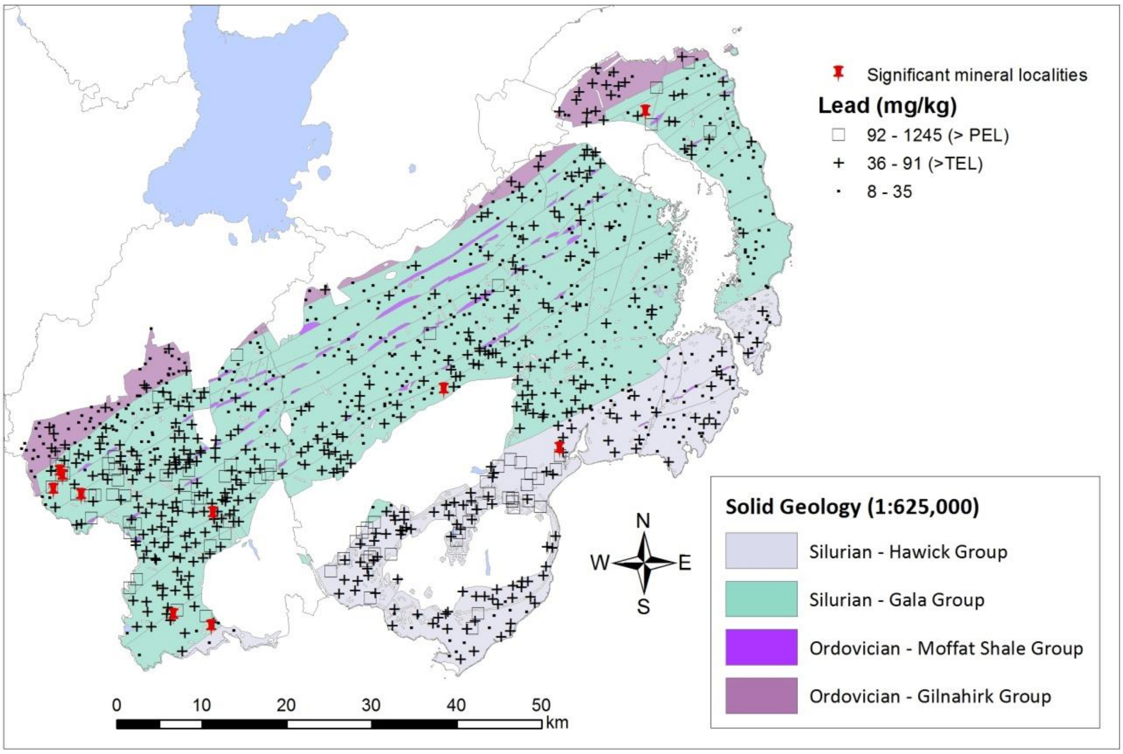
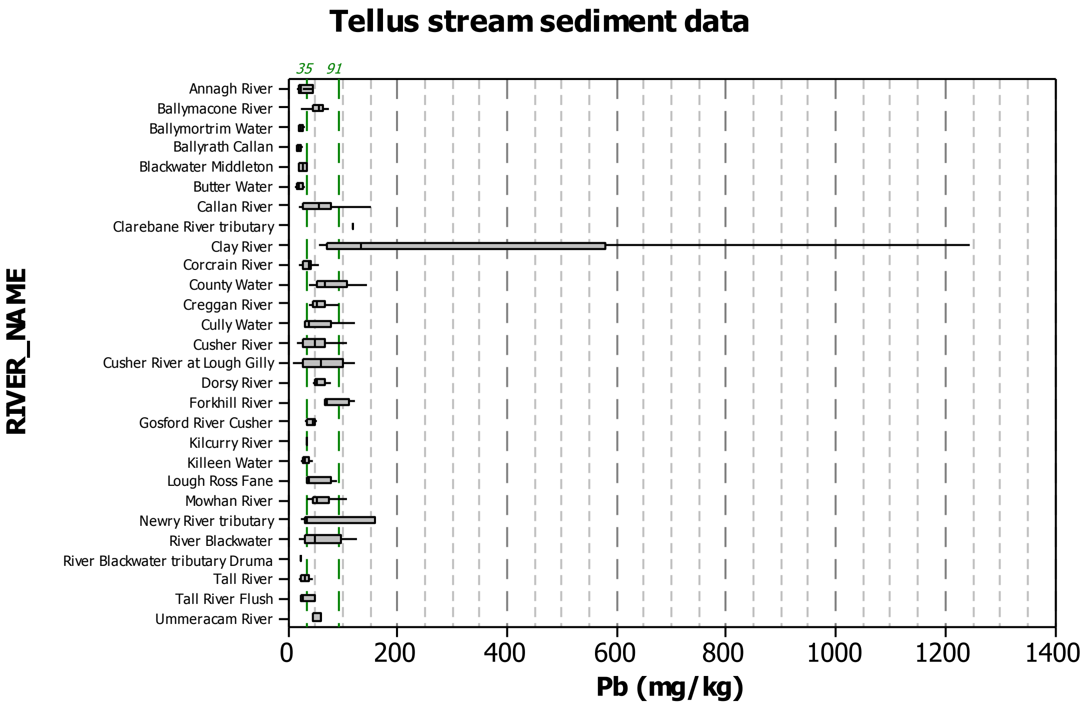


Figure 6: Example map of sample sites compared to TEL and PEL: Pb in stream sediments.



TEL and PEL shown in green (cf. Table 2)

Figure 7: Example EDA boxplot of stream sediment data with surface water catchments: Pb in Co Armagh.

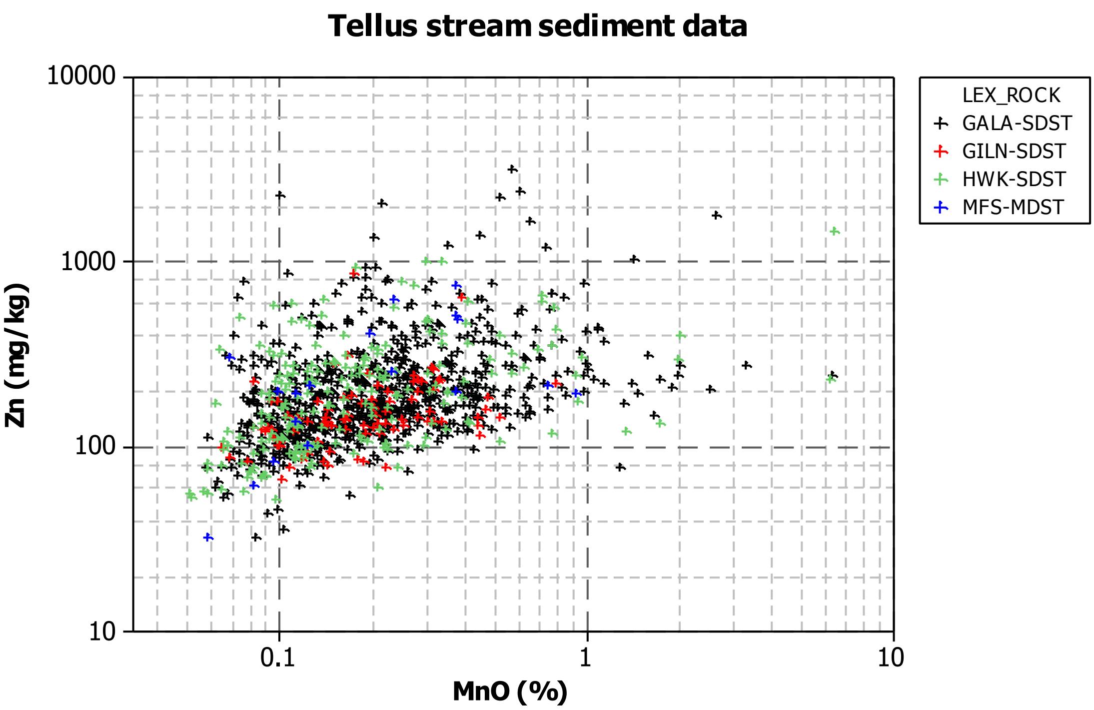


Figure 8: Comparison of Zn and Mn (as MnO) concentrations.

3.3 STATISTICAL ANALYSIS AND OUTPUTS

3.3.1 Selection of data population threshold values using probability graphs

Statistical analysis of the data and selection of threshold values between anomalous and background geochemical data is undertaken using the method of Sinclair (1976a) and applied using the 'PROBPLOT' code (Stanley, 1987). This method takes multi-modal distributed data and allows the user to iteratively identify the component populations, until the point is reached at which the modelled populations recombine to describe the distribution of the input data. This method offers the benefit of analysing the data independent of location information. There is, therefore, no requirement to have complete knowledge of all mining activity locations for its success, and when the groupings are projected in a GIS they offer an independent means of checking the success of those classifications by comparison with available spatial information.

Fundamental to this method is the fact that the cumulative frequency distribution of a normally or log-normally distributed data set plotted on a probability scale defines a straight line. The mean value estimate of this population can be read as the ordinate value corresponding to the 50 percentile and the values of the mean plus or minus one standard deviation can be estimated by the ordinate values corresponding to the 16 and 84 percentile, respectively. Commonly, however, geochemical data do not plot as a straight line on probability plots, but have a curvature with an inflection point. Such patterns result from the presence of two or more populations within the data. The Probplot method provides a procedure for estimating the constituent populations from real mixed populations. It derives thresholds which split the data into populations for which statistical parameters (e.g. mean and standard deviation) are calculated. This splitting of the polymodal data provides a more robust calculation of these thresholds than simply assuming that a value (such as the 95th percentile) will always represent such a threshold in the polymodal data (Sinclair, 1974). This method will generally require >100 sample points although methods do exist where there are not, and does assume that the resolved populations are (log)-normally distributed. From this is computed the mean ($\overline{x}$) and standard deviation (σ) of those data, and population thresholds are calculated based upon

minimum threshold = $\overline{x} - 2\sigma$

maximum threshold = $\overline{x} + 2\sigma$

Overlaps may occur between these populations. In this case, any given value in the overlapping range cannot be assigned to a specific population. Following this procedure in mineralised areas it is possible ad example to separate the background population from the anomalous "mineralised – mining contaminated" population, and to calculate the background threshold as the upper limit within the background concentration. Further details on this method can be found in the following references as well as those cited above (Sinclair, 1976b; Rose et al., 1979).

3.3.2 R-script

The 'PROBPLOT' code (Stanley, 1987) has been reproduced in an 'R' script environment, and used in conjunction with the 'R' package *mixtools* (Benaglia et al., 2009). The R code and basic operator instructions are attached in the Appendix.

The operation of the R package is illustrated using simulated data. Table 8 shows simulated data used to illustrate the operation of the program. The simulated data consist of a mixture of three normally distributed populations (the original data in the first three rows of Table 8).

Table 8: Data and outputs used to illustrate the PROBPLOT program

Data stets	Type	Number of points	Mean	Standard deviation	% of population
population1	original	150	50	20	17.6
population2	original	400	90	10	47.1
population3	original	300	180	15	35.3
population1	hand selected	90	39.5	12.6	10.6
population2	hand selected	463	87.4	11.0	54.5
population3	hand selected	297	176.9	14.6	35.0
population1	optimised	124	45.8	16.2	14.5
population2	optimised	426	88.8	10.1	50.2
population3	optimised	300	177.3	15.1	35.3

The program plots out the data as a cumulative frequency curve (Figure 9) the x-axis of the plot is scaled so that a normal distribution will appear as a straight line. The user selects the points on the graph where there are breaks between populations (red arrows on Figure 9).

