LANDCARE RESEARCH
MANAAKI WHENUA

Background soil concentrations of selected trace elements and organic contaminants in New Zealand.

Envirolink Tools Grant: C09X1402

Background soil concentrations of selected trace elements and organic contaminants in New Zealand.

Jo-Anne E Cavanagh, Stephen McNeill, Cecilia Arienti

Landcare Research

Mark Rattenbury

GNS Science

Prepared for Advisory Group:

Regional Council Waste and Contaminated Land Forum
Land Monitoring Forum
Land Managers Group
Ministry for the Environment
Ministry for Primary Industries

November 2015

Landcare Research, Gerald Street, PO Box 69040, Lincoln 7640, New Zealand,
Ph +64 3 321 9999, Fax +64 3 321 9998, www.landcareresearch.co.nz

Reviewed by:	*Approved for release by:*
Ian Lynn Soil Scientist Landcare Research	Chris Phillips Portfolio Leader – Managing Land and Water Landcare Research
Landcare Research Contract Report:	LC2440

Disclaimer

This report has been prepared by Landcare Research for the Advisory Group ((Regional Council Waste and Contaminated Land Forum, Land Monitoring Forum, Land Managers Group, Ministry for the Environment, and Ministry for Primary Industries). If used by other parties, no warranty or representation is given as to its accuracy and no liability is accepted for loss or damage arising directly or indirectly from reliance on the information in it.

Contents

Summary

Project and Client

- This project undertook extensive analysis using spatial databases and existing data to identify key factors influencing trace elements and develop predictive relationships to enable the determination of background concentrations across New Zealand. This report also includes considerations of the requirements to develop and use a database for trace element data and soil properties to aid in the further assessment of background concentrations, and the relationship with other soil quality information. This project was undertaken as part of Envirolink Tools Project C09X1402 for regional councils.

Objectives

- To develop a methodology and determine background concentrations of trace elements and relevant organic contaminants across New Zealand based on existing data
- To establish database requirements for linking trace element data with soil quality data

Methods

- Existing data on selected trace element and organic contaminant concentrations were compiled from regional council soil quality monitoring data and regional studies to determine background soil concentrations. Additional data from a recently completed grid-based soil geochemistry sampling program in Southland and Otago conducted by GNS Science were also used.
- For trace element data, context information for each sampling location was obtained from three spatial databases: the pedological S-Map, fundamental soils data in the Land Resource Information System, and the digital QMAP geological map of New Zealand. Data from only the most recent sampling of a given site was used for subsequent data analysis.
- The statistical package 'R' was used to assess the influence of land use and individual pedological and geological parameters on the key trace elements (arsenic, cadmium, chromium, copper, lead, nickel, and zinc) using regression analysis, after initial testing for spatial correlation of the data for each dataset. Initial predictions of background concentrations from the individual datasets were also compared.
- Final estimates of background concentrations were based on the two datasets (regional council data and GNS Science Southland-Otago data) being combined, and aggregating some land-use classes to provide a larger background dataset that was subsequently analysed using regression analysis, after initial testing for spatial correlation of the data.

- Minerals exploration data from New Zealand Petroleum & Minerals and analytical research data from GNS Science's PETLAB database were used to identify areas of mineralisation for the individual elements.

Results and conclusions

- Land-use effects were evident for most trace elements considered in the regional council dataset, with the effects observed differing for the individual trace elements. Some land-use effects were evident in the GNS Science Southland-Otago dataset although the differences were small. Further, due to the small number of samples (24) in the background class, the robustness of that model (and therefore comparison of the influence of land use) was limited.
- Analysis of datasets indicated that a rock group-based grouping (*Chemical4*, Q-Map) provided the best fit for trace elements data for the regional council data while a geochemically-based lithostratigraphic grouping provided the best fit for the GNS Science Southland-Otago data. This difference between the datasets is attributed to the relatively clustered spatial distribution of sampling points in the regional council dataset compared to the grid-based Otago-Southland dataset.
- After combining the individual datasets, and including data from additional land-use categories, spatial correlation was evident, requiring the use of Generalised Least Squares modelling to develop predicted concentration distributions. A rock-group-based parameter, *Chemical4*, was found to provide the best fit for the combined data and was used to generate predicted background concentration distribution (described by the effective median, 5th and 95th percentile estimates) for the individual trace elements for the individual *Chemical4* subgroups.
- These predictions provide a first-pass estimate of trace element background concentrations across most of New Zealand. Predictions for *Chemical 4* subgroups with few underlying samples (n<30) are considered less reliable and for n<10, unreliable.
- The predicted concentration distributions for many *Chemical 4* subgroups fall within a similar range for a given trace element and could be grouped together to facilitate easier application of this data. The volcanic subgroups basalt, scoria and tuff typically had higher concentrations of Cu, Cr, Ni and Zn while schist and semischist subgroups typically had lower concentrations, although sample numbers in these groups are low.
- Within surficial depositional groupings, such as gravel and sand, there is evidence for a potential influence of elevated concentrations of trace elements arising from upstream erosion of mineralised areas.
- Where the area of a *Chemical 4* subgroup intersects with mineralised areas, identified from minerals exploration and geoscience research data, or depositional zones downstream of mineralised areas, measured trace element concentrations may be higher than the predicted concentration range.
- Provisional ambient concentrations for benzo(a)pyrene in urban areas, provincial towns and rural areas were determined from data from three regions.

- Data on DDT residue concentrations were compiled, although an absence of individual site data did not allow any substantive analysis. While some generalised comments about DDT residue concentrations in relation to land use may be made, DDT residue concentrations are likely to be highly variable within a given land use. Whether action is ultimately required depends on whether any effect might occur. In this context, ecological considerations are the most relevant, including potential bioaccumulation in the food chain.

Recommendations

- Further work is required to enable practical use of the predicted background concentration distributions developed in this project. This could include simplification (i.e. grouping of sub-groups with similar concentration distributions), and/or provision of detailed information via the internet to enable the background concentration distribution of a location of interest to be specifically determined. This could be achieved through a downloadable dataset or an interactive online map. Such a system could be delivered through the LRIS portal (https://lris.scinfo.org.nz/) or the Our Environment website (http://ourenvironment.scinfo.org.nz/home).
- Additional sampling and analysis is required to further develop and refine these estimates of background concentrations of trace elements, particularly in areas for which no or limited data are available.
- Additional sampling and analysis is required to develop more robust estimates of ambient concentrations of polycyclic aromatic hydrocarbons (PAHs) in urban and provincial areas.
- A grid-based approach is recommended for determination of ambient PAH concentrations in urban areas. A combination of grid-based and targeted sampling is recommended outside urban areas.

1 Introduction

Nationally and internationally there is an increased focus on the determination of 'background' concentrations of contaminants for the purposes of managing land. This focus has often been the result of legislative imperatives; for example, Johnson and Demetriades (2011) highlight nine European Commission (EC) Directives driving demand for harmonised geochemical baseline data across European political borders. In New Zealand, the National Environmental Standard for assessing and managing contaminants in soil to protect human health (hereafter referred to as NES), places an increased focus on 'background' soil concentrations, as the NES does not apply 'if contaminants in or on the piece of land of interest are at, or below background concentrations'. In addition, in developing the soil contaminant standards (SCS) for arsenic and cadmium, consideration was given to their background concentrations of (considered as 99th percentile concentration of arsenic and cadmium in soils collected from around the country and thought not to have been affected by anthropogenic activities) (MfE 2011a). For cadmium, the background concentration (once again defined as the 99th percentile concentration of cadmium in soils collected from around the country and thought not to have been affected by anthropogenic activities) is used to define the trigger value for the first tier of the Tiered Fertiliser Management System for Cadmium (MAF 2011). Finally, and of most relevance to the current Envirolink Tools Project 'Background concentrations and soil guideline values for the protection of ecological receptors' (Eco-SGV tools project), the development of soil guideline values for the protection of ecological receptors increasingly uses an 'added risk approach' for naturally occurring elements. This approach was proposed by Crommentuijn et al. (1997) and assumes that species are fully adapted to the natural background concentration and therefore only the added anthropogenic fraction should be regulated or controlled. This approach is used in the Netherlands in the development of intervention values for managing contaminated land, in REACH guidance (EC 2008) on conducting chemical safety assessments for naturally occurring substances, and more recently in Australia, for the development of Ecological Investigation Levels (EIL, SCEW 2013).

However, there is a lack of national guidance on how to determine background concentrations in soils, and how this should be considered in the context of managing land. Furthermore, there is confusion over the term 'background', which has a complex and varied usage in different areas of science or for different purposes (e.g. Matschullat et al. 2000; Reimann & Garrett 2005; Reimann et al. 2005). This includes background as being only naturally occurring concentrations, or being naturally occurring concentrations plus concentrations arising from diffuse anthropogenic contamination.

Previous New Zealand studies on background soil concentrations have used different pedological and geological groupings to define background concentrations (ARC 2001; URS 2003; Tonkin & Taylor 2006, 2007a,b; McDowell et al. 2013) and do not allow for the assessment of inter-regional similarities and differences in background concentrations and the factors influencing these. This in turn, limits the ability to predict likely concentrations in locations for which no data is available. A previous report (Cavanagh 2013) provided an overview of international and national approaches to determining background concentrations, outlined the sources of existing data on trace element concentrations in New Zealand and provided recommendations for a nationally consistent approach to determining background soil concentrations across New Zealand. Specifically, it was recommended that more extensive analysis (including the use of spatial tools such as S-Map) of existing data should be undertaken, to identify key factors influencing trace elements and to identify whether any predictive relationships can be developed. This existing data should include studies on regional background soil concentrations as well as soil quality monitoring data collected by regional councils.

Spatial tools are increasingly used internationally to determine background soils information (e.g. Lado et al. 2008; Diez et al. 2009; Jarva et al. 2010; Cave et al. 2012) or to use background concentration information (e.g. Sheppard et al. 2009). Often geostatistical analyses are undertaken and used to define relevant 'domains' or groupings where background concentrations are similar. Such tools enable the extrapolation of collected data to areas where data have not been collected. Whether the tools are geologically based or soils based appears to depend more on what databases are available, as opposed to a rigorous assessment of the factors controlling background concentrations. For example, in the UK a Soil-Parent Material Model (SPMM) has been developed (Lawley 2009) based on the origin of the parent materials, dominant mineralogy and texture. This is available at a 1:50,000 scale and was used by Ander et al. (2011) alongside mineralisation and historical mining databases to determine 'normal' (background concentrations) across the UK for the purposes of managing contaminated land. In contrast, Sheppard et al. (2009) use a spatial system based on soil classifications, to provide a trace element index to assess the sustainability of Canadian agriculture. Lado et al. (2008) use a geological database and various other databases, including landcover, nightlights (as a measure of urbanisation) and infrastructure databases, to explain heavy metal concentrations across 26 European countries.