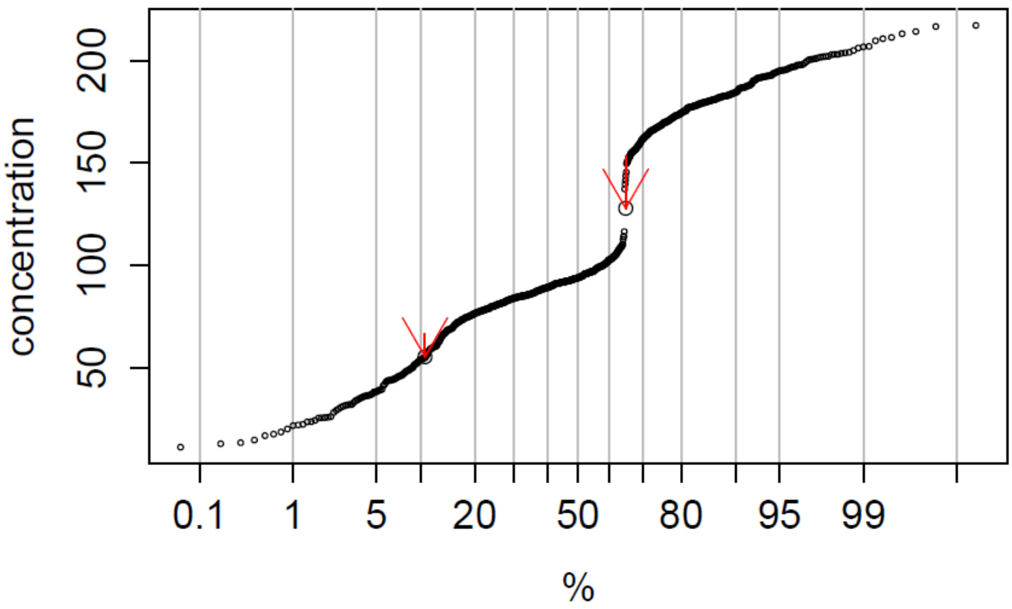


Figure 9: Example PROBPLOT cumulative frequency plot.

The program uses the selected locations to calculate the underlying normally distributed populations of data and adds these to the original data plot (Figure 10 red lines). The program estimates the fit to the original data based on the hand selected estimations of the populations (Figure 10 green line) and then uses an Expectation–maximization algorithm to optimise the fit to the data using the hand selected point as starting values for the optimisation (Figure 10 blue line). The optimised fit to the data is illustrated as a histogram of the original data with overlaid probability density plots of the estimates of the underlying populations.

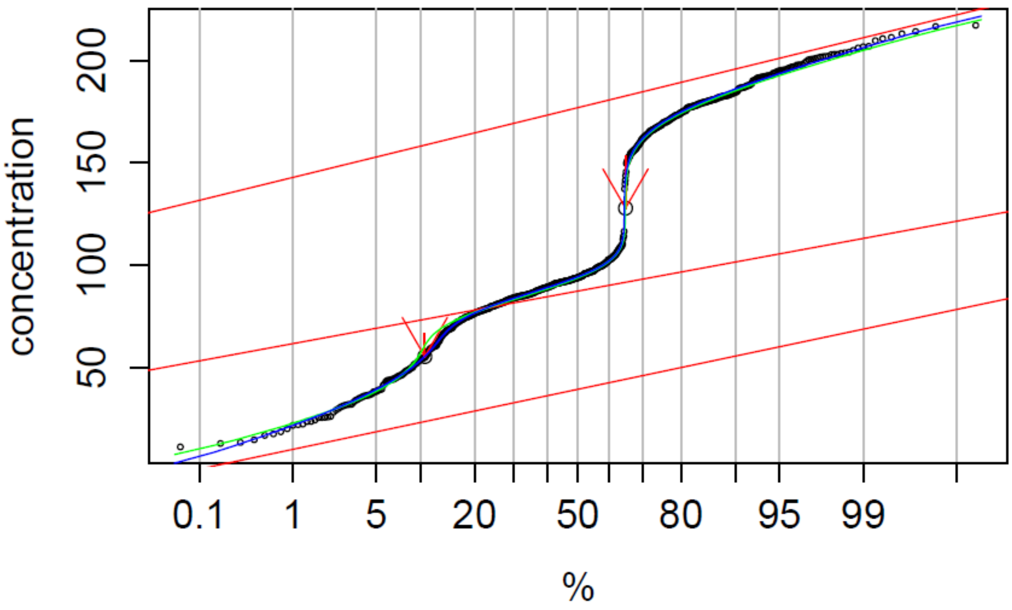


Figure 10: Probplot curve with estimated underlying populations (red lines), the hand selected fit (green line) and the optimised fit (blue line).

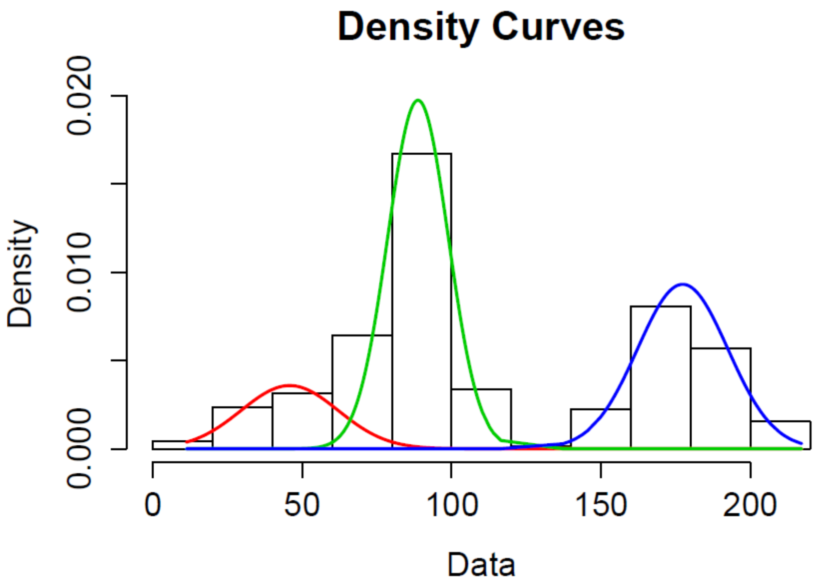


Figure 11: Optimised fit of the underlying distributions to the original data.

Table 8 compares the estimated mean, standard deviation and percentage of the population for both the hand estimated and the optimised outputs from the program to the original test data. Both methods give good estimates of the test data with slightly improved results for the optimised method. The results for both methods are less accurate where there is a greater degree of overlap between populations (populations 1 and 2, Table 8).

Whilst this example illustrates that R-PROBPLOT program is accurately carrying out the required functions, it is still in a fairly rudimentary form and needs further development and testing to ensure that it works for real data in a variety of scenarios.

3.3.3 Lead data population

The distribution of the stream sediment Pb data is shown in Figure 12 and it can be seen that 50% of the data exceed the TEL and 8% the PEL.

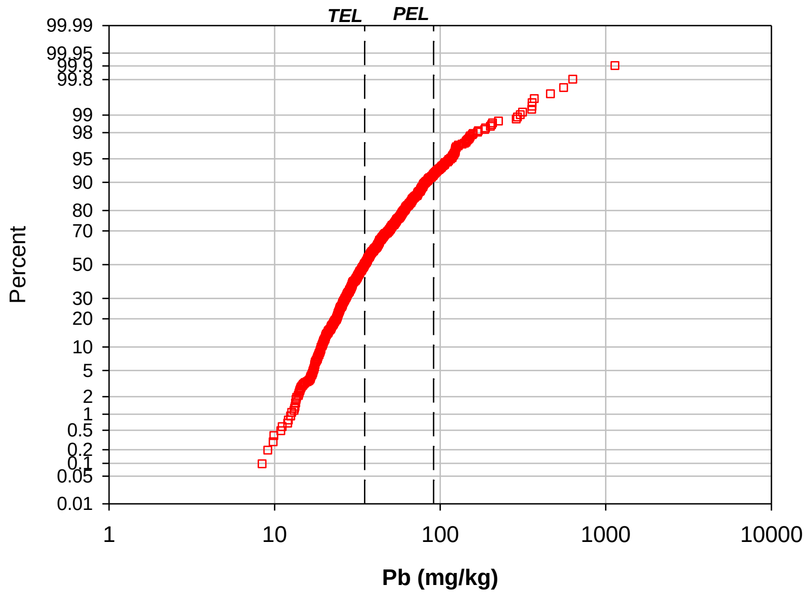


Figure 12: Probability distribution of stream sediment Pb concentrations with TEL and PEL (cf. Table 2).

When these data are partitioned into their component populations using the method of Sinclair (1976a) and applied using the 'PROBPLOT' code (Stanley, 1987), reproduced in an 'R' script environment, these seem to be best described by a four population model (Figure 13), with population mean values and thresholds ($\overline{x} \pm 2\sigma$) as shown in Table 9. Population 1 and 4 cannot be delineated with much precision because of the small percentage of total data that each represent.

In choosing thresholds for distinction between anomalous and background values there is no need to consider the bottom population 1. The critical part of the probability plot curve is the upper part, which is partitioned in populations 3 (thresholds 27.6 - 119.3) and 4 (thresholds 65.4 - 829.4), respectively. Applying the criteria to define each population as the mean ($\overline{x}$) ± 2σ (section 2.4), some overlaps occur between the four populations. Additional classes defined as "overlapping populations" are therefore created where data could form part of the data population above or below, as shown in Table 10. The two populations 3 and 4 overlap and an intermediate population 3-4 can be defined with thresholds 65.4 - 119.3 (Table 10).

The partitioned populations can then be inspected in relation to understanding different background populations, and how these may be controlled by local variations in mineralogy, or mineral extraction legacy across the landscape of a geological unit. Figure 14 shows the spatial distribution of the Pb populations as described above. It can be seen that the method successfully separates the highly anomalous data in population 4* from the rest of the data. This population is found to correspond to the most significant mining localities in the area. Population 3* can be interpreted as mineralised background of the Silurian-Ordovician terrane. The intermediate population (3-4) mostly clusters around the anomalous values and will contain both anomalous and background values. The highly anomalous populations in the Silurian Hawick Group form a halo around the Paleogene Mourne Mountain central complex. The other modelled Pb populations with relatively lower Pb concentrations cannot be clearly attributed to the different geological units of the Southern Upland- Down- Longford terrane.

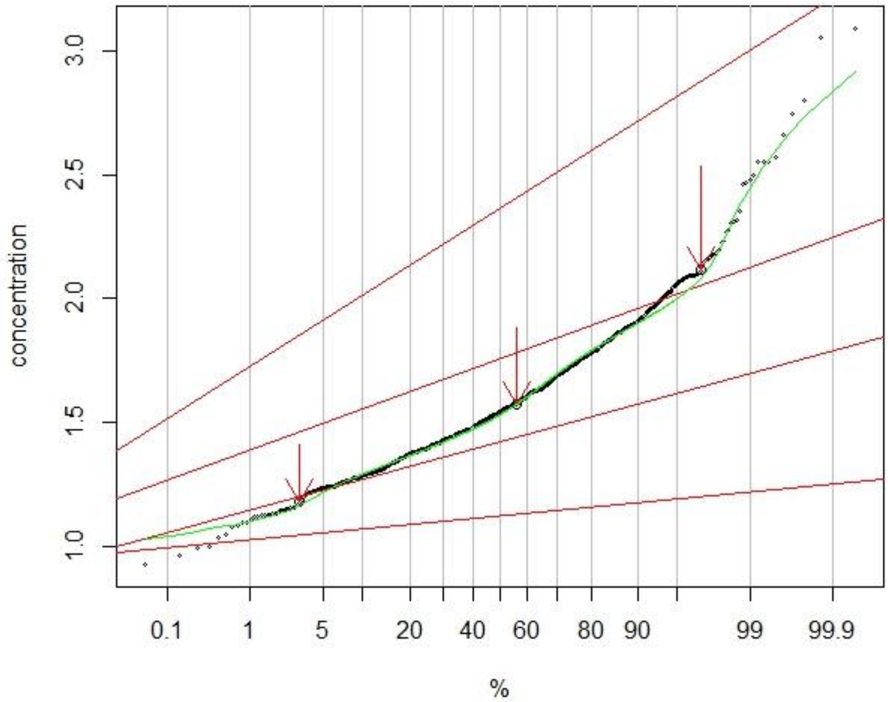


Figure 13: Probability plot showing log_{10}-Pb original data, plotted as black circles, with four populations (red line) partitioned using the partitioning procedures of Sinclair (1976a). Red arrows indicate inflection points where the modelled populations join. The modelled populations are recombined proportionally (green line) to compare with the original data.

Table 9: Statistical description of partitioned populations, Pb (mg/kg) in stream sediments

Pb Population	Mean	%	-sd	+sd	Thresholds	
					Min	Max
1	13.2	2.8	12.0	14.6	10.9	16.0
2	26.4	53.5	20.1	34.7	15.3	45.6
3	57.3	41.0	39.8	82.7	27.6	119.3
4	232.9	2.7	123.4	439.6	65.4	829.4

Table 10: Thresholds of partitioned populations including overlapping populations, Pb (mg/kg) in stream sediments

Pb Population	Thresholds	
	Min	Max
1	11.0	15
2*	15	27.6
2-3	27.6	45.6
3*	45.6	65.4
3-4	65.4	119.3
4*	119.3	829.4

*The symbol indicates that the original population from Table 9 has been redefined for its thresholds on the basis of the overlapping populations.

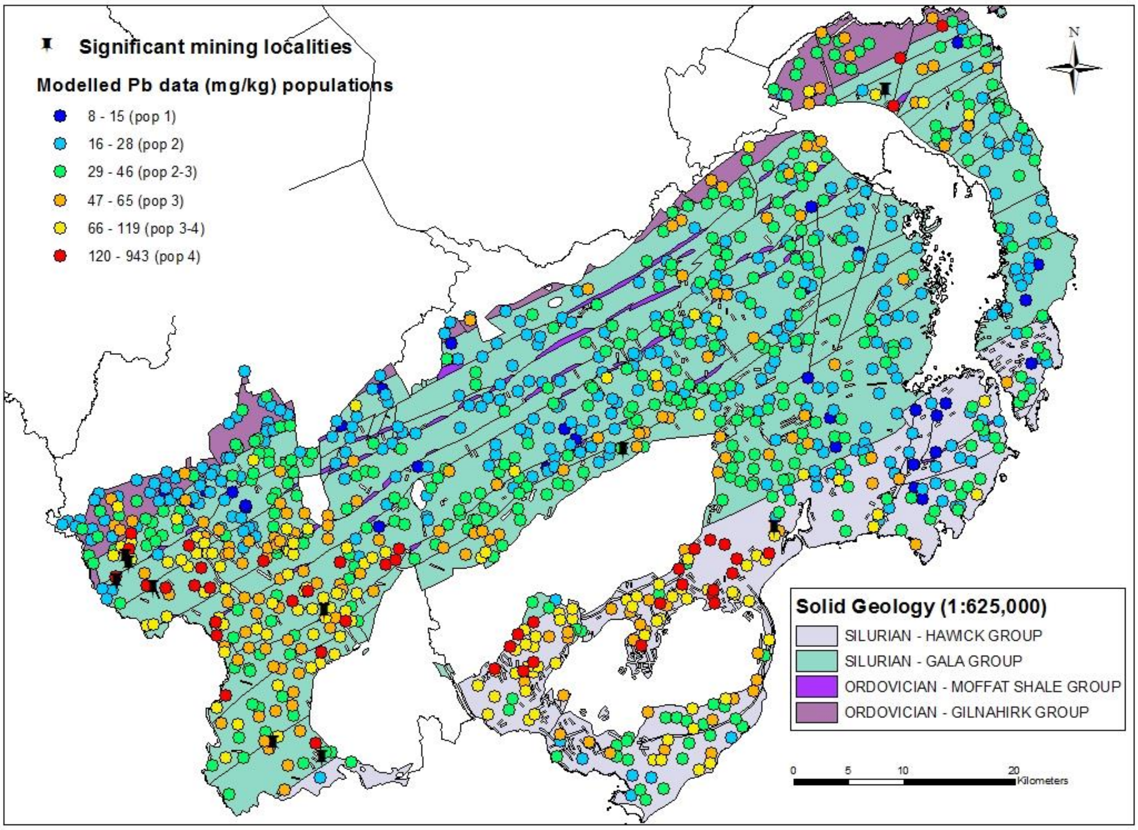


Figure 14: Map of classification of Pb stream sediment concentrations: based on modelled populations shown in Table 10 with additional "overlapping" populations reflecting data which may be drawn from either of the surrounding data populations.

3.3.4 Zinc data population

The modelling outputs for Zn (Figure 15, Table 11 and Table 12), using the protocol as described above, have produced four data populations. When examined spatially (Figure 16) populations 3 and 3-4 that are likely to reflect high mineralised background concentrations are seen in South Armagh, along with isolated higher concentrations (population 4). The latter may reflect either direct intersection of a mineral vein, concentration magnification associated with iron/manganese oxides, or proximity to historical mining activities. The majority of the outcrop can be seen to fall into lower data population concentration thresholds (populations 1 to 2).

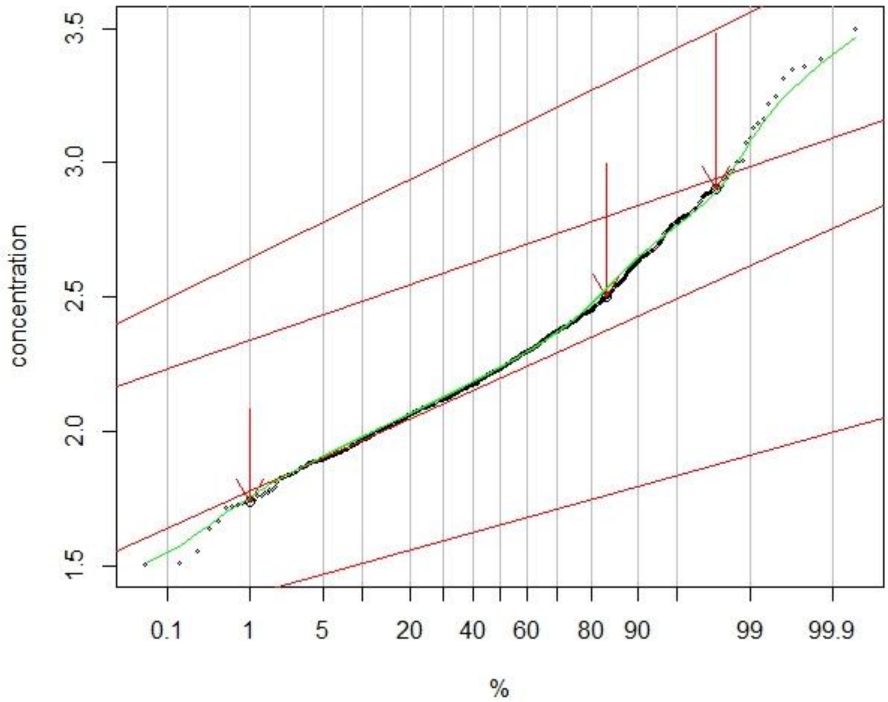


Figure 15: Probability plot showing log_{10}-Zn original data, plotted as black circles, with four populations (red line), partitioned using the partitioning procedures of Sinclair (1976a). Red arrows indicate inflection points where the modelled populations join. The modelled populations are recombined proportionally (green line) to compare with original data.

Table 11: Statistical description of partitioned populations, Zn (mg/kg) in stream sediments

Zn Population	Mean	%	-sd	+sd	Thresholds	
					Min	Max
1	45.0	0.5	34.8	58.1	27.0	75.1
2	157.9	83.4	104.3	239.1	68.9	362.1
3	461.3	14.3	334.7	635.9	242.8	876.5
4	1264.7	1.8	804.6	1988.1	511.8	3125.3

Table 12: Thresholds of partitioned populations including overlapping populations, Zn (mg/kg) in stream sediments

Zn Population	Thresholds	
	Min	Max
1*	27	68.9
1-2	68.9	75.1
2*	75.1	242.8
2-3	242.8	362.1
3*	362.1	511.8
3-4	511.8	876.5
4*	876.5	3125.3

*The symbol indicates that the original population from Table 11 has been redefined for its thresholds on the basis of the overlapping populations.

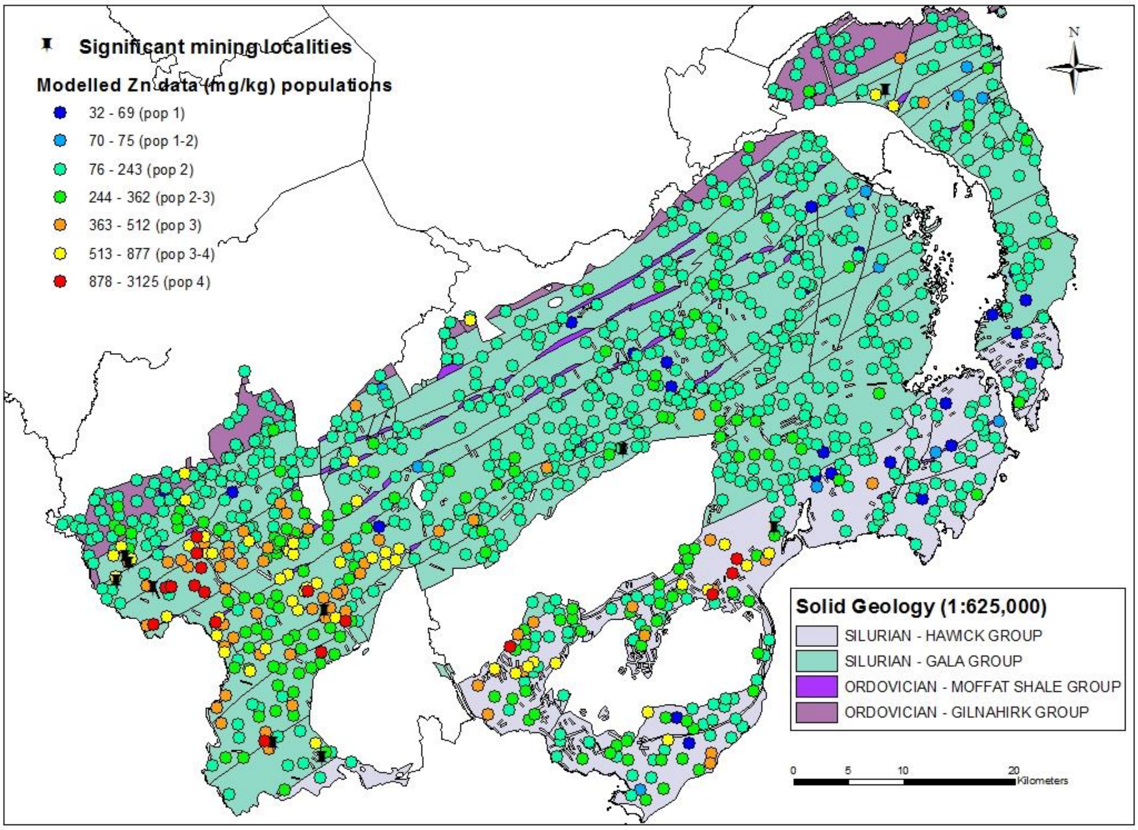


Figure 16: Map of classification of Zn stream sediment concentrations: based on modelled populations shown in Table 12 with additional "overlapping" populations reflecting data which may be drawn from either of the surrounding data populations.

3.3.5 Arsenic data population

A similar distribution to the base metals Pb and Zn is seen for As (Figure 17, Table 13, Table 14, Figure 18), with higher concentrations around the historical mining districts, although the proportion of samples falling within the lower concentration data populations is higher.