The focus of this report is to undertake more extensive analysis using spatial databases and existing New Zealand data to potentially identify key factors influencing trace elements and to identify whether any predictive relationships can be developed. Cavanagh (2013) also suggested that from a longer-term perspective, development of a spatial database as a central repository for all current and future data on trace elements and soil properties should be a priority to aid the further assessment of background concentrations. As the current project involves the collation of existing information, this report includes considerations of the requirements to develop and use a database for trace element data. This report provides the technical background to the determination of background soil concentrations, and guidance on the use of this information alongside the Eco-SGV will be provided towards the end of the project (June 2016).

2 Background

Soil guideline values developed to protect soil biota (Eco-SGVs) provide a useful means to readily assess potential environmental impact. Some soil guideline values already exist, such as within the Timber Treatment Guidelines (MfE 1997) or Biosolids Guidelines (NZWWA 2003), but these are for a limited number of contaminants and are based on inconsistent methodologies. The absence of national Eco-SGVs has resulted in inconsistency and a lack of clarity around protection of ecological receptors in soil, and a lack of focus on ensuring this protection in territorial and regional/unitary council functions.

The Envirolink Tools Project 'Background concentrations and soil guideline values for the protection of ecological receptors' (Eco-SGV tools project) has the following aims:

- Develop nationally agreed methodologies for determining background soil concentrations of naturally occurring elements, and ecological soil guideline values (Eco-SGVs) for the protection of soil biota, such as soil microbes, plants and soil invertebrates.
- Use existing data to determine background concentrations and Eco-SGVs for multiple land-use scenarios.
- Develop clear guidance on applying Eco-SGVs for different purposes to ensure they are applied correctly.
- Identify requirements for a database that enables ongoing input of trace element concentrations and links to existing soil quality databases (e.g. SINDI https://sindi.landcareresearch.co.nz/).

In essence, this project aims to develop Eco-SGVs for the most commonly encountered contaminants and establish agreed methods for derivation such that values can subsequently be developed for other contaminants of concern as needed. Determination of background soil concentrations are included within this project as methodologies for deriving Eco-SGVs may include their use, or they may be used as criteria for ensuring environmental protection (e.g. cleanfill criteria). Further detail on the methodologies proposed for use, and the context for use of background soil concentration information within the Eco-SGV tools project is provided in Cavanagh (2015a). The current report focusses on determination of background concentrations.

Only naturally occurring substances are considered in the context of background concentrations. This includes the range of trace elements, but also applies to some organic compounds, in particular polycyclic aromatic hydrocarbons (PAHs), which can have natural origins (e.g. bushfires). There are numerous terms used to define the 'background' concentrations of chemical substances in soil, including normal, typical, baseline, ambient, characteristic, natural, background and widespread. These terms are often used interchangeably or can be defined differently in different contexts (see Matschullat et al. (2000), Reimann & Garrett (2005) and Reimann et al. (2005) for more detailed discussion).

For the purposes of this report, the following definitions are used:

Natural background – The concentrations of naturally occurring elements derived/originating from natural processes in the environment as close as possible to natural conditions, exclusive of specific anthropogenic activities or sources. The term may also be referred to as the geochemical background. Attributable to mineral content derived from parent materials, and influence of soil-forming processes. This is also the definition of 'background' used in New Zealand Contaminated Land Management Guidance (MfE 2011b).

Ambient background – The concentrations of chemical substances in the environment that are representative of the area surrounding the site that are not attributable to a single identifiable source. This can include contaminants from historical activities and widespread diffuse impacts (e.g. fallout from motor vehicles). Referred to as 'normal' concentrations in the UK (DEFRA 2012).

Baseline – The soil concentrations of chemical substances in a specified location at a given point in time. Baseline concentrations are analogous to natural background concentrations where the specified locality is not influenced by diffuse anthropogenic sources, or ambient concentrations when the specified locality is influenced by diffuse anthropogenic sources. In contrast to ambient and natural background concentrations, baseline concentrations also include concentrations in locations known to be influenced by land use (e.g. agricultural land use).

Threshold – Upper limit of background variation (Reimann & Garrett 2005).

2.1 Related projects

There are two related projects that are currently being undertaken ('Beneficial use of organic waste') or nearing completion ('Land disposal guidelines') for which the determination of background soil concentrations and development of Eco-SGVs have relevance. This section provides a brief overview of the current status of the two projects, and identifies the relationship between the information generated in the Eco-SGV tools project and waste acceptance criteria/soil limits used by these projects.

As consistency in updated soil limits and Eco-SGVs is required to avoid confusion among regulators and industry, it is intended that the Eco-SGV tools project complements, rather than conflicts with this other work. Specifically, it is anticipated that the *application* of waste criteria/soil limits is specified within the particular guidelines, but that the *methodology* or information (e.g. background soil concentrations) developed in this Eco-SGV tools project is used to inform the criteria or limit-setting where these relate to background soil concentrations or protection of ecological receptors.

2.1.1 Land disposal guidelines

The Land Disposal Technical Guidelines were released for public consultation in July 2013 (WasteMINZ 2013) and are due to be published on the Wasteminz website in early 2016 (Nic Quilty, WasteMINZ, pers. comm.). The following is taken from a near-final version provided by Paul Evans (CEO,WasteMINZ).

The Land Disposal Technical Guidelines consider landfills classified into four types:

- Class 4 Landfill – Cleanfill
- Class 3 Landfill – Managed/Controlled Fill
- Class 2 Landfill – Construction and Demolition Landfill or Industrial Waste Landfill
- Class 1 Landfill – Municipal Solid Waste Landfill or Industrial Waste Landfill

Of most relevance to the Eco-SGV tools project are Classes 3 and 4, as no liners are required for these landfills, enabling direct contact of the surrounding soil with the landfilled materials. Class 4 landfills accept materials such as virgin excavated natural materials (VENM), which include soils, clays, gravels and rocks, and limited amounts of inert manufactured materials (e.g. concrete, brick, tiles) and incidental or attached biodegradable materials (e.g. vegetation). The definition of cleanfill states that 'when discharged to the environment clean fill material will not have a detectable effect relative to the background', and regional background concentrations are the specified waste acceptance limits to be used (section 5). Appendix C in the guidelines provides an overview of the development of waste acceptance criteria, which includes consideration of leaching potential, human health exposure, and exposure of ecological receptors, and Appendix G provides Class 4 waste acceptance criteria – as examples of regional background concentrations for key inorganic elements, and specified criteria for selected organic contaminants.

It should, however, be noted that approaches used by regional councils to date for cleanfill criteria have been variable. Specifically, they may be based on background concentrations alone or a combination of background concentrations and Eco-SGVs.

A Class 3 landfill accepts managed/controlled fill materials, which are considered to be predominantly cleanfill materials but also other inert materials and soils with chemical contaminants in excess of local background concentrations, but with specified maximum total concentrations (section 5). Appendix C identifies the exposure pathways, relevant criteria for each pathway (value and source), and the limiting exposure pathway. The final criteria are provided in Appendix F and are a mix of criteria for the protection of human health, ecological receptors, and aquatic receptors.

2.1.2 Guidelines on the beneficial use of organic waste

A guideline to facilitate the beneficial use of organic waste – which includes an update of the soil limits to protect human health and the environment in the Biosolids Guidelines (NZWWA 2003) – is currently being developed through industry and research groups (NZWater, WasteMINZ, Centre for Integrated Biowaste Research (CIBR), and the Land Treatment Collective (LTC)) with an advisory group including the Ministry for the Environment, Ministry for Primary Industries and Ministry of Health. This project has reviewed contaminants of concern (metals, pathogens and organic contaminants) for the application of organic wastes to land to identify the specific contaminants of concern, and relevant existing national and international soil guideline values. A draft guideline has been developed for the project's advisory group and is currently being further developed prior to release for public comment (N. Walmsly, NZWater, pers. comm.). In this new guideline, contaminant limits are set only for the organic waste material, as opposed to waste and soil (in contrast to the Biosolids Guidelines, which set soil limits for contaminants).

3 Trace element concentrations - Methods

3.1 Data

3.1.1 Regional council data

Trace element concentrations have been obtained primarily from regional council soil quality monitoring programmes or specific projects to determine background concentrations in different regions. In addition, data for surface soils (topsoil or AH horizon) was obtained from the National Soils Database.

Various regional councils provided data from soil quality monitoring and specific studies (Figure 1):

- Northland Regional Council state-of-environment (SOE) soil quality monitoring, 2007–2012
- Auckland SOE and specific soil quality monitoring studies (2000–2013), and data from urban parks and reserves (1999–2000; ARC 2001)
- Waikato Regional Council SOE and specific studies, including soil quality monitoring transects (1998–2013)
- Bay of Plenty soil quality monitoring data (1999–2012)
- Hawke's Bay Regional Council soil quality monitoring (2000–2013)
- Taranaki Regional Council sampling (2000–2012)
- Greater Wellington Regional Council soil quality monitoring (1999–2013), and data from urban parks (2003; URS 2003)
- Tasman District Council soil quality monitoring data (2007–2012)
- Marlborough District Council soil quality monitoring, (2000-2014) and specific studies
- Environment Canterbury Regional Council soil quality monitoring as part of the Arable and Pastoral regional monitoring programme (1998–2013 Lawrence-Smith & Tregurtha 2013); background concentration studies (Tonkin &Taylor 2006, 2007a,b)
- Southland Regional Council soil quality monitoring data (2013).

Samples collected for soil quality monitoring were typically collected following Hill & Sparling (2009), whereby approximately 25 subsamples (0–10 cm) are collected along a 50-m transect to form a single composite sample. Some specific studies used a grid sampling process to provide a composite sample comprised of subsamples collected from the centre and points of a grid, around a 10–15 m square (e.g ARC 2001, URS 2003).

Analytical methods included

- extraction by refluxing with a mixture of concentrated perchloric and nitric acid and analysis by flame atomic absorption spectroscopy (FAAS) or graphite furnace atomic absorption spectroscopy (GFAAS)
- extraction by nitric and hydrochloric acid followed by analysis by inductively coupled optical emission spectroscopy (ICP-OES) or inductively coupled plasma mass spectroscopy (ICP-MS)
- non-destructive analysis of soil by X-ray fluorescence (XRF). The relationship between XRF and aqua regia extractions was examined for a subset of data for which both results were available. The relationships developed were used to convert XRF data to equivalent aqua regia concentrations for subsequent analysis where only XRF data was available.