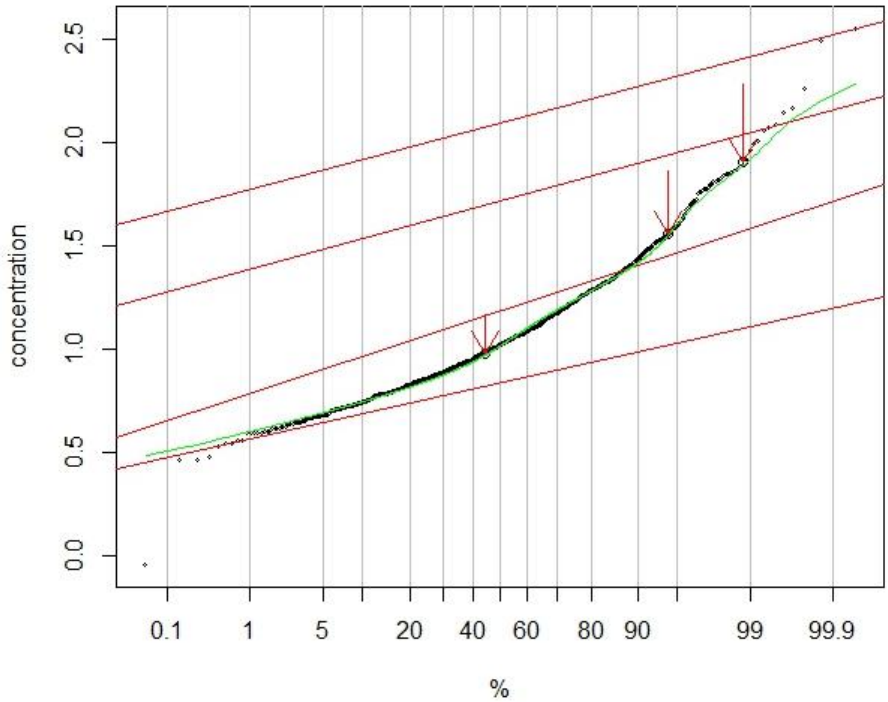


Figure 17: Probability plot showing log_{10}-As original data, plotted as black circles, with four populations (red line), partitioned using the partitioning procedures of Sinclair (1976a). Red arrows indicate inflection points where the modelled populations join. The modelled populations are recombined proportionally (green line) to compare with original data.

Table 13: Statistical description of partitioned populations, As (mg/kg) in stream sediments

As Population	Mean	%	-sd	+sd	Thresholds	
					Min	Max
1	6.8	44.6	5.2	9.0	4.0	11.7
2	15.2	49.6	10.3	22.7	6.9	33.7
3	51.9	5.1	37.4	72.1	26.9	100.1
4	124.0	0.7	90.3	170.4	65.7	234.0

Table 14: Thresholds of partitioned populations including overlapping populations, As (mg/kg) in stream sediments

As Population	Thresholds	
	Min	Max
1*	4.0	6.9
1-2	6.9	11.7
2*	11.7	26.9
2-3	26.9	33.7
3*	33.7	65.7
3-4	65.7	100.1
4*	100.1	234.0

*The symbol indicates that the original population from Table 13 has been redefined for its thresholds on the basis of the overlapping populations.

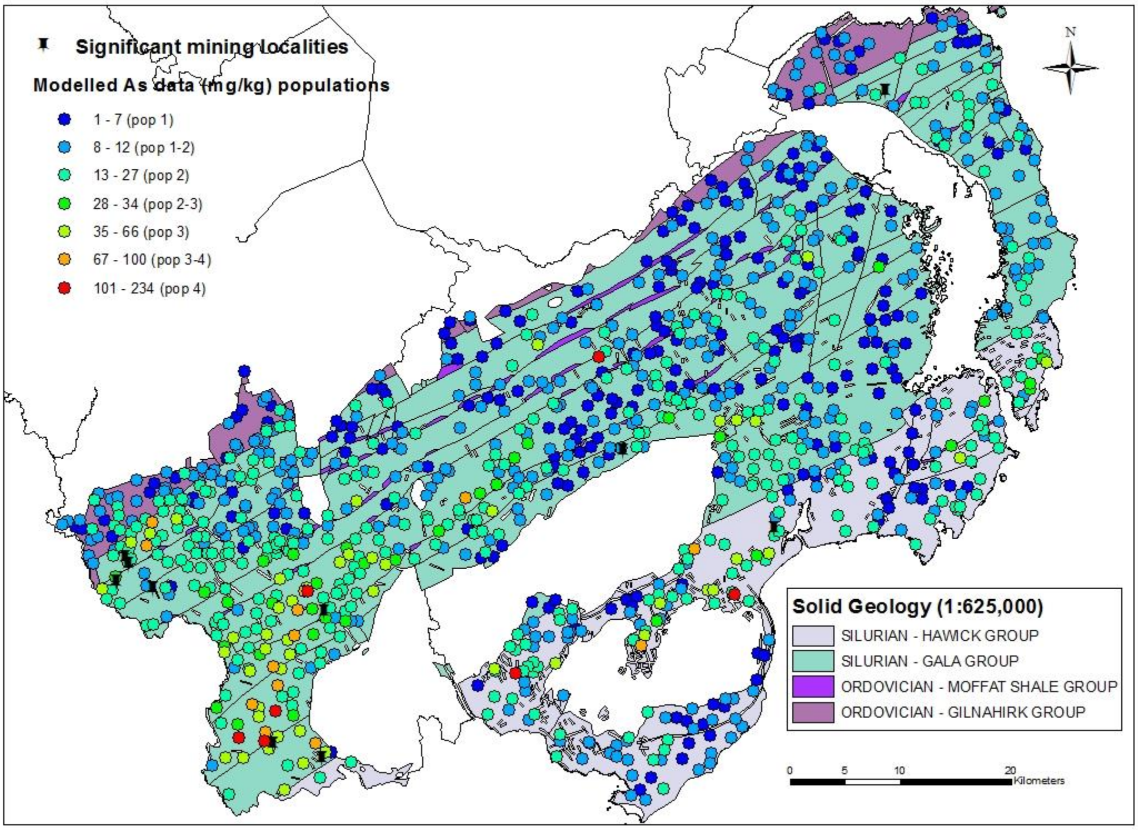


Figure 18: Map of classification of As stream sediment concentrations: based on modelled populations shown in Table 14 with additional "overlapping" populations reflecting data which may be drawn from either of the surrounding data populations.

3.3.6 Chromium data population

Data from modelling the Cr concentrations are shown in Figure 19, Table 15 and Table 16. The distribution of the modelled populations in Figure 20 indicates that there is no spatial relationship apparent between higher concentrations and the distribution of major mineral localities, which are often the site of historical mining activities. There is instead a greater correspondence between population 4 with the highest Cr concentrations and the northern Silurian Gala Group outcrop in Co. Armagh and the Ordovician Moffat Shale Group in northern Co. Down, reflecting more mafic meta-sediments, responsible for the Cr enrichment.

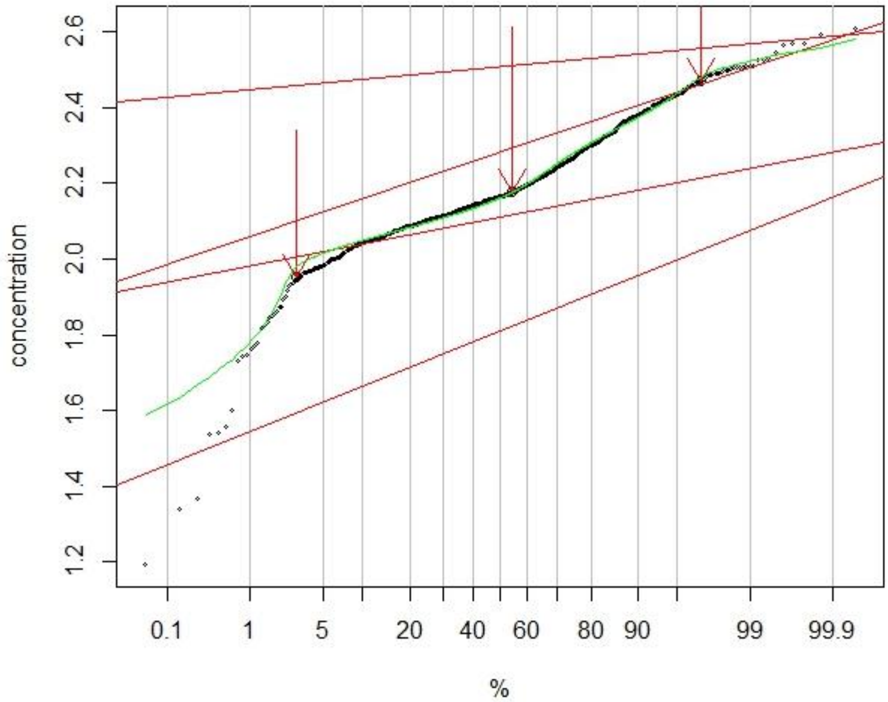


Figure 19: Probability plot showing log_{10}-Cr original data, plotted as black circles, with four populations (red line), partitioned using the partitioning procedures of Sinclair (1976a). Red arrows indicate inflection points where the modelled populations join. The modelled populations are recombined proportionally (green line) to compare with original data.

Table 15: Statistical description of partitioned populations, Cr (mg/kg) in stream sediments

Cr Population	Mean	%	-sd	+sd	Thresholds	
					Min	Max
1	64.8	2.6	49.8	84.3	38.3	109.6
2	128.6	52.0	113.2	146.1	99.7	166.0
3	192.2	42.7	154.1	239.7	123.5	299.1
4	321.5	2.7	302.9	341.1	285.5	361.9

Table 16: Thresholds of partitioned populations including overlapping populations, Cr (mg/kg) in stream sediments

Cr Population	Thresholds	
	Min	Max
1*	38.3	99.7
1-2	99.7	109.6
2*	109.6	123.5
2-3	123.5	166.0
3*	166.0	285.5
3-4	285.5	299.1
4*	299.1	361.9

*The symbol indicates that the original population from Table 15 has been redefined for its thresholds on the basis of the overlapping populations.

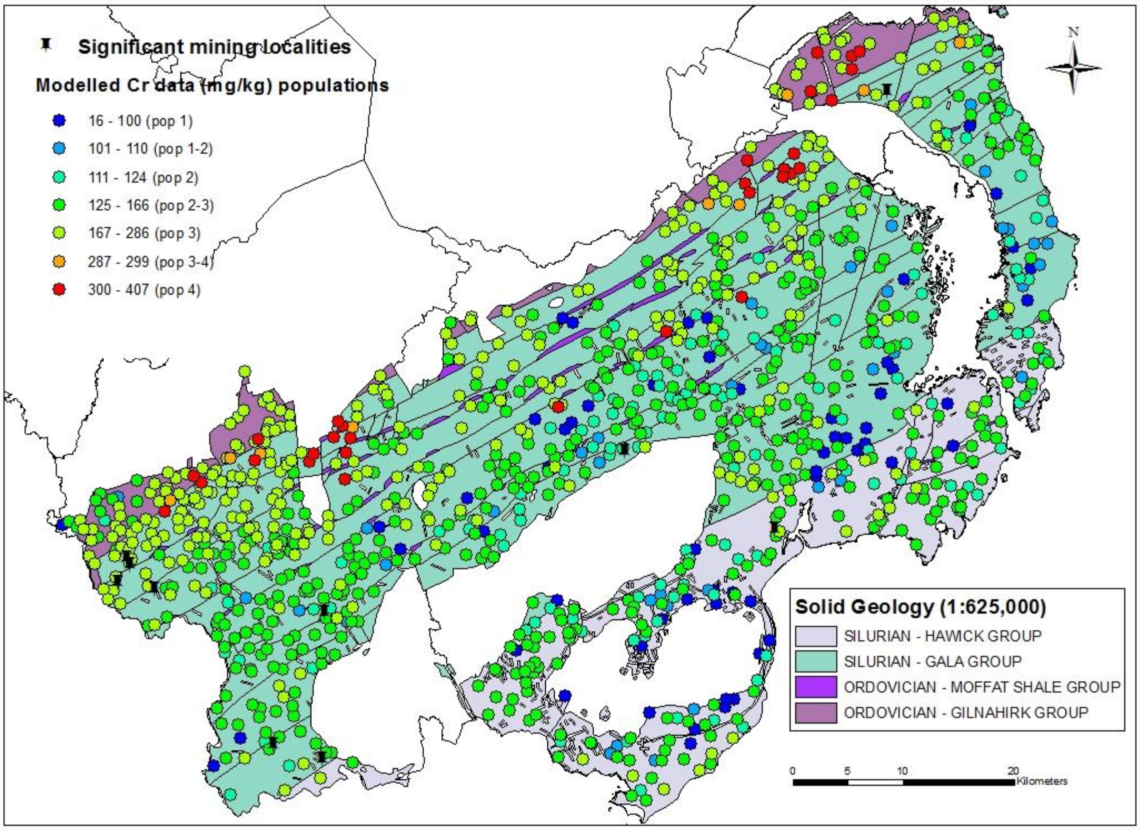


Figure 20: Map of classification of Cr stream sediment concentrations: based on modelled populations shown in Table 16 with additional "overlapping" populations reflecting data which may be drawn from either of the surrounding data populations.

3.3.7 Nickel data population

Data from modelling the Ni concentrations values into four populations are similarly shown in Figure 21, Table 17 and Table 18. The distribution map (Figure 22) of the modelled populations shows a very similar distribution as Cr; this is not unexpected given the chemical similarities between these two elements. It is shown the impact on background concentrations of the greater mafic rock inputs to the Ordovician sediments (e.g. east of Belfast), as well as areas falling into a higher background concentration in west Armagh. The background concentration of these sediments is thus greater than those from streams overlying the Silurian lithologies of Co. Down.

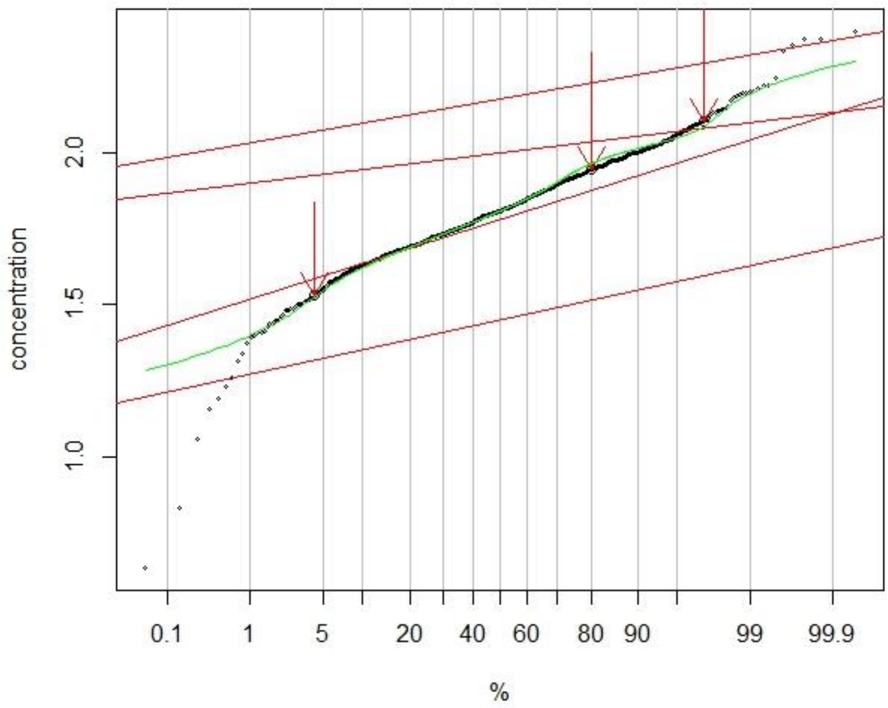


Figure 21: Probability plot showing $\log_{10}$-Ni original data, plotted as black circles, with four populations (red line), partitioned using the partitioning procedures of Sinclair (1976a). Red arrows indicate inflection points where the modelled populations join. The modelled populations are recombined proportionally (green line) to compare with original data.

Table 17: Statistical description of partitioned populations, Ni (mg/kg) in stream sediments

Ni Population	Mean	%	-sd	+sd	Thresholds	
					Min	Max
1	28.1	3.9	23.5	33.6	19.7	40.1
2	60.0	76.3	46.3	77.7	35.7	100.7
3	99.4	17.3	90.1	109.6	81.7	120.9
4	150.4	2.5	130.3	173.6	112.9	200.3

Table 18: Thresholds of partitioned populations including overlapping populations, Ni (mg/kg) in stream sediments

Ni Population	Thresholds	
	Min	Max
1*	19.7	35.7
1-2	35.7	40.1
2*	40.1	81.7
2-3	81.7	100.7
3*	100.7	112.9
3-4	112.9	120.9
4*	120.9	200.3

*The symbol indicates that the original population from Table 17 has been redefined for its thresholds on the basis of the overlapping populations.

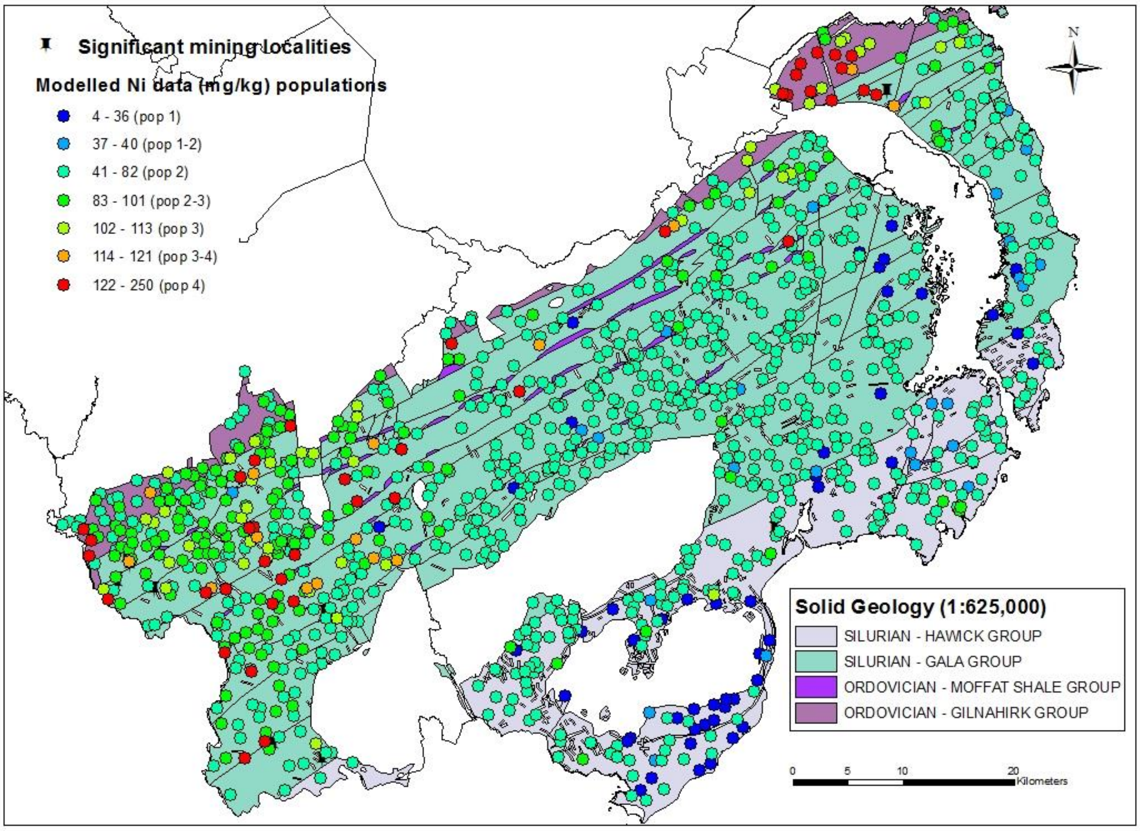


Figure 22: Map of classification of Ni stream sediment concentrations: based on modelled populations shown in Table 18 with additional "overlapping" populations reflecting data which may be drawn from either of the surrounding data populations.

4 Summary of findings

International literature has usefully provided both methodologies and examples of 'background' or environmental quality concentration thresholds for comparison to the Tellus stream sediment data. The lack of formal regulation but likelihood of future implementation makes this work particularly timely.

Exploratory data analysis showed that the probability plot method of Sinclair (1976a) adapted into an R-script environment could be applied to the stream sediment data overlying the Ordovician-Silurian bedrock in Counties Down and Armagh. Data analysis has primarily focused on elements for which there are sediment quality standards derived in other jurisdictions, which may be adopted in the UK regulatory framework in future (Table 2) and for which concentrations are typically higher in relation to these standards (Table 7).

Calculations have been made for Pb, Zn, As, Cr and Ni using the principles of the ProbPlot program. Each element distribution curve has been partitioned in four populations; the contributing populations can then be inspected in relation to understanding different background populations, and how these may be controlled by local variations in mineralogy or mineral extraction legacy across the landscape of a geological unit.

Where data exceed sediment quality standards these populations give some assistance in identifying where natural background concentrations (due to mineralogical variations in the catchment geology) may contribute more of the ‘contaminant’. This is designed to aid the decision-making process in relation to why quality standards may have failed, or if there is any merit in ‘remediation’ of a natural ecosystem (cf. section 2.3). Separation of these more widespread, potentially natural, high concentrations from the data populations which reflect very high concentrations, perhaps more likely to arise from point sources, could also help in targeting key sites for further investigation.

5 Suggestions for future research

Regional geochemical data can be used to assess stream sediment background concentrations in both mineralised and unmineralised areas. Using probability plot approaches to define data populations shows where higher concentration areas may be from mineralised, but unmined regions. This is being worked up into a peer-review publication, although we will undertake further model validation work (i.e. compare R script to original model software) to allow us to publish the R-script compilation that we have used.

This project has demonstrated that these approaches can quantify statistical population thresholds, using the derived mean and standard deviations. In addition, we suggest this research as opened up future research options which may be of direct benefit to the Tellus and/or Tellus Border region:

- cross-border evaluation of levelled data from Tellus and Tellus Border in the Monaghan-Antrim mining area;
- evaluation of mineralised background concentrations in other areas of historical mining; *and,*
- comparison with catchment soil data which would also reflect any underlying mineralisation and/or historical mining waste re-distribution.

6 References

Alpers, C N, Nordstrom, D.K. 2000. Estimation of pre-mining conditions for trace metal mobility in mineralized areas - an overviewof the 5th International Conference on Acid Rock Drainage (ICARD 2000), Littleton, Colorado, Society for Mining, Metallurgy, and Exploration, 463-472.

Ander, E L, and Casper, S. 2008. Determination of Metal Background Reference Concentrations: Feasibility Study. Environment Agency SC050063.

Applied Geochemistry Research Group. 1973. Provisional geochemical atlas of Northern Ireland. Applied Geochemistry Research Group, Imperial College, Technical Report 60.

Benaglia, T, Chauveau, D, Hunter, D R, and Young, D S. 2009. mixtools: An R Package for Analyzing Finite Mixture Models. Journal of Statistical Software, Vol. 32.

Bird, G, Brewer, P A, Macklin, M G, Nikolova, M, Kotsev, T, Mollov, M, and Swain, C. 2010. Dispersal of Contaminant Metals in the Mining-Affected Danube and Maritsa Drainage Basins, Bulgaria, Eastern Europe. Water Air and Soil Pollution, Vol. 206, 105-127.

Breward, N, Stone, P, Flight, D, and Anderson, T B. 2011. Regional geochemical comparisons from the Lower Palaeozoic, Southern Uplands-Down-Longford terrane in Northern Ireland and Scotland. Scottish Journal of Geology, Vol. 47, 33-43.

Butler, B A. 2009. Effect of pH, ionic strength, dissolved organic carbon, time, and particle size on metals release from mine drainage impacted streambed sediments. Water Research, Vol. 43, 1392-1402.