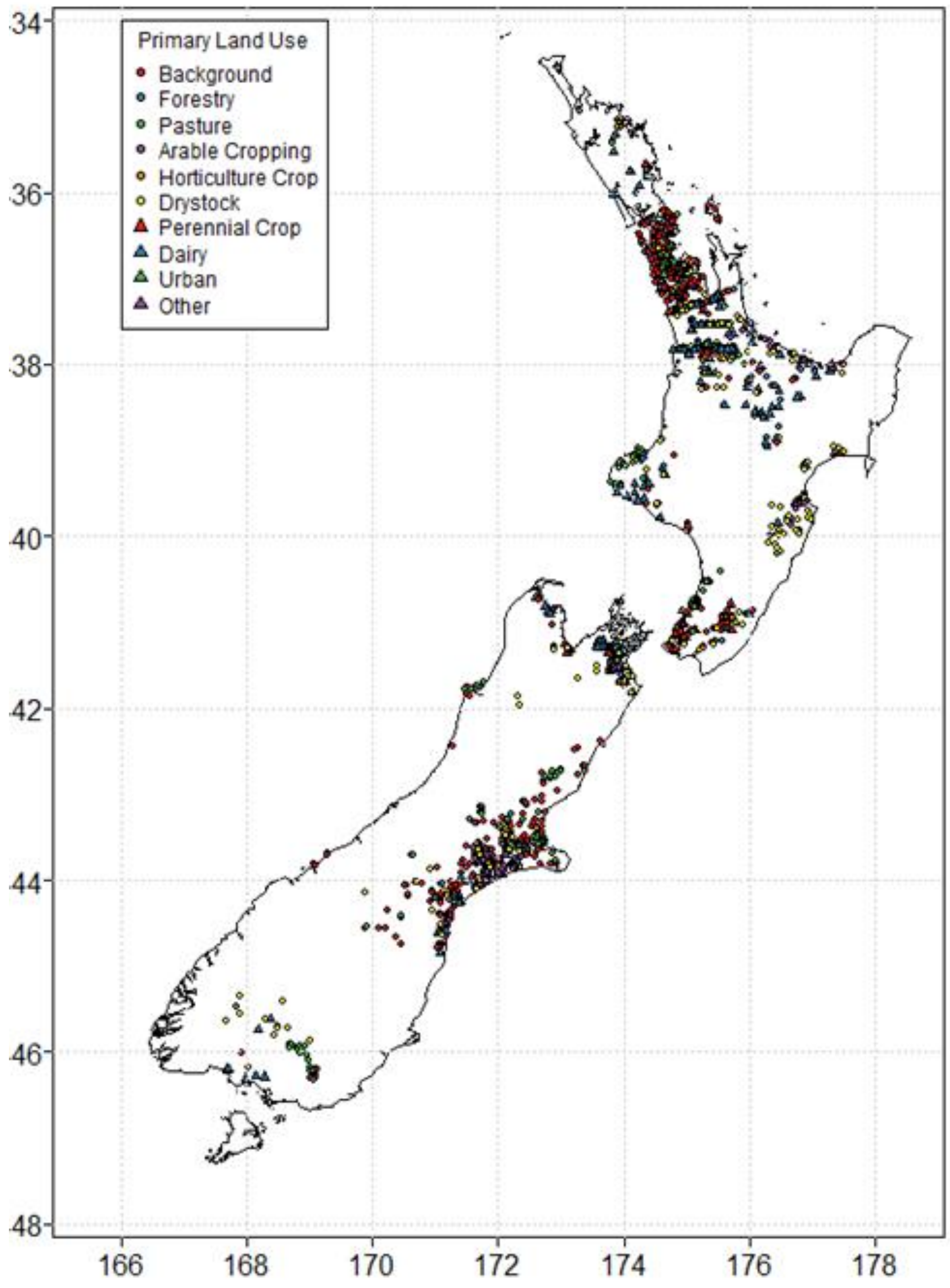

Figure 1 Locations of regional council data, shown as a function of land use.

3.1.2 Grid-based sampling – Southland and Otago

Geochemical data from grid-based sampling undertaken in southern New Zealand by GNS Science were also analysed. The survey collected two samples from 370 sites at 0–30 and 50–70-cm depths spaced on an 8-km grid (Figure 2). These samples were analysed for all major and trace elements using aqua regia partial digest ICP-MS and whole sample XRF (Rattenbury et al. 2014; Martin et al. 2015). Data from the 0–30-cm samples only was used in the current study. These data were initially analysed separately from the regional council data to identify what parameters provided the greatest predictive power.

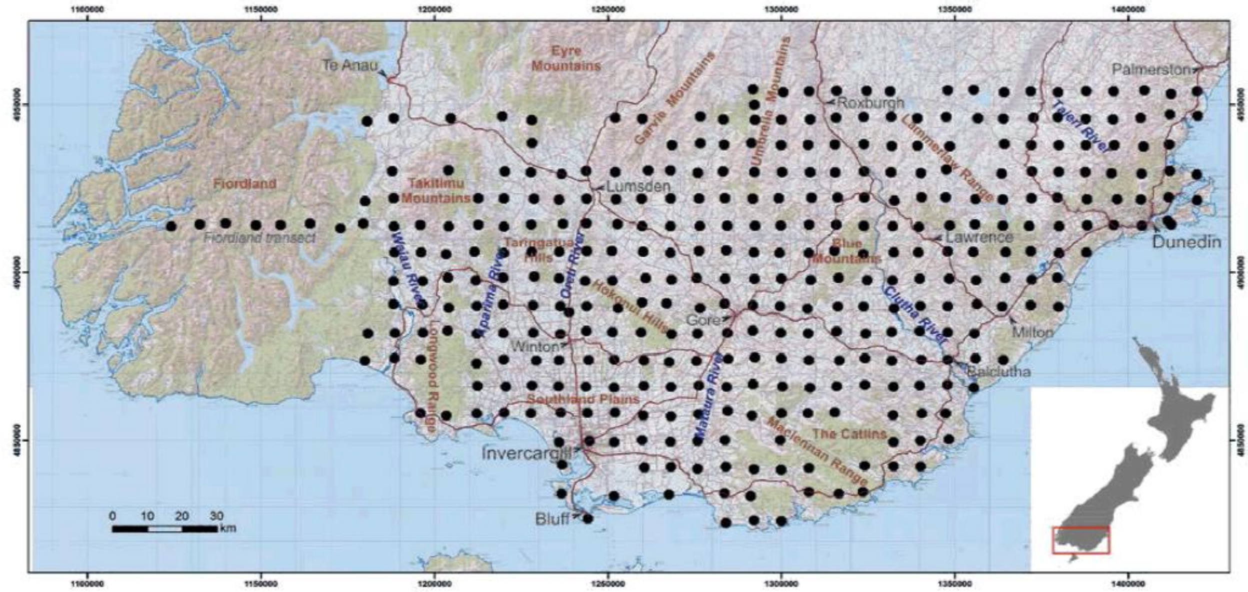

Figure 2 Location of sampling undertaken by GNS Science Southland and Otago data, 8km grid.

3.1.3 Minerals exploration data and research geoanalytical data

To identify areas that may have naturally occurring high concentrations of trace elements, data were also retrieved from the New Zealand Petroleum & Minerals (NZP&M) Open File Metallic Minerals Geochemical Database (Crown Minerals 2009). Trace element concentration data is available for rock, soil and stream sediment samples that have been collected for mineral exploration purposes. Mineral exploration involves collection of many types of rock, soil and sediment samples. This study has restricted its analysis to trace element data that could be useful for identifying elevated concentrations in soil samples discussed in sections 3.1.1 and 3.1.2:

- For rock data – rock chunks from outcrop rock (RC); outcrop rock (OC), subcrop rock (SC) sample and rock grab samples (Grab)
- For soils data – soil sediment taken from one of the near-surface horizons (Soil), and composite sieve soil sample (CompSL).

Data from stream sediment taken from a catchment area (Stream), bulk unsieved stream sediments (bulk) and bulk alluvial samples (alluv) were used only to plot the specific location of samples, and associated concentration, provided in Appendix 3.

In addition, the geoscience research community routinely analyses chemical composition of rocks and minerals to characterise samples and determine their paragenesis. These analytical data are typically published with their interpretation. The Petlab Geoanalytical Database is a Nationally Significant Database hosted by GNS Science (http://pet.gns.cri.nz) that stores physical and chemical property measurements for New Zealand rocks and minerals. The database includes more than 17 000 samples with analyses of one or more major elements and more than 12 000 samples with analyses of one or more trace elements. Around 90% of the analyses were by the X-ray fluorescence (XRF) method with

ICP-MS and atomic absorption (AAS) accounted for the bulk of the remaining methods. About 80% of the analyses were made in the last 30 years from many different laboratories. The sampled material is very eclectic and project based. Samples were commonly taken for their representativeness, that is, to geochemically characterise a rock type, but in many cases the samples were chosen because they were unusual or could establish the extent of variation.

Table 1 Number of analyses of target trace elements from New Zealand samples available in the Petlab Geoanalytical Database (as of September 2015). As, Arsenic; Cd, cadmium; Cr, chromium; Cu, copper; Ni, Nickel; Pb, lead; Zn, zinc

Element	**As**	**Cd**	**Cr**	**Cu**	**Ni**	**Pb**	**Zn**
Number of analyses	2727	189	9749	8018	9843	9755	9538
By XRF method	2697	97	9538	7854	9566	9509	9382

Data from the NZP&M Geochemical Database and PETLAB are presented as maps to illustrate currently identified areas of elevated concentrations in relation to the spatial distribution of soil samples. Specifically, high-end cut-off points (typically the 90th or 95th percentile concentration of the Crown Minerals rock or soil data) were identified for each element and used to define areas in which soil concentrations may be elevated above the predicted background concentrations for that location. The high-end cut-off points were selected on the basis of elevation relative to the surface soil monitoring data with the cut-off points higher for rock data than soil data. Exploration soil data cut-off points tended to be around the 95th percentile of surface soil monitoring data.

3.2 Data collation and databases used

3.2.1 Data collation

Trace element concentration data were provided as spreadsheets, in varying formats and were compiled into a single data file using ‘R’ (R Core Team 2015) to enable statistical analysis. This required conversion of all locational data to a consistent spatial coordinate system (New Zealand Transverse Mercator). The regional council dataset included data from resampling of sites over time, although for determination of background concentrations, data from only the most recent sampling were used. Use of the most recent data provides the greatest likelihood of identifying whether land-use effects are evident. These data were extracted using R, and subject to some initial data checking. For example, data from Auckland urban sampling locations, such as high traffic and industrial areas, were excluded from analysis. Some individual data points were in fact replicate samples from the same location; where this occurred, concentrations were averaged. In a number of instances these replicate data points represented resampling of sites with anomalously high concentrations found during the study on background concentrations in the Auckland region. In some other cases, sites were excluded due to anomalous concentrations related to anthropogenic use, for example a site in the Tasman region received wastewater and was observed to have high concentrations of trace elements; another orchard site was excluded due to apparent contamination with arsenic and lead, based on previous analysis (Cavanagh 2015).

3.2.2 Land use

Land use classification was based on observations made at the time of sampling for the soil datasets, and required consistent categorisation prior to analysis (Table 2 and also section 6.4 for further discussion on land use categorisation). The Land use category 1 was used for determining background concentrations, with categories 2 and 3 provided to illustrate the different land use or land covers captured within the different land use 1 categories.

Table 2 Land use categorisation for describing trace element concentrations, the number of sites in each land use

Land use category 1	Land use category 2	Land use category 3	Comment on land use
Dairy	Dairy	Dairy	
Dry stock	Sheep and beef	Sheep	Includes sheep and beef, deer, goat and is likely to be a mix of intensive and extensive systems. Ideally intensive and extensive systems could be identified with extensive (low input) systems including lifestyle blocks
		Beef	
	Deer	Deer	
	Other	Other	
Pasture	Unspecified pasture, pasture seed crops	Unspecified pasture, pasture seed crops	
Horticulture[1]	Horticulture	Crop type	Market gardens, vegetable crops
Arable cropping[1]	Arable cropping	Crop type	Includes grain crops, hay, fodder crops
Perennial crop	Orchard	Crop type	Stonefruit, berry fruit, kiwifruit, grapes
	Vineyard		
Forestry	Plantation	Tree type	
Undisturbed[2]	Indigenous forest, native scrub	Indigenous forest, native scrub	
	Reserve in non-urban areas	Reserve in non-urban areas	
	Native tussock, not used for grazing	Native tussock, not used for grazing	
Urban	Parks and reserves	Parks and reserves	
	Other	Specify	
Other		Specify	

[1] Arable crop and horticultural crop categories can overlap, e.g. potatoes can be considered as either.

[2] land anticipated to have not been disturbed, or minimally disturbed by anthropogenic land use.

Land use categorisation for the GNS Science Southland-Otago dataset was also based on observations at the time of visitation, although a slightly different categorisation was used. The categories and number of sites in each category is shown in Table 3.

Table 3 Categories and number of sites in the individual land-use categories in the GNS Science Southland-Otago dataset

Land use	Number of sites
Undisturbed	27
Cropping	5
Forestry	40
Other	3
Pasture	294
Urban	8

3.2.3 Spatial databases

The site-based trace element data was complemented with data from four GIS-based polygon spatial databases that regionally and/or nationally delineate areas that are interpreted to have distinct common pedological and/or geological properties. Underlying geology is generally regarded as a major contributor to the geochemical signals in soils and surficial material. The rocks are composed of minerals and these have defined chemical compositions, most commonly as silicates but many minerals naturally contain the trace elements of interest in this study. Because of this, the parameters extracted from the different databases provide some assessment of the geological origins of the soil. The databases used and the parameters extracted from each database are given below (further details are provided in Appendix 1):

- **LRIS** (The Land Resources information System; http://lris.scinfo.org.nz/) is a means for the public to access environmental data held by Landcare Research. Data layers available include NZLRI fundamental soil layers (FSLs), vegetation data layers, and land cover. The NZLRI (FSL) is a spatial database that describes land on the basis of five characteristics including rock type. The FSLs are based on 16 key attributes for soil selected through a consultation process with stakeholders. These attributes fall into three groups: soil fertility/toxicity, soil physical properties (particularly those related to soil moisture), and topography/climate. Data is mapped at a 1:50 000 scale.

  Top-rock is a parameter that describes the near surface lithology (rock type) and was used to investigate variation in trace element composition in different soils in the current study. Given the number of classifications of *top-rock* (Appendix 1), the classifications provided in the FSL were grouped into ultramafics, igneous, surficial, weak sedimentary rock, strong sedimentary rock, and metamorphic rocks to facilitate analyses. *Soil order* was also extracted from the FSL and used in analyses.

- **S-Map** (http://smap.landcareresearch.co.nz/home) is a spatial database for New Zealand soils designed to provide quantitative soil information for modellers and the best-available soil data for land managers and policy analysts (Lilburne et al. 2012). S-Map includes linkages to the National Soils Database (NSD) and QMAP, a geological spatial database developed by GNS (see below). Data is mapped at a 1:50 000 scale or finer in some locations.

  The *rock-type-of-fines* parameter describes the rock class of the fine (<2 mm) soil material (to a depth of 100 cm). For organic soils, a *rock-type-of-fines* attribute is only defined for soils that have a significant mineral content.

- **QMAP** (http://www.gns.cri.nz/qmap) is a database of geological maps developed by GNS Science over the period 1993–2012 (Rattenbury & Isaac 2012). This includes a national seamless GIS component based on the 21 published geological maps at 1:250 000 (Heron 2014). The key component of the QMAP for this project is the geological units polygon layer. This layer defines the extent of discrete mapping entities that are primarily based on a combination of rock type and emplacement age. Each geological unit has a degree of lithological homogeneity placed on it, in some cases embodied in a formal stratigraphic name. Logically, each unit should have a signature geochemical composition based on the proportion and chemical composition of its constituent minerals and this signature is commonly reflected in their overlying soils.

  The geological units polygon layer has many attribute fields that describe the composition, age and stratigraphic affiliation of each unit. For this study, the lithological fields, *rock-group* and *rock-class* were used, along with six derivative fields that provide some delineation based on differences in source material and associated chemistry (*Chemical1-5*). These terms *rock-group* and *rock-class* are hierarchical in that *rock-class* contains highly generalised rock types (for example, ‘mafic extrusive’), while *rock-group* can be a very specific (‘basalt’). Some rock types such as ‘sandstone’ (*rock-group*) and its generalised equivalent ‘clastic sediment’ (*rock-class*) are very broad in a lithological and mineralogical sense. These broad categories can be chemically diverse depending on the nature of their source material. To counteract this chemical diversity in some *rock-group* and *rock-class* categories, six derivative fields were constructed based on the high level stratigraphic affiliation, age and lithological composition of each stratigraphic unit. *Chemical1* has the most detailed stratigraphic affiliations (43 categories) based on Mortimer et al. (2014), C*hemical2* is more generalised (24 categories) and *Chemical3* even more so (9 categories). *Chemical5* is based on *Chemical1* but subdivides the Miocene and younger sedimentary rocks and sediments (Maui and Pakihi supergroups, Mortimer et al. 2014) on a five-part regional basis forming 51 categories. *Chemical4* is based on *rock-group* but subdivides the Miocene and younger sedimentary rocks and sediments (Maui and Pakihi supergroups, Mortimer et al. 2014) forming 72 categories and *Chemical4b* is a variant on *rock-class* where the broad ‘clastic sediment’ field has been subdivided on a three-fold, high-level stratigraphy-age basis to create 14 categories.

3.3 Data analyses

Statistical analyses were undertaken using R version 3.0.2 (R Core Team 2015).

3.3.1 Preliminary analysis

The regional council dataset and the GNS Science Southland and Otago dataset were initially analysed separately to determine whether different information was provided by the respective datasets and whether it was appropriate to combine the datasets to provide a larger dataset to develop predictive models. Linear regression models were developed for arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), lead (Pb), nickel (Ni), zinc (Zn) with the land use and the variables shown in Table 3 for the regional council dataset and the GNS Science Southland-Otago dataset. No interactions between explanatory variables were allowed, since the intersection of, for example, *land use* and *top rock* is incomplete, so a model could only be built for a small subset of the dataset. Thus, the various factors are additive, so the effect of land use and the various explanatory factors can be considered separately. The models follow the form:

$$\text{Log Concentration} = a + b \text{ Land use1} + c \text{ parameter} \qquad (1)$$

The intercept, **a,** can be interpreted as the log-concentration for the reference levels of land use and the specific parameter being assessed (e.g. *soil order*, *top-rock*, etc, Table 3). The coefficients **b** and **c** reflect the change in log-concentration for a land use or parameter other than the reference level. It should be noted that the choice of the reference level for the land use and the parameter of choice is arbitrary; choosing a different reference changes the values of the coefficients in the model, but does not alter the predicted values or their accuracy.

Regression analyses were undertaken using the mean, and median and 95th percentile using quantile regression (Koenker 2005).

Table 4 Reference levels for regression models

Parameter	Reference condition
Land Use	Background
Soil Order	Brown
Top Rock	Igneous
Rock Group	Breccia
Rock Class	Clastic sediment
Chemical1	BrookStT (Brook Street Terrane)
Chemical2	BrookStT (Brook Street Terrane)
Chemical3	AustralSupCenT (Austral Superprovince central Terrane)
Chemical4	Breccia
Chemical5	BrookStT (Brook Street Terrane)
Chemical4b	Clastic sediment

Model selection

The linear regression models were primarily compared using the Akaike information criterion (AIC). The AIC is a measure of the relative quality of statistical models for a given set of data that balances the goodness-of-fit of a model and its complexity (Burnham & Anderson 2002), with the preferred model being that with the lowest AIC. To use the AIC, the same number of samples must be used in each model, thus the datasets for *soil order, rock group, rock class* and *Chemical 1-5* were pruned to the same number of samples. However, the AIC was not used to compare the models involving *rock-type-of-fines* and *top-rock* since the latter factors have fewer samples in them. Instead, the root mean square (RMS) error was used. The R-squared value (in percent) was also determined.

The $\exp((\mathrm{AIC}_{\min} - \mathrm{AIC}_i)/2)$ can be interpreted as the relative probability that the *i*th model minimises the (estimated) information loss (Burnham & Anderson 2002) and can be used to determine whether there are significant differences in the predictive ability of different models. Generally speaking, if the AIC difference is less than two units there is no substantial difference in information lost between the two models.

Land-use effect

To determine the effect of land use for the regional council data, the model with the smallest AIC was used to generate the log-concentrations for the different land-use classes over and above the effect of the undisturbed land-use class. The 95% confidence interval of the land-use coefficient was then used to determine which land-use classes were significantly different from background. Specifically, if the 95% confidence interval overlaps with 0, the trace element concentrations for the specific land use are not significantly different from background. Table 5 provides a summary of the number of samples in the undisturbed land use category and the total number of data points in other land use categories for each element.

Table 5 Summary of all data points, and the number of samples from the undisturbed land use category for the most recent sampling at a given location (recent dataset) for the Regional Council dataset

Element	Number of samples	Number of samples from undisturbed land use category
Arsenic	1287	256
Copper	1480	289
Cadmium	1876	286
Chromium	1480	288
Nickel	1479	288
Lead	1490	287
Zinc	1515	289

For the GNS Science Southland-Otago dataset, land-use classes were aggregated as shown in Table 6 because of the low numbers in the Undisturbed land-use category (27, Table 3).