Butt, C R M, and Nichol, I. 1979. The identification of various types of geochemical stream sediment anomalies in Northern Ireland. Journal of Geochemical Exploration, Vol. 11, 13-32.

Byrne, P, Wood, P J, and Reid, I. 2013. The Impairment of River Systems by Metal Mine Contamination: A Review Including Remediation Options. Critical Reviews in Environmental Science and Technology, Vol. 42, 2017-2077.

Cole, G A J. 1922. Memoir and map of localities of minerals of economic importance and metalliferous mines in Ireland, Mem Geol Survey Ireland 155.

EC. 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. [OJ L327/1, 22.12.2000]

EC. 2011. Guidance Document No. 27 for Deriving Environmental Quality Standards. Common Implementation Strategy for the Water Framework Directive (2000/60/EC) - Technical Report - 2011 - 055.

EC. 1976. Council Directive 76/464/EEC of 4 May 1976 on pollution caused by certain dangerous substances discharged into the aquatic environment of the Community as amended by Council Directive 91/692/EEC (further amended by Council Regulation 1882/2003/EC), and Directive 2000/60/EC (further amended by Decision 2455/2001/EC).

EPA and GSI. 2009. Historic Mine Sites – Inventory and Risk Classification.

Hudson-Edwards, K A, Macklin, M G, Brewer, P A, and Dennis, I A. 2008. Assessment of Metal Mining-Contaminated River Sediments in England and Wales. Environment Agency, SC030136/4.

Matschullat, J, Ottenstein, R, and Reimann, C. 2000. Geochemical background - can we calculate it? . Environmental Geology, Vol. 39, 990-1000.

Mitchell, W I, (editor) 2004. The Geology of Northern Ireland - Our Natural Foundation. (Belfast: The Geological Survey of Northern Ireland.)

Moles, N R, Smyth, D, Maher, C E, Beattie, E H, and Kelly, M. 2004. Dispersion of cerussite-rich tailings and plant uptake of heavy meals at historical lead mines near Newtownards, Northern Ireland. Transactions of the Institute of Mining and Metallurgy Section B Applied Earth Sciences, Vol. 113, B21-B30.

Rose, A W, Hawkes, H E, and Webb, J S. 1979. Geochemistry in mineral exploration (2nd edition edition). (London: Academic Press.) ISBN ISBN 0-12-596252-5

Runkel, R L, Kimball, B A, Walton-Day, K, and Verplanck, P L. 2007. A simulation-based approach for estimating premining water quality: Red Mountain Creek, Colorado. Applied Geochemistry, Vol. 22, 1899-1918.

Runnels, D D, Shepherd, T A, and Angino, E E. 1992. Metals in water- Determining natural background concentrations in mineralized areas. Environ. Sci. Technol., Vol. 26, 2316-2323.

SedNet. 2004. Contaminated Sediments in European River Basins. SedNet Final summary report.

Sinclair, A J. 1974. Selection of threshold values in geochemical data using probability graphs. Journal of Geochemical Exploration, Vol. 3, 129-149.

Sinclair, A J. 1976a. Applications of probability graphs in mineral exploration. Special Volume no. 4. (Vancouver, BC, Canada: The Association of Exploration Geochemists.)

Sinclair, A J. 1976b. Univariate analysis. 59-81 in Statistics and data analysis in geochemical prospecting. Howarth, R J (editor). Handbook of Exploration Geochemistry, Volume 2. (Amsterdam: Elsevier Scientific Publishing Company.)

Stanley, C R. 1987. PROBPLOT. An interactive computer program to fit mixtures of normal (or log-normal) distributions with maximum likelihood optimization procedures. The Association of Exploration Geochemists.

7 List of References in Project EndNote Bibliography

Alpers, C N, Nordstrom, D.K. 2000. Estimation of pre-mining conditions for trace metal mobility in mineralized areas - an overviewof the 5th International Conference on Acid Rock Drainage (ICARD 2000), Littleton, Colorado, Society for Mining, Metallurgy, and Exploration, 463-472.

Ander, E L, and Casper, S. 2008. Determination of Metal Background Reference Concentrations: Feasibility Study. Environment Agency SC050063.

Ankley, G T, Di Toro, D M, Hansen, D J, and Berry, W J. 1996. Technical basis and proposal for deriving sediment quality criteria for metals. Environmental Toxicology and Chemistry, Vol. 15, 2056-2066.

Applied Geochemistry Research Group. 1973. Provisional geochemical atlas of Northern Ireland. Applied Geochemistry Research Group, Imperial College, Technical Report 60.

Barsby, A, McKinley, J M, Ofterdinger, U, Young, M, Cave, M R, and Wragg, J. 2012. Bioaccessibility of trace elements in soils in Northern Ireland. Science of The Total Environment, Vol. 433, 398-417.

Benaglia, T, Chauveau, D, Hunter, D R, and Young, D S. 2009. mixtools: An R Package for Analyzing Finite Mixture Models. Journal of Statistical Software, Vol. 32.

Bird, G, Brewer, P A, Macklin, M G, Balteanu, D, Serban, M, Driga, B, and Zaharia, S. 2008. River system recovery following the Novat-Rosu tailings dam failure, Maramures County, Romania. Applied Geochemistry, Vol. 23, 3498-3518.

Bird, G, Brewer, P A, Macklin, M G, Nikolova, M, Kotsev, T, Mollov, M, and Swain, C. 2010. Dispersal of Contaminant Metals in the Mining-Affected Danube and Maritsa Drainage Basins, Bulgaria, Eastern Europe. Water Air and Soil Pollution, Vol. 206, 105-127.

Breward, N, Stone, P, Flight, D, and Anderson, T B. 2011. Regional geochemical comparisons from the Lower Palaeozoic, Southern Uplands-Down-Longford terrane in Northern Ireland and Scotland. Scottish Journal of Geology, Vol. 47, 33-43.

Brydie, J R, and Polya, D A. 2003. Metal dispersion in sediments and waters of the River Conwy draining the Llanrwst Mining Field, North Wales. Mineralogical Magazine, Vol. 67, 289-304.

Butler, B A. 2009. Effect of pH, ionic strength, dissolved organic carbon, time, and particle size on metals release from mine drainage impacted streambed sediments. Water Research, Vol. 43, 1392-1402.

Butt, C R M, and Nichol, I. 1979. The identification of various types of geochemical stream sediment anomalies in Northern Ireland. Journal of Geochemical Exploration, Vol. 11, 13-32.

Byrne, P, Reid, I, and Wood, P J. 2010. Sediment geochemistry of streams draining abandoned lead/zinc mines in central Wales: the Afon Twymyn. Journal of Soils and Sediments, Vol. 10, 683-697.

Byrne, P, Wood, P J, and Reid, I. 2013. The Impairment of River Systems by Metal Mine Contamination: A Review Including Remediation Options. Critical Reviews in Environmental Science and Technology, Vol. 42, 2017-2077.

Cole, G A J. 1922. Memoir and map of localities of minerals of economic importance and metalliferous mines in Ireland, Mem Geol Survey Ireland 155.

Davies, B E. 1983. Graphical estimation of the normal lead content of some British soils. Geoderma, Vol. 29, 67-75.

Dennis, I A, Coulthard, T J, Brewer, P, and Macklin, M G. 2009. The role of floodplains in attenuating contaminated sediment fluxes in formerly mined drainage basins. Earth Surface Processes and Landforms, Vol. 34, 453-466.

European Communities. 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. [OJ L327/1, 22.12.2000]

European Communities. 2011. Common Implementation Strategy for the Water Framework Directive (2000/60/EC) - Guidance Document No. 27 Technical Guidance For Deriving Environmental Quality Standards. Technical Report - 2011 - 055.

European Council. 1976. Council Directive 76/464/EEC of 4 May 1976 on pollution caused by certain dangerous substances discharged into the aquatic environment of the Community as amended by Council Directive 91/692/EEC (further amended by Council Regulation 1882/2003/EC), and Directive 2000/60/EC (further amended by Decision 2455/2001/EC).

European Parliament. 2006. Directive 2006/21/EC of the European Parliament and of the Council on the management of waste from extractive industries and amending Directive 2004/35/EC, Official Journal of the European Union.

EPA and GSI. 2009. Historic Mine Sites – Inventory and Risk Classification.

Hudson-Edwards, K A. 2003. Sources, mineralogy, chemistry and fate of heavy metal-bearing particles in mining-affected river systems. Mineralogical Magazine, Vol. 67, 205-217.

Hudson-Edwards, K A, Macklin, M G, Brewer, P A, and Dennis, I A. 2008. Assessment of Metal Mining-Contaminated River Sediments in England and Wales. Environment Agency, SC030136/4.

MacKenzie, A B, and Pulford, I D. 2002. Investigation of contaminant metal dispersal from a disused mine site at Tyndrum, Scotland, using concentration gradients and stable Pb isotope ratios. Applied Geochemistry, Vol. 17, 1093-1103.

Macklin, M G, Brewer, P A, Hudson-Edwards, K A, Bird, G, Coulthard, T J, Dennis, I A, Lechler, P J, Miller, J R, and Turner, J N. 2006. A geomorphological approach to the management of rivers contaminated by metal mining. Geomorphology, Vol. 79, 423-447.

Matschullat, J, Ottenstein, R, and Reimann, C. 2000. Geochemical background - can we calculate it? . Environmental Geology, Vol. 39, 990-1000.

Mitchell, W I, (editor) 2004. The Geology of Northern Ireland - Our Natural Foundation. (Belfast: The Geological Survey of Northern Ireland.)

Moles, N R, Betz, S M, McCready, A J, and Murphy, P J. 2003. Replacement and authigenic mineralogy of metal contaminants in stream and estuarine sediments at Newtownards, Northern Ireland. Mineralogical Magazine, Vol. 67, 305-324.

Moles, N R, Smyth, D, Maher, C E, Beattie, E H, and Kelly, M. 2004. Dispersion of cerussite-rich tailings and plant uptake of heavy meals at historical lead mines near Newtownards, Northern Ireland. Transactions of the Institute of Mining and Metallurgy Section B Applied Earth Sciences, Vol. 113, B21-B30.

Nimick, D A, Gurrieri, J T, and Furniss, G. 2009. An empirical method for estimating instream pre-mining pH and dissolved Cu concentration in catchments with acidic drainage and ferricrete. Applied Geochemistry, Vol. 24, 106-119.

Northern Ireland Environmental Agency. 2009. River Basin Management Plans – Groundwater Classification - Mine Workings.

Parsons, M B, Little, M E, Desbarats, A J, Percival, J B, LeBlanc, K W G, Vaive, J E, and Pelchat, P. 2013. Baseline geochemistry and controls on mine drainage composition in the Bridge River Mining District, British Columbia Geological Survey of Canada, Open File 6435.

Pulford, I D, MacKenzie, A B, Donatello, S, and Hastings, L. 2009. Source term characterisation using concentration trends and geochemical associations of Pb and Zn in river sediments in the vicinity of a disused mine site: Implications for contaminant metal dispersion processes. Environmental Pollution, Vol. 157, 1649-1656.

Reimann, C, and de Caritat, P. 2005. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. Science of The Total Environment, Vol. 337, 91-107.

Reimann, C, Filzmoser, P, and Garrett, R G. 2005. Background and threshold: critical comparison of methods of determination. Science of The Total Environment, Vol. 346, 1-16.

Reimann, C, and Garrett, R G. 2005. Geochemical background - concept and reality. Science of The Total Environment, Vol. 350, 12-27.

Rose, A W, Hawkes, H E, and Webb, J S. 1979. Geochemistry in mineral exploration (2nd edition edition). (London: Academic Press.) ISBN ISBN 0-12-596252-5

Runkel, R L, Kimball, B A, Walton-Day, K, and Verplanck, P L. 2007. A simulation-based approach for estimating premining water quality: Red Mountain Creek, Colorado. Applied Geochemistry, Vol. 22, 1899-1918.

Runnels, D D, Shepherd, T A, and Angino, E E. 1992. Metals in water- Determining natural background concentrations in mineralized areas. Environ. Sci. Technol., Vol. 26, 2316-2323.

SedNet. 2004. Contaminated Sediments in European River Basins. SedNet Final summary report.

Sinclair, A J. 1974. Selection of threshold values in geochemical data using probability graphs. Journal of Geochemical Exploration, Vol. 3, 129-149.

Sinclair, A J. 1976a. Univariate analysis. 59-81 in Statistics and data analysis in geochemical prospecting. Howarth, R J (editor). Handbook of Exploration Geochemistry, Volume 2. (Amsterdam: Elsevier Scientific Publishing Company.)

Sinclair, A J. 1976b. Applications of probability graphs in mineral exploration. Special Volume no. 4. (Vancouver, BC, Canada: The Association of Exploration Geochemists.)

Sinclair, A J. 1991. A fundamental approach to threshold estimation in exploration geochemistry: probability plots revisited. Journal of Geochemical Exploration, Vol. 41, 1-22.

Smyth, D. 2007. Methods used in the Tellus Geochemical Mapping of Northern Ireland. British Geological Survey

Stanley, C R. 1987. PROBPLOT. An interactive computer program to fit mixtures of normal (or log-normal) distributions with maximum likelihood optimization procedures. . The Association of Exploration Geochemists.

Stanley, C R, and Sinclair, A J. 1989. Comparison of probability plots and the gap statistic in the selection of thresholds for exploration geochemistry data. Journal of Geochemical Exploration, Vol. 32, 355-357.

Turner, J N, Brewer, P A, and Macklin, M G. 2008. Fluvial-controlled metal and As mobilisation, dispersal and storage in the Rio Guadiamar, SW Spain and its implications for long-term contaminant fluxes to the Donana wetlands. Science of The Total Environment, Vol. 394, 144-161.

UKTAG. 2012. Proposals for further environmental quality standards for specific pollutants.

Wilson, H E. 1972. Regional geology of Northern Ireland (1st edition). (Belfast: Her Majesty's Stationery Office.)

Yager, D B, Mast, M, Verplanck, P J, Bove, D J, Wright, W G, and Hageman, P L. 2000. Natural versus mining-related water quality degradation to tributaries draining Mount Moly, Silverton, Coloradoof the 5th International Conference Acid Rock Drainage, Society for Mining, Metallurgy, and Exploration, 535-547.

8 Appendix

Table A 1: Mineral type and number of occurrence in Northern Ireland (GSNI Database)

Alkali Feldspar	10		Iron/Aluminium Oxide	60
Anhydrite	3		Lignite	21
Apatite	1		Limonite	5
Arsenopyrite	3		Magnetite	8
Azurite	5		Malachite	20
Baryte	60		Manganese	1
Beryl	5		Manganese Oxide	1
Calcite	4		Marcasite	1
Celestine	1		Molybdenite	5
Chalcopyrite	48		Olivine	1
Chloride	16		Pentlandite	1
Chrysotile	1		Pyrite	45
Copper	2		Pyrolusite	1
Corundum	1		Pyromorphite	1
Covelline	1		Pyrrhotite	10
Diamond	1		Quartz	2
Dolomite	11		Rutile	3
Fluorite	2		Siderite	7
Galena	67		Silica (Diatomite)	25
Gold	51		Smithsonite	1
Gypsum	10		Sphalerite	21
Hematite	29		Sphene	1
Hydrozincite	1		Stibnite	1
Ilmenite	1		Topaz	3
Iron	15		Tourmaline	1

Table A 2: List of named mines in the GSNI database

Mine name	Easting	Northing
ANNAGHER	284297	367120
ANNAGHONE COLLIERY	284753	373358
ANTHRACITE MINE	318788	443139
ARCHED MINE	318821	443162
ARDCLINIS MINES	327209	424033
AUGHNAGURGAN MINE	286655	331367
BALLYBOLEY MINE	332758	397343
BALLYCRAIG MINE	288487	439271
BALLYLAGAN MINE	287884	436737
BALLYLIG MINE	317925	409524
BALLYNABARNISH MINE	324613	383844
BALLYNAKILLY FIRECLAY MINE	285278	364650
BALLYVOY MINE	315482	441908
BARROW MINE	279406	380495
BAY MINES	325284	423748
BELLEEK MINE	194666	359397
BIRCH TREE MINE	317590	443724
BLACK PIT MINE	342866	389418
BLACKPARK MINE	315665	441824
BRACKAVILLE MINE	283966	367158
BURLEIGH HILL MINE	339853	389911
CALDWELL MINE	193887	359285
CAPPAGH COPPER MINE	267368	367529
CARGACLOGHER MINE	284568	332819
CARGAN MINE	316974	418388
CARRICKFERGUS/INTERNATIONAL	342853	389519
CARRICKGALLOGLY MINE	298318	328933
CARRICKMORE IRON MINE	316371	442508
CASTLE CALDWELL MINE	200804	363890
CASTLEWARD LEAD MINE	357675	350072
CHAPMAN'S MINE	318803	443046
CLAY MINE	282698	331055
CLEGNAGH BAUXITE MINE	302446	443633
CLONETRACE MINE	317095	410236
COALISLAND COLLIERY	284720	367162
COALPITS MINE	331138	413627
COLLEGE MINE	280675	333305
CONGO COLLIERY	280040	365504
CONLIG/WHITESPOTS LEAD MINE	349151	377094
CORR FIRECLAY MINE	285524	364401
CORREEN MINE	314778	408750
CRAIGFAD MINE	317096	442069
CRATLEY COLLIERY	285532	373423
CREENAGH COLLIERY	283776	365716
CREENAGH FIRECLAY MINE	285242	364854
CREGGAN MINE	293972	317076

Mine name	Easting	Northing
CROMMELIN MINE	314856	420786
CULLINANE MINE	327025	414546
DEEHOMMED IRON MINE	325627	343733
DERRAGHADOAN COLLIERY	279458	365301
DERRY (GORTNASKEA) MINE	283542	366440
DERRY CLAY MINE	283040	366735
DERRY FIRECLAY MINE	283080	366850
DERRY MINE	283022	366675
DERRY PLANTATION MINE	283133	366724
DERRYNOOSE LEAD MINE	279518	331872
DOON COLLIERY	314101	441588
DRUMGLASS COLLIERY	280383	364450
DRUMGLASS NEW COLLIERY	279943	365357
DRUMREAGH MINE	282870	366990
DUNCRUE MINE	339210	389286
DUNGANNON COLLIERY	280389	365396
DUNGONNELL MINES	317819	417094
DUNLUCE AND GLENTASK MINES	290907	440653
ELGINNY MINE	316607	409614
ESSATHOHAN BAUXITE MINE	318900	421991
EVISHACROW MINES	317171	419562
FALBANE COLLIERY	313932	441522
FRENCH PARK MINE	339237	389482
GLEBE MINE	329463	413173
GLENARIFF MINES	321885	420203
GLENARM MINE	328316	414571
GLENRAVEL MINES	315891	419356
GLORE MINE	329317	413493
GOBB COLLIERY	315871	442102
GOLDNAMUCK COLLIERY	314970	441910
GOODMAN'S MINE	318764	443014
GRIFFIN COLLIERY	315669	442049
IRISH HILL AND STRAID MINES	333170	391861
ISLANDMORE MINE	287315	438054
KILLYGREEN MINE	287961	436242
KILMONAGHAN MINE	306061	333245
KNOCKBOY MINES	314585	408510
LAGGLASS COLLIERY	315035	441812
LEITRIM MINE	329727	317531
LEMNAGH MORE MINES	301106	443376
LEWIN COLLIERY	279372	365175
LIBBERT WEST MINE	331120	413665
LISNASTRANE MINE	284035	367408
LURGABOY COLLIERY	281239	364208
LYLES HILL MINE	324587	382942
MAIDEN MOUNT MINE	339166	389778
MOUNT CASHEL MINES	317088	415815
NELSON COLLIERY	316664	442910

Mine name	Easting	Northing
NO.5 ANNAGHER	284333	367176
NORTH STAR COLLIERY	314316	441555
PARKMORE IRON ORE MINES	318488	420798
POLLARD COLLIERY	315509	441972
PORTNAGREE COLLIERY	316368	442479
RATHKENNY MINE	312967	411598
RATHMULLAN LEAD MINE	347509	337571
ROSSBEG COLLIERY	279916	365244
ROSSBEG PIT (DUFFYS)	279677	364781
SALT PANS COLLIERY	313415	441318
SKERRY MINE	313954	418480
SKERRY NO.2 MINE	313986	419292
TENNANT MINE	343134	389321
TROSTAN MINES	318604	424324
TUFTARNEY BAUXITE MINE	315846	417906
TULLYDONNELL COPPER MINE	297695	315461
TULLYNAWOOD MINE	286442	329721
TULLYRATTY LEAD MINE	356563	348520
ULSTER FIRECLAY MINE	284141	366862
URBALREAGH BAUXITE MINE	289263	439624
URBALREAGH IRON ORE MINE	289297	439632
WEST MINE COLLIERY	314846	441874
WHITE MINE	318853	442743
WHITE MINE COLLIERY	313862	441468

Table A 3: List of river basins in Northern Ireland

Ards Strangford
Ballinderry River
Belfast Lough North
Belfast Lough South
Bradoge River
Carlingford Lough North
Castletown River
Drowes River
Fane River
Larne Lough
Lough Foyle East & Benon
Lough Foyle South
Lough Neagh & Peripheral
Lower Bann
Lower Erne
Main River
Moyola River
Neagh Bann
Newry River
North Coast (Skerries)
North East Coast

North Western
Rathlin
River Blackwater
River Bush
River Faughan
River Finn
River Foyle
River Lagan
River Mourne
River Roe
Six Mile Water
South East Coast
Upper Bann
Upper Erne

R PROBPLOT Programme 25/06/2014 – Mark Cave

1 Reading in your data - copy your data from an excel sheet remembering to include a header row, then run the PROBPLOT read program. This should read your data in the correct format for the PROBPLOT program. The program assumes that you are providing the data as log to base 10 transformed values of the original concentration data.

2 Run Section 1 of the PROBPLOT5 program, this sets up the program functions and sorts the data ready for plotting

3 Run Section 2, this prints out the cumulative frequency plot.

4 Run Section 3. Use the cursor to mark on the plot where you think the breaks in the curve occur (up to a maximum of three breaks). Select the point and click moving from low to high percentiles. When finished click on the "Finish" button at the top of the plot.

4 Depending on the number of breaks select either Section 4, 6 or 8. Run the appropriate section and it will plot on the lines which relate to the breaks you have chosen and then it will add a theoretical line of fit to the plot based on the populations you have chosen.

5 Now run sections 5, 7 or 9 depending on the number of populations you have chosen stopping just after the line "lines(fty,ftmn,col="blue")". This runs the mixtools program which refines your first guess to optimise the fit and plots on a blue line with the optimised fit.

6 Now run the last four lines of the section to give you summary statistics of the optimised fit and plots out the density plot with the optimised distributions being shown

7 Run section 10 of the program. This organises the data into two data sets which give the summary statistics on the derived underlying populations after first transforming the data back from their log10 form. The data are saved to the default drive as two csv files called "Original Estimate.csv" and "Optimised Estimate.csv".

8 If you are not happy with the fit and wish to go back and select new starting conditions then rerun the program from the line "### x contains the data being studied####" towards the end of Section 1.