Table 6 Aggregation of land-use classes for the GNS Science Southland-Otago dataset

Element	Aggregated land use	N
Arsenic	Aggregate all classes, except urban	338
Copper	Aggregate all classes, except urban and cropping	300
Cadmium	Aggregate undisturbed plus 'other' plus 'forestry'	65
Chromium	Aggregate all classes, except urban	338
Nickel	Aggregate all classes, except urban	338
Lead	Aggregate all classes, except urban	338
Zinc	Aggregate all classes, except urban	338

Predicted background concentrations and initial model comparison

Once the relevant 95th quantile regression model for a given element has been selected based on the AIC, trace element concentrations for the preferred explanatory variable for the undisturbed land use (regional council dataset) or aggregated land use (GNS Science Southland-Otago dataset) were predicted and compared.

3.3.2 Final analysis

The preliminary analyses showed that there was no bias in the predictions developed from the different datasets (see section 4.3). Thus, to maximise the geographical spread and therefore predictive ability of the analysis, the two datasets were combined for further analysis. Based on analysis of the land-use effect (section 3.3.1, 4.1.1, 4.2), land-use categories for the regional council dataset were aggregated for the different elements as shown in Table 7, and for the GNS Science Southland-Otago dataset as shown in Table 8. The dataset was pruned so that there was common data involving *rock class, rock group, soil order*, and the Chemical variables. The S-map r*ock-type-of-fines* and FSL *top rock* variables were not included in this set since we did not have complete coverage over a sufficiently large set of samples. This dataset then underwent further analysis.

Table 7 Land-use aggregation for regional council dataset

Element	Background aggregation rule	N
Arsenic	Combine all data except horticulture, arable, and urban	1048
Copper	Combine undisturbed, urban, forestry, and drystock	766
Cadmium	Combine undisturbed and forestry	393
Chromium	Combine all data	1474
Nickel	Combine all data except urban	1359
Lead	Combine all data except urban and horticulture	1301
Zinc	Combine undisturbed and forestry	446

Table 8 Land-use aggregation for GNS Science Southland-Otago dataset

Element	Background aggregation rule	N
Arsenic	Combine all data, except cropping and urban	334
Copper	Combine all data, except cropping	342
Cadmium	Combine undisturbed, ‘other’ and forestry	65
Chromium	Combine all data	347
Nickel	Combine all data	347
Lead	Combine all data except urban and cropping	334
Zinc	Combine undisturbed, forestry, and ‘other’	65

Spatial correlation

A common assumption made in estimating the model coefficients in equation 1 is to assume that the different samples are independent and identical in their distribution. The mathematical solution to the problem of estimating the coefficients is then quite simple. When data are spatially referenced, this effectively requires samples to be uniformly spread across the country (either randomly or over a regular grid). Clearly, this is not the case in this study; some regions are heavily sampled while others have relatively few samples. In addition, gathering more samples does not necessarily contribute independent new information, for if two points are closely spaced, one might expect the log-concentration of the sample values to be closely related. One way to quantify this spatial association is to estimate the spatial autocorrelation as a function of point-to-point distance, or equivalently the semivariance (the two are linked by a mathematical relationship).

Figure 3 shows plots of the semivariance of arsenic by point-to-point distance. At very small point-to-point distances the log-concentration is strongly correlated, while at point-to-point distances greater than a critical distance (for arsenic, about 32 km) the two points are essentially independent. The assessment of whether or not spatial correlation is evident in the data for a given element is not straightforward, since it depends on the type of non-linear spatial model chosen to represent the correlation structure, and the estimated parameters of the chosen model. For each element, the model was chosen that gave the best fit to the empirical spatial semivariance plot, and the parameters were estimated using maximum likelihood.

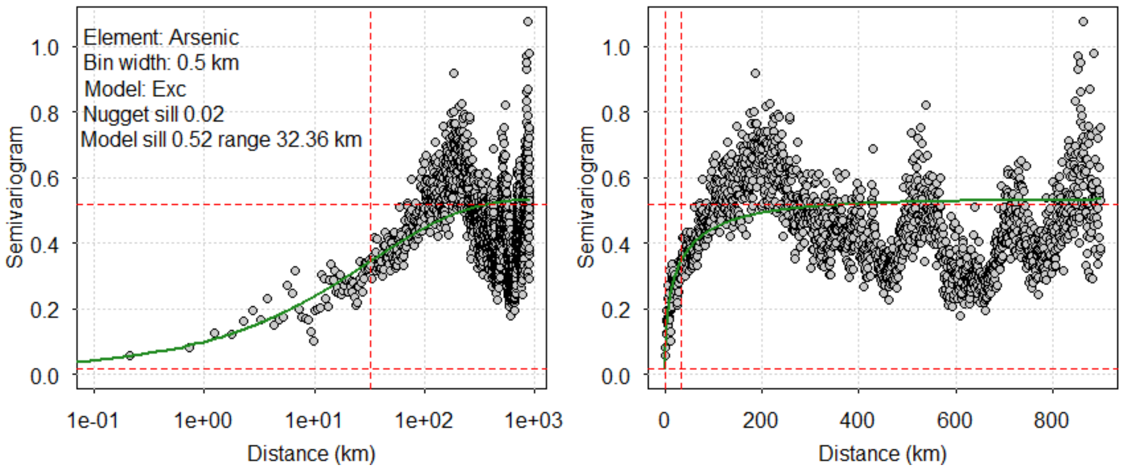

Figure 3 Plots of the semivariance of arsenic using the combined data, as a function of point-to-point distance. The plot on the left has the point distance in log-units, while the plot on the right uses linear units.

The interpretation of the spatial semi-variogram is not simple, since the empirical data seldom follows the 'best' model at very short or very long distances (or both), and the interpretation attempts to balance the conformance of the data to the model with an assessment of the range distance and the relative sizes of the model and nugget sills. For example, the arsenic data follows the model shape well (but poorly for distances over 100 km), and the relative size of the model and nugget sill (0.52 versus 0.02) and the moderate range (32.4 km) suggest that there is some evidence for spatial correlation (Table 9). By contrast, the fit for copper data is relatively poor for short distances, and the relatively short range (2.3 km) suggests no compelling evidence for spatial correlation (Table 9).

Table 9 Interpretation of spatial correlation for each element

Element	Range (km)	Sill	Nugget	Interpretation
Arsenic	32.4	0.52	0.02	Evidence of spatial correlation
Copper	2.3	0.44	0.10	Little evidence of spatial correlation
Cadmium	10.2	0.49	0.19	Little evidence of spatial correlation
Chromium	25.3	0.71	0.00	Evidence of spatial correlation
Nickel	13.6	0.65	0.00	Little evidence of spatial correlation
Lead	3.1	0.28	0.00	Little evidence of spatial correlation
Zinc	128	0.50	0.25	Little evidence of spatial correlation

The observation that spatial autocorrelation is significant with the combined datasets but not significant with the regional council and GNS Science datasets alone evidently comes from the added power of the two datasets, plus the effect of aggregating different land-use classes to form an estimate of the background concentration. The evidence for spatial correlation makes it necessary to account for the spatial autocorrelation in subsequent regressions. Failing to account for spatial autocorrelation will likely result in over-optimistic predictions (i.e. the estimated confidence intervals will be too narrow). If no elements in the study showed compelling evidence of spatial correlation, it would be reasonable to assume that the points were independent, and a much simpler regression procedure could be employed. In this study, at least two of the elements (arsenic and chromium) showed evidence of spatial correlation, so a revised regression procedure was required.

A key point is that taking spatial correlation into account, even when there is little evidence for it, is benign; the regression procedure simply takes longer to complete, but the estimated values are correct. The reverse procedure (treating the points as independent when they are not) is not valid, since in that case the regression procedure assumptions are invalid (Bain & Englehardt 1992).

Model development

The model for concentration is of the form

$$\text{Log Concentration} = a + b \text{ parameter} + E \qquad (2)$$

Where **log concentration** is the log-transformed concentration of the trace element, **a** is the overall mean log-concentration level of the element for the reference level of the single explanatory parameter (*Rock Class, Rock Group, Soil Order*, or the *Chemical* variables), **b** is the change in log-concentration of the element when the explanatory variable is changed from the reference level to another level. Finally, E is the uncertainty, which is Gaussian with a mean zero and variance σ^2_{TE}. The regression analysis yields values for **a** and **b**.

The model for log-concentration is the same whether or not spatial autocorrelation is evident in the data (i.e. whether or not the sample points are independent); however, the solution method is different. Where the samples are not independent (as is the case here), we use Generalised Least Squares (GLS), which is a generalisation of the ordinary least squares solution procedure, intended to take into account the point-to-point correlation. The detailed mathematical steps for this solution method are beyond the scope of this report, but they are described in detail in Pinheiro and Bates (2000). This approach has been used previously for modelling soil carbon (Beare et al. 2014).

The GLS method requires that the response residuals (the difference between fitted and actual log-concentration) is Gaussian distributed, which is a good assumption. This also means the mean and the median of the log-concentration distribution are coincident, and that the 95% prediction interval can be calculated. The 95% prediction interval provides the log-concentration interval within which you would find a predicted concentration value with a 0.95 probability, and thus provides an estimate of the spread of predicted concentration distribution (i.e. the 5th and 95th percentile). Specifically, the 95% prediction interval was inferred using the predicted mean and the estimated standard error approximated using the delta method (Oehlert 1992).

Data processing using generalised least squares (GLS) can be slow, depending on the amount of data. To minimise processing time, and to test the effect of correcting for spatial autocorrelation, models were first fit assuming samples were independent, with the best fit model selected as described in section 3.3.1. *Chemical4* was most often the favoured model (Table 10) and therefore was used to generate predictions from the GLS model directly.

Table 10 Summary of the favoured models based on the median and 95th percentile concentrations for the aggregated land use regional council and GNS Science Southland-Otago dataset, selected using AIC criterion

Element	Favoured variable, 50% quantile model	Favoured variable, 95% quantile model
Arsenic	Soil Order	Chemical4
Copper	Chemical5	Chemical5
Cadmium	Chemical4	Chemical5
Chromium	Chemical4	Chemical4
Nickel	Chemical4	Chemical4
Lead	Soil Order	Chemical4
Zinc	Chemical4	Chemical4

4 Trace element analyses - results

4.1 Preliminary analysis - Regional Council data

The distribution of concentrations on a land use and regional basis is shown in Figures 4–6.