R programs

R program to read data into R by copying and pasting from Excel

```
# program to read in dat for PROBPLOT
# run this one you have copied the data from excel
# remember to include a header row

read.excel <- function(header=TRUE,...) {
  read.table("clipboard",sep="\t",header=header,...)
}

ELS=read.excel()

#this bit puts the column of your data set into variable x
#ready for use in the PROBPLOT program

x<-ELS[,1]

# change the "1" to the appropriate column number
```

PROBPLOT program

```
## Use this data to test out the method
#tst<-rnorm(400,90,10)
#tst2<-rnorm(300,180,15)
#tst1<-rnorm(300,300,30)
#st3<-rnorm(150,50,20)
#x<-c(tst2,tst,tst3)
########################################

#########################################
#########################################
##### SECTION 1 #########################
#########################################
#########################################
library(mixtools)
##########################################
##########################################
######function to calc slope, int , mean and SD
#########################################
dist.vals<-function(x){
  xl <- quantile(x, c(0.25, 0.75))
yl <- qnorm(c(0.25, 0.75))
slope <- diff(xl)/diff(yl)
int <- xl[1] - slope * yl[1]
mn<-(slope*0)+int
sd<-(slope*1+int)-mn
return(c(int,slope,mn,sd))
}
#########################################
#########################################

#########################################
#########################################
### function to get the theoretical fit##
#########################################
fit.line<-function(df){
```

```
  n.pop<-nrow(df)
  n<-vector(length=n.pop)
  smp<-rep(list(list()), 3 )
  # calculated 1000 fits and take a mean
  for (j in 1:1000){
    for (i in 1:n.pop){
      n[i]<-round(1000*df[i,4]/100)
      smp[[i]]<-rnorm(n[i],df[i,1],df[i,2])
    }
    if (j==1) {mat<-matrix(nrow=length(unlist(smp)),ncol=1000)}
    mat[,j]<-sort(unlist(smp))
  }
  # take the mean
  mn<-rowMeans(mat)
  # get the y probability values
  y<-qnorm(ppoints(length(mn)))
  ft<-data.frame(y=y,mn=mn)
  return(ft)
}
###########################################################
###########################################################
###########################################################

### x contains the data being studied####
### This first part plots out the data in the from aof a probability
plot

df<-data.frame(x=sort(x),y=qnorm(ppoints(length(x))))
probs <- c(0.01, 0.05, seq(0.1, 0.9, by = 0.1), 0.95, 0.99)
probs<- c(0.001, probs, 0.999)
qprobs<-qnorm(probs)

# get fitted values for % of data
dat<-data.frame(probs=probs,qprobs=qprobs)
#plot(qprobs,probs)
M<-loess(probs~qprobs,span=0.4,data=dat)
Mnw<-loess(qprobs~probs,span=0.4,data=dat)
##################################################
##################################################
##########END OF SECTION 1########################
##################################################
##################################################

##################################################
##################################################
###########SECTION 2##############################
##################################################
##################################################

plot(df$y, df$x, axes = FALSE, type = "n", xlim = range(c(df$y,
qprobs)),
     xlab = "%", ylab = "concentration")
box()
abline(v = qprobs, col = "grey")
axis(2)
axis(1, at = qprobs, labels = 100 * probs)
points(df$y, df$x,cex=0.4,pty=3)

################################################
```

```
################################################
########### END OF SECTION 2####################
################################################
################################################

################################################
################################################
############ SECTION 3 #########################
################################################
################################################


##### Run this line and choose the breaks using the
######cursor then press finish
coords <- locator(type="p")
##locs stores the locations chosen in locs
locs<-unlist(coords)
## n.pop gives the number of populations
n.pop<-(length(locs)/2)+1
# plot on the chosen break points
pt<-length(locs)/2
apos.x<-locs[1:pt]
apos.y<-locs[(pt+1):(pt*2)]
arrows(apos.x,apos.y*1.2,apos.x,apos.y,col="red")
###############################################
###############################################
######## END OF SECTION 3 #####################
###############################################
###############################################


##########################################
########### SECTION 4 ####################
##########################################
# run this if the number of populations= 2
##########################################
##########################################
pop1<-subset(df,y<=locs[1])
pop2<-subset(df,y>locs[1])
# calculate the population statistics
pop1.vals<-dist.vals(pop1$x)
#add the line to the plot
abline(pop1.vals[1],pop1.vals[2],col="red")
# calculate the population statistics
pop2.vals<-dist.vals(pop2$x)
#add the line to the plot
abline(pop2.vals[1],pop2.vals[2],col="red")
# combine the data
# calculate the percentages
pop.2<-data.frame(mean=c(pop1.vals[3],pop2.vals[3]),
                  SD=c(pop1.vals[4],pop2.vals[4]))
qprobs<-c(locs[1],NA)
pop.2<-cbind(pop.2,qprobs)
pcnt<-predict(M,pop.2)
pc.vals<-c((pcnt[1]*100),(1-pcnt[1])*100)
pop.2<-cbind(pop.2,pc.vals)
ft<-fit.line(pop.2)
lines(ft$y,ft$mn,col="green")
##################################################
```

```
##################################################
####### END OF SECTION 4 #########################
##################################################
##################################################

#################################################
#################################################
####### SECTION 5 ###############################
#################################################
#################################################
# do mixtools model based on estimated parameters
M2 <- normalmixEM(x, lambda = pop.2[,4]/100, mu = pop.2[,1]
                  , sigma = pop.2[,2])
pop.2.opt<-data.frame(mn=M2$mu)
pop.2.opt<-cbind(pop.2.opt,M2$sigma,c(NA,NA),M2$lambda*100)
ft<-fit.line(pop.2.opt)
lines(ft$y,ft$mn,col="blue")
summary(M2) # get a summary of optimised conditions
plot(M2,density=T) # plot the mixture density
out1<-pop.2
out2<-pop.2.opt
################################################
################################################
########## END OF SECTION 5#####################
################################################
################################################


###############################################
###############################################
##########SECTION 6 ###########################
###############################################
###############################################
# run this if the number of populations= 3
pop1<-subset(df,y<=locs[1])
pop2<-subset(df,y>locs[1] & y<=locs[2])
pop3<-subset(df,y>locs[2])
# calculate the population staistics
pop1.vals<-dist.vals(pop1$x)
#add the line to the plot
abline(pop1.vals[1],pop1.vals[2],col="red")
# calculate the population staistics
pop2.vals<-dist.vals(pop2$x)
#add the line to the plot
abline(pop2.vals[1],pop2.vals[2],col="red")
pop3.vals<-dist.vals(pop3$x)
#add the line to the plot
abline(pop3.vals[1],pop3.vals[2],col="red")
# combine the data
pop.3<-data.frame(mean=c(pop1.vals[3],pop2.vals[3],pop3.vals[3]),
                  SD=c(pop1.vals[4],pop2.vals[4],pop3.vals[4]))
qprobs<-c(locs[1],locs[2],NA)
pop.3<-cbind(pop.3,qprobs)
pcnt<-predict(M,pop.3)
pc.vals<-c((pcnt[1]*100),(pcnt[2]-pcnt[1])*100,(1-pcnt[2])*100)
pop.3<-cbind(pop.3,pc.vals)

ft<-fit.line(pop.3)
lines(ft$y,ft$mn,col="green")
```

```
##########################################################
##########################################################
######### END OF SECTION 6 ###############################
##########################################################
##########################################################

##########################################################
##########################################################
########## SECTION 7 #####################################
##########################################################
##########################################################
# do mixtools model based on estimated parameters
M3 <- normalmixEM(x, lambda = pop.3[,4]/100, mu = pop.3[,1]
                  , sigma = pop.3[,2])
pop.3.opt<-data.frame(mn=M3$mu)
pop.3.opt<-cbind(pop.3.opt,M3$sigma,c(NA,NA,NA),M3$lambda*100)
ft<-fit.line(pop.3.opt)
lines(ft$y,ft$mn,col="blue")
summary(M3)
plot(M3,density=T)
out1<-pop.3
out2<-pop.3.opt
########################################################
########################################################
######## END OF SECTION 7 ##############################
########################################################
########################################################


#######################################################
#######################################################
######### SECTION 8 ###################################
#######################################################
#######################################################

##########################################
# run this if the number of populations= 4
pop1<-subset(df,y<=locs[1])
pop2<-subset(df,y>locs[1] & y<=locs[2])
pop3<-subset(df,y>locs[2] & y<=locs[3])
pop4<-subset(df,y>locs[3])
# calculate the population staistics
pop1.vals<-dist.vals(pop1$x)
#add the line to the plot
abline(pop1.vals[1],pop1.vals[2],col="red")
# calculate the population staistics
pop2.vals<-dist.vals(pop2$x)
#add the line to the plot
abline(pop2.vals[1],pop2.vals[2],col="red")
pop3.vals<-dist.vals(pop3$x)
#add the line to the plot
abline(pop3.vals[1],pop3.vals[2],col="red")
pop4.vals<-dist.vals(pop4$x)
#add the line to the plot
abline(pop4.vals[1],pop4.vals[2],col="red")
# combine the data
pop.4<-
data.frame(mean=c(pop1.vals[3],pop2.vals[3],pop3.vals[3],pop4.vals[3])
,
```

```
SD=c(pop1.vals[4],pop2.vals[4],pop3.vals[4],pop4.vals[4]))
qprobs<-c(locs[1],locs[2],locs[3],NA)
pop.4<-cbind(pop.4,qprobs)
pcnt<-predict(M,pop.4)
pc.vals<-c((pcnt[1]*100),(pcnt[2]-pcnt[1])*100,
           (pcnt[3]-pcnt[2])*100,(1-pcnt[3])*100)
pop.4<-cbind(pop.4,pc.vals)

ft<-fit.line(pop.4)
lines(ft$y,ft$mn,col="green")
#########################################################
#########################################################
######### END OF SECTION  8 #############################
#########################################################
#########################################################

#########################################################
#########################################################
######### SECTION  9 ####################################
#########################################################
#########################################################
# do mixtools model based on estimated parameters
M4 <- normalmixEM(x, lambda = pop.4[,4]/100, mu = pop.4[,1]
                  , sigma = pop.4[,2])
pop.4.opt<-data.frame(mn=M4$mu[1:4])
pop.4.opt<-cbind(pop.4.opt,M4$sigma,c(NA,NA,NA,NA),M4$lambda*100)
ft<-fit.line(pop.4.opt)
lines(ft$y,ft$mn,col="blue")
summary(M4)
plot(M4,density=T)
out1<-pop.4
out2<-pop.4.opt
######################################################
######################################################
######### END OF SECTION 9 ###########################
######################################################
######################################################

######################################################
######################################################
######### SECTION 10 #################################
######################################################
######################################################
#now put together the data and export to csv
# first summarise orig and log data
origlog<-cbind(out1[,c(1,2,4)])
names(origlog)<-c("mean","sd","percentage")
optlog<-cbind(out2[c(1,2,4)])
names(optlog)<-c("mean","sd","percentage")
# now convert back from log values
orig<-origlog
orig$mean<-10^orig$mean
orig.low1<-10^(origlog[,1]-origlog[,2])
orig.high1<-10^(origlog[,1]+origlog[,2])
orig.low2<-10^(origlog[,1]-(2*origlog[,2]))
orig.high2<-10^(origlog[,1]+(2*origlog[,2]))
orig<-as.data.frame(cbind(orig[,1],orig[,3],orig.low1,orig.high1,
            orig.low2,orig.high2))
```

```
pops<-row.names(orig)
orig<-cbind(pops,orig)
names(orig)<-c("Population","Mean","Percentage"," -sd"
               ," +sd","-2sd","+2sd")
write.csv(orig,file="Original Estimate.csv",row.names=FALSE)

#now put together the data and export to csv
# first summarise optimised  and log data
# now convert back from log values
opt<-optlog
opt$mean<-10^opt$mean
opt.low1<-10^(optlog[,1]-optlog[,2])
opt.high1<-10^(optlog[,1]+optlog[,2])
opt.low2<-10^(optlog[,1]-(2*optlog[,2]))
opt.high2<-10^(optlog[,1]+(2*optlog[,2]))
opt<-as.data.frame(cbind(opt[,1],opt[,3],opt.low1,opt.high1,
                          opt.low2,opt.high2))
pops<-row.names(opt)
opt<-cbind(pops,opt)
names(opt)<-c("Population","Mean","Percentage"," -sd"
               ," +sd","-2sd","+2sd")
write.csv(opt,file="Optimised Estimate.csv",row.names=FALSE)
######################################################
######################################################
######### END OF SECTION 10 ##########################
######################################################
######################################################
```