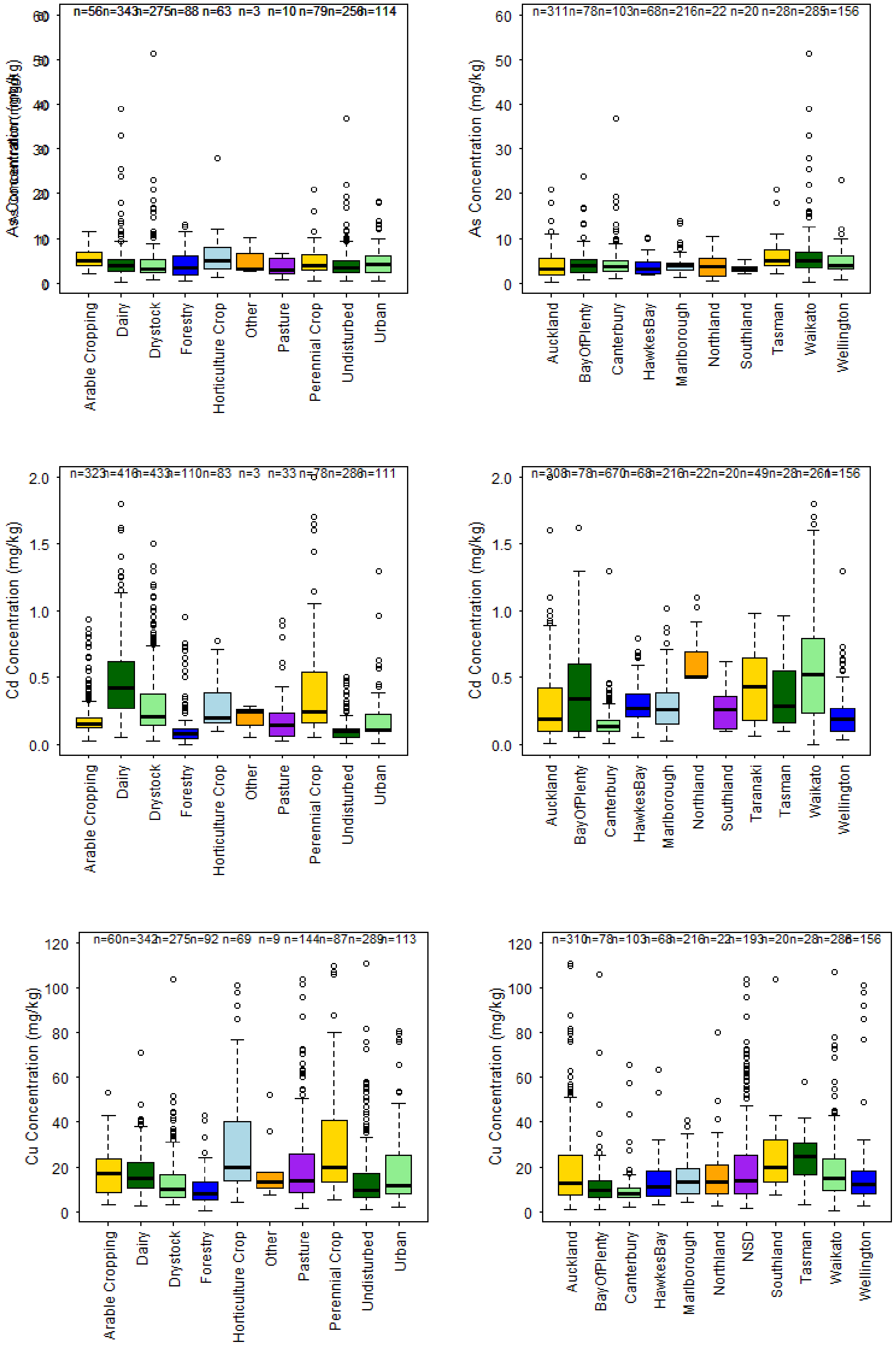

Figure 4 Boxplots of Arsenic (As), cadmium (Cd) and copper (Cu) concentrations from the regional council dataset grouped by land use and region. Whiskers represent maximum or minimum observations within 1.5 × inter-quartile range; outlying data shown as open circles.

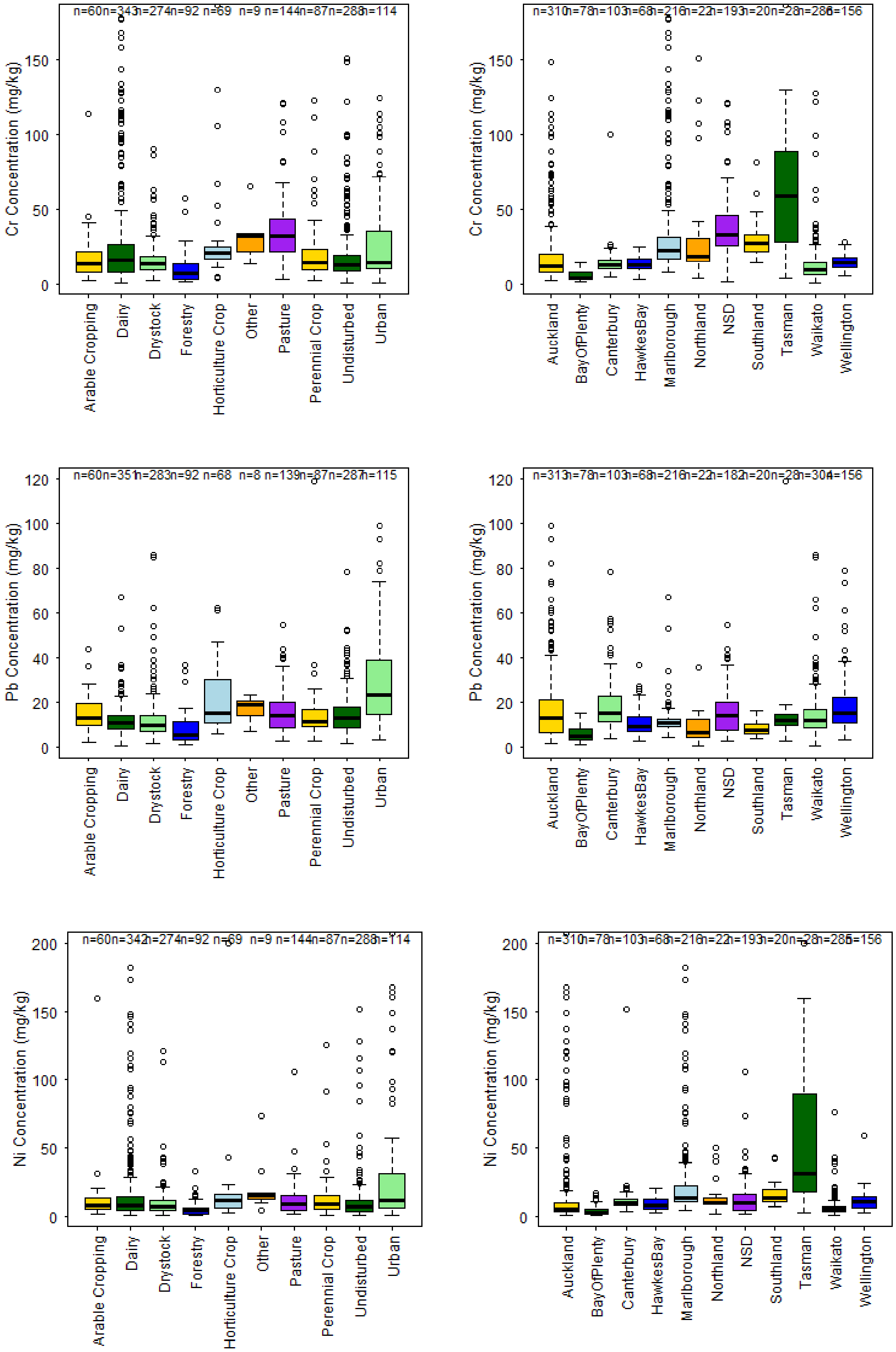

Figure 5 Boxplots of chromium (Cr), lead (Pb) and nickel (Ni) concentrations from the regional council dataset grouped by land use and region. Whiskers represent maximum or minimum observations within 1.5 × inter-quartile range; outlying data shown as open circles.

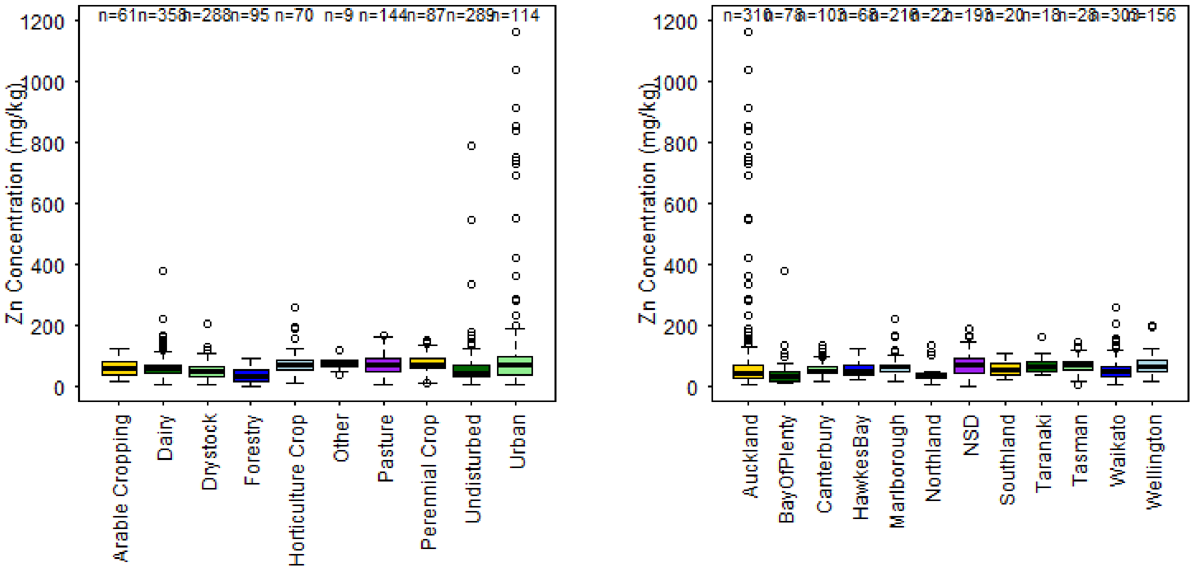

Figure 6 Boxplots of zinc (Zn) concentrations from the regional council dataset grouped by land use and region. Whiskers represent maximum or minimum observations within 1.5 × inter-quartile range; outlying data shown as open circles.

Table 11 provides a summary of the favoured models for each element. *Chemical4*, based on *Rock Group*, was most commonly the favoured model on the basis that the AIC was consistently the smallest. If RMS error were to be used, then *Rock-type-of-fines* would have been selected (Appendix 2).

Table 11 Summary of the favoured models based on the conditional mean and 95th percentile for undisturbed land use for trace elements from regional council datasets, selected using the AIC criterion

Element	Mean	95th
Arsenic	Soil Order	Chemical5
Cadmium	Soil Order	Chemical4
Chromium	Chemical4	Chemical4
Copper	Soil Order	Soil Order
Lead	Soil Order	Soil Order
Nickel	Chemical4	Chemical5
Zinc	Chemical5	Chemical4

4.1.1 Effect of land use

Figure 8 illustrates the land-use effect using the model with the smallest AIC for the 95th quantile model (Table 11). There were significant differences in the concentration of the different elements with land use. The elevated concentration of As in arable and horticultural crops suggests an influence from the historic use of lead arsenate. This influence was also evident by the significant elevation of Pb in horticultural land. The elevated concentration of Cu in perennial, arable, and horticultural crops is consistent with the usage of copper-based fungicides. Significantly elevated Cu concentrations were also observed for dairy and pastoral land uses, and may be attributable to use of Cu fertilisers on pastoral land. The elevated concentration of Cr for various land uses was unexpected, and probably reflects the location the land use, rather than the land use itself. Similarly, the lower concentrations of all elements except As in forestry land are anticipated to be a consequence of the location of the land use rather than the land use itself. Elevated concentrations of Pb in urban environments may indicate lead-based paints or residual leaded petrol. High Zn concentrations were present in some background samples from Auckland (Figure 7), so the data were reanalysed excluding the Auckland data (Figure 8). This revealed some significant land-use effects, with elevated concentrations observed for most land uses except undisturbed and forestry. This may reflect the use of Zn fertilisers, in addition to the use of Zn for the treatment of facial eczema.

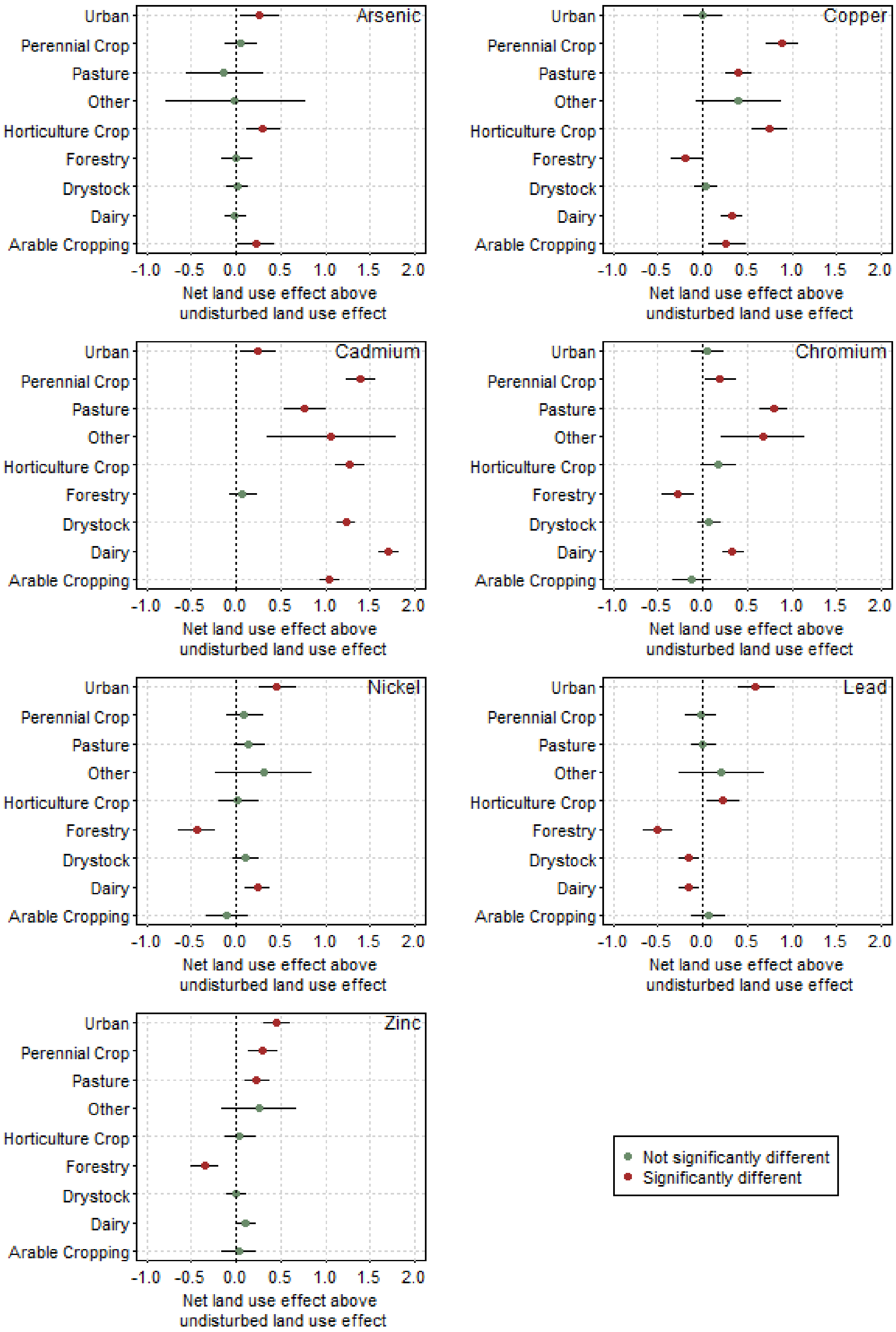

Figure 7 Summary of the difference in trace element concentrations of different land uses in relation to the undisturbed land use class for the regional council dataset. 0.0 represents no difference from undisturbed land use, lines represent the 95% confidence interval of the mean.

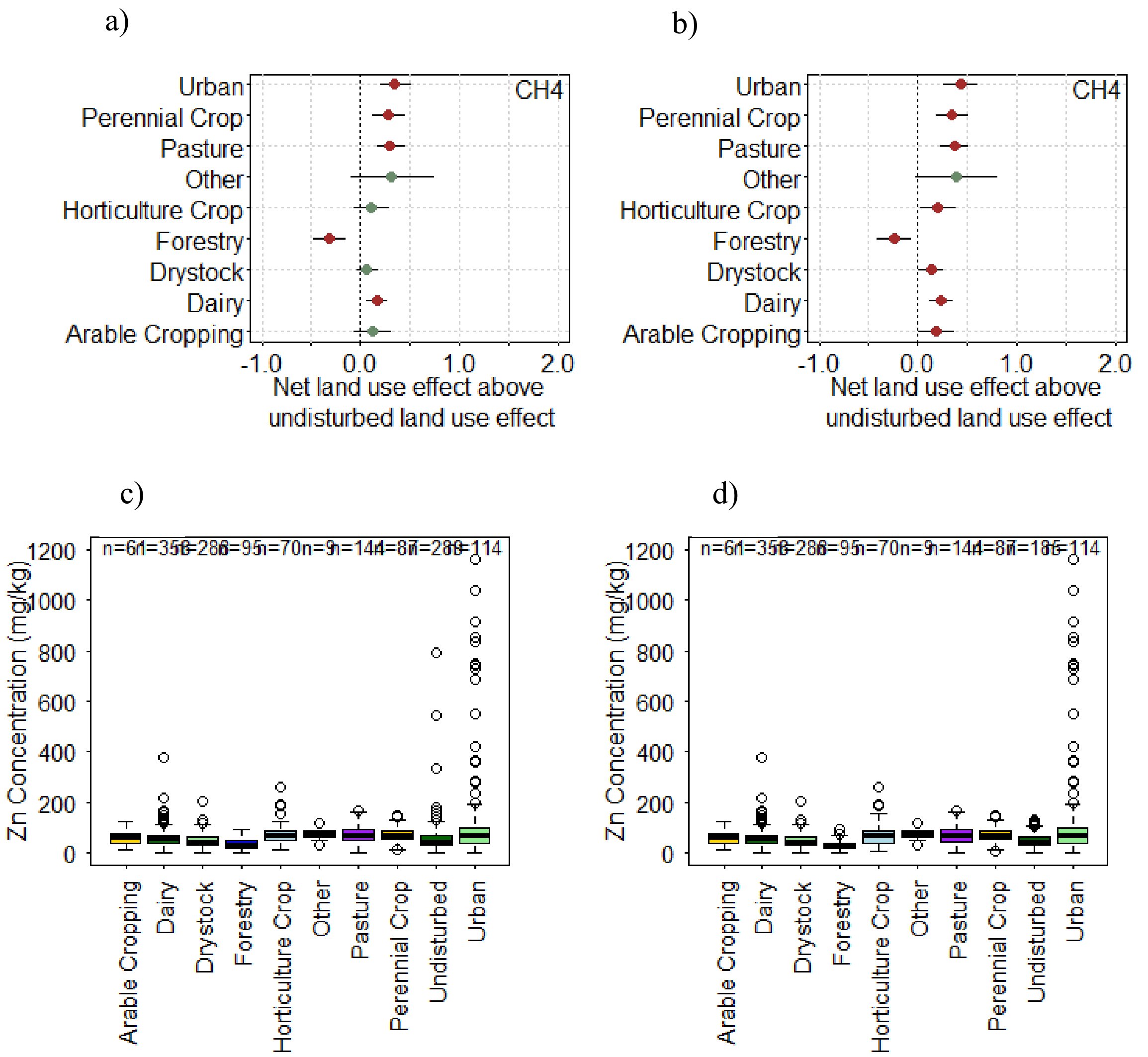

Figure 8 Summary of the difference in predicted zinc (Zn) concentrations of different land uses in relation to the undisturbed land use class based on the chemical 4 (CH4) models for a) all data b) with Auckland undisturbed data removed. 0.0 represents no difference from undisturbed land use, lines represent the 95% confidence interval of the mean. Also shown are boxplots of zinc (Zn) concentrations from the regional council dataset grouped by land use for c) all data d) with Auckland undisturbed data removed. Whiskers represent maximum or minimum observations within 1.5 × inter-quartile range; outlying data shown as open circles.

4.2 Preliminary analysis – Southland-Otago data

While some significant effects of land use were observed in the GNS Science Southland-Otago dataset, these differences were typically very small (Figure 9). Further, the robustness of the background soil model was considered to be low, given the low number of samples in this class (27, Table 3) and as illustrated by the comparatively wide confidence limits in Figure 10. Hence data were aggregated for subsequent analysis (Table 6).

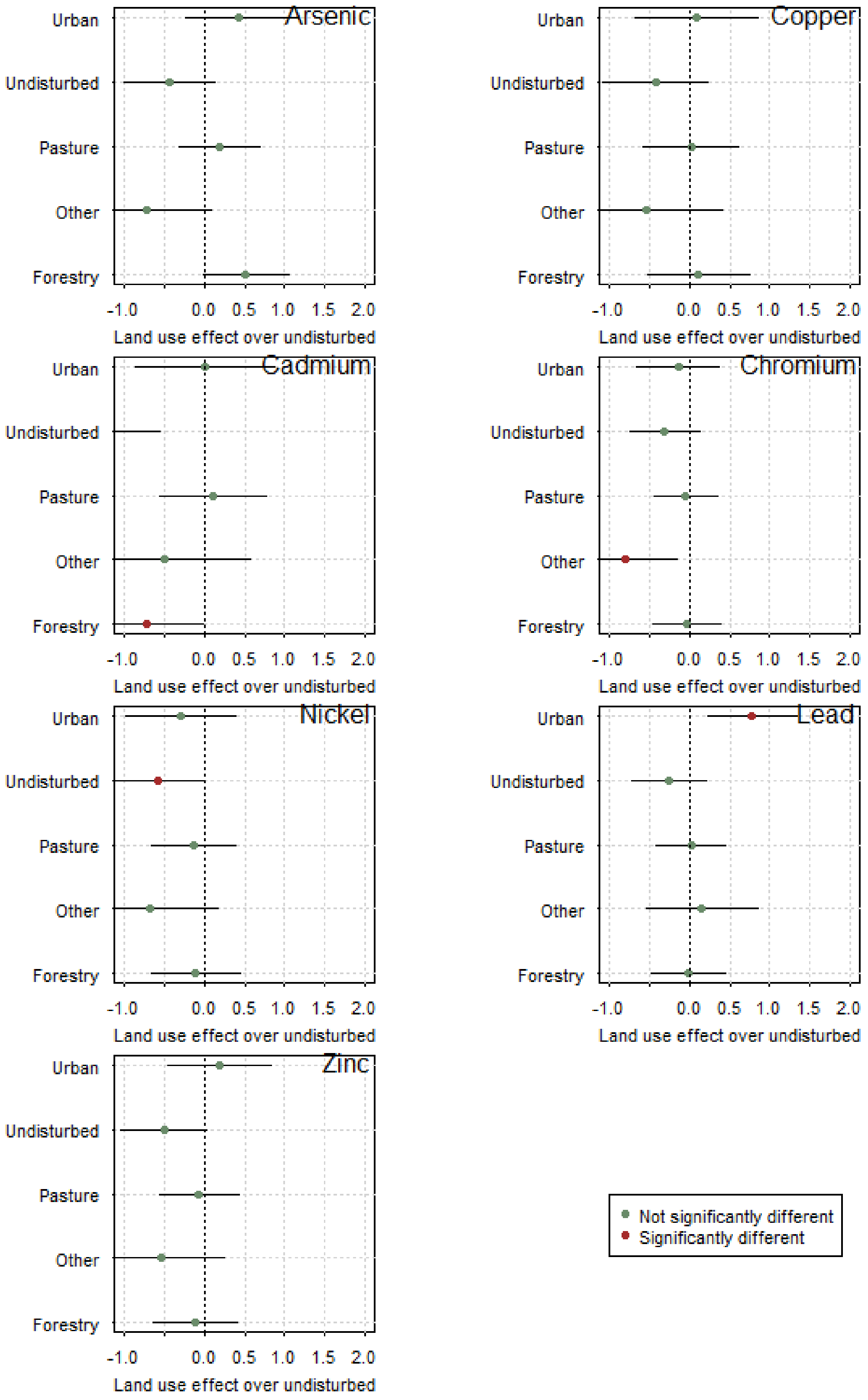

Figure 9 Summary of trace element concentrations from the GNS Science Southland-Otago dataset with respect to different land uses. 0.0 represents no difference from undisturbed land use, lines represent the 95th confidence interval of the mean.

Using the aggregated land-use class dataset, a range of models was identified as being the best explanation for the different trace element concentrations, with general consistency shown between the primary parameter identified in the conditional mean and 95th percentile models (Table 13). Details of the comparison are provided in Appendix 2.

Table 12 Summary of the favoured models for the GNS Science Southland-Otago dataset based on the conditional mean and 95th percentile, selected using AIC criterion

Element	Mean	95th percentile
Arsenic	Chemical5	Chemical5
Cadmium	Chemical2	Chemical5
Chromium	Chemical5	Chemical1
Copper	Chemical5	Chemical5
Lead	Chemical4	Rock Class
Nickel	Chemical5	Chemical1
Zinc	Chemical2	Rock Class

4.3 Comparison between models – preliminary analyses

Using the favoured models for the predicted 95th percentile concentrations, the predicted background median concentrations for each element from both the regional council and GNS Science Southland-Otago datasets, were compared (Appendix 3). The confidence interval for factor levels with many samples are quite narrow and the estimates are therefore considered reliable, whereas factor levels with few samples will have wide and possibly unreliable estimates for the median. Predictions were also compared directly by using the same model to provide predictions of the median concentrations for the regional council and GNS-Science Southland-Otago data (Appendix 2).

These results showed that there was no bias in the estimates developed from the two datasets, with the exception of Cd, which showed a right bias in the estimates (i.e. greater concentrations predicted from the regional council data). This is attributed to the greater amount of data from agricultural land that showed elevated concentrations in the regional council data (Figures 4–6).

4.4 Final data analysis

The final data analysis was undertaken using the combined regional council and GNS Science Southland-Otago datasets, and aggregated land-use data to maximise the amount of data for analysis (see section 3.3.2). Testing for spatial correlation revealed that, in contrast to the individual datasets, the data was spatially correlated for As and Cr (Figure 10).

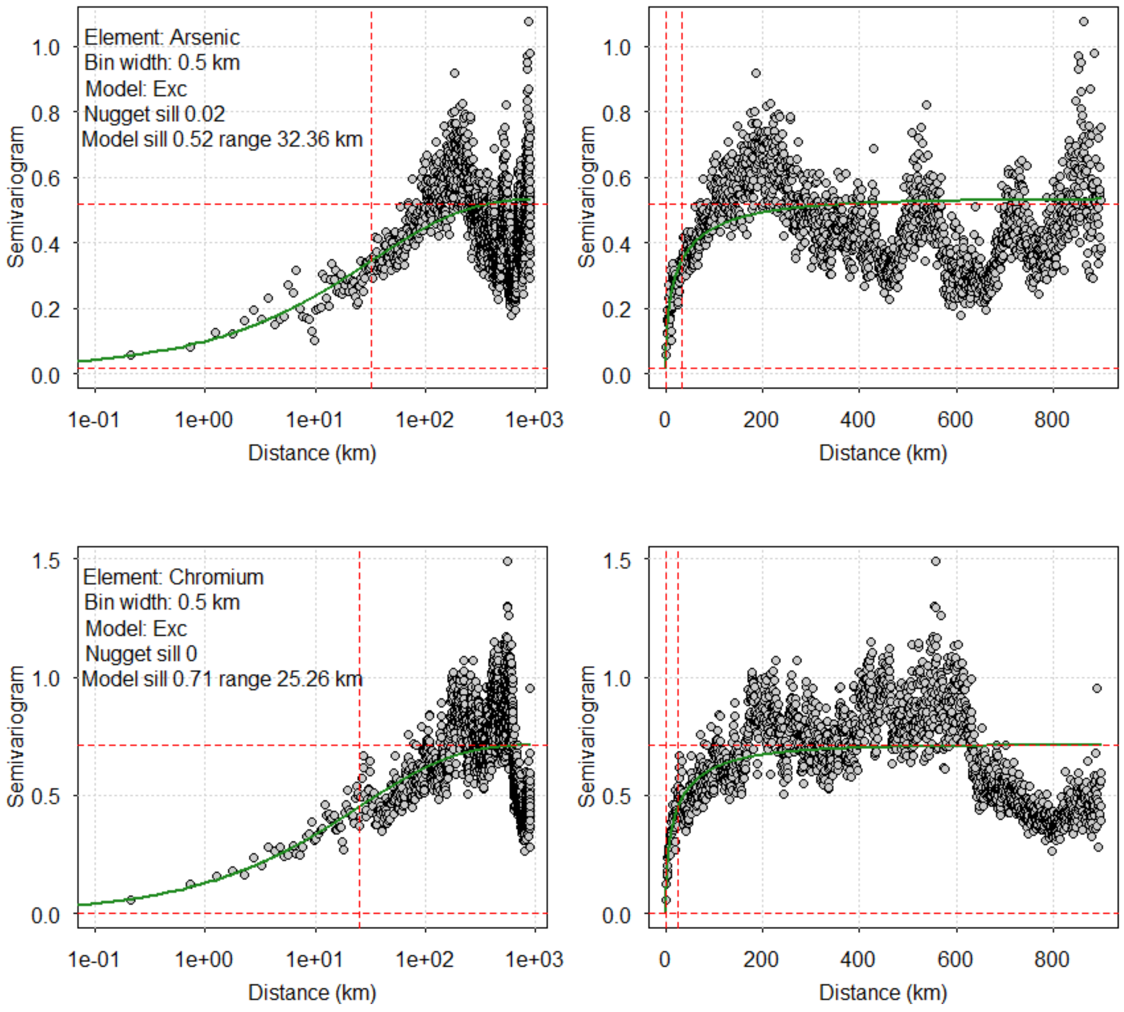

Figure 10 Semivariograms for arsenic and chromium showing spatial correlation over varying distances (range).

The same general models are used to develop predictions with and without a correction for spatial autocorrelation; however, the mathematics underlying the model predictions differs. In the first instance, the model selected for use was based on assuming no spatial autocorrelation, which revealed that *Chemical4* is favoured most often (Table 12). Thus estimates based on this variable were used to compare predictions from the 50% quantile estimates (*without* a correction for spatial autocorrelation) and the GLS mean estimates (*with* a correction for spatial autocorrelation). Example comparisons are shown for arsenic and chromium (Figure 11). Generally, the predictions based on the median and the GLS method were quite close, although there were some exceptions. For example, the prediction for a *Chemical4* value of 'metasediment' was much higher for the GLS model, suggesting that this *Chemical4* factor level was strongly clustered (and indeed all points were located close together in the south-west of the South Island), although the number of samples on which this model was based was low (n = 3). In general, GLS predictions also typically increased the difference between median and 95th quantile estimates, although some reductions did occur (e.g. for Rhyolite). Finally, estimates produced from fewer samples should be considered as indicative rather than definitive.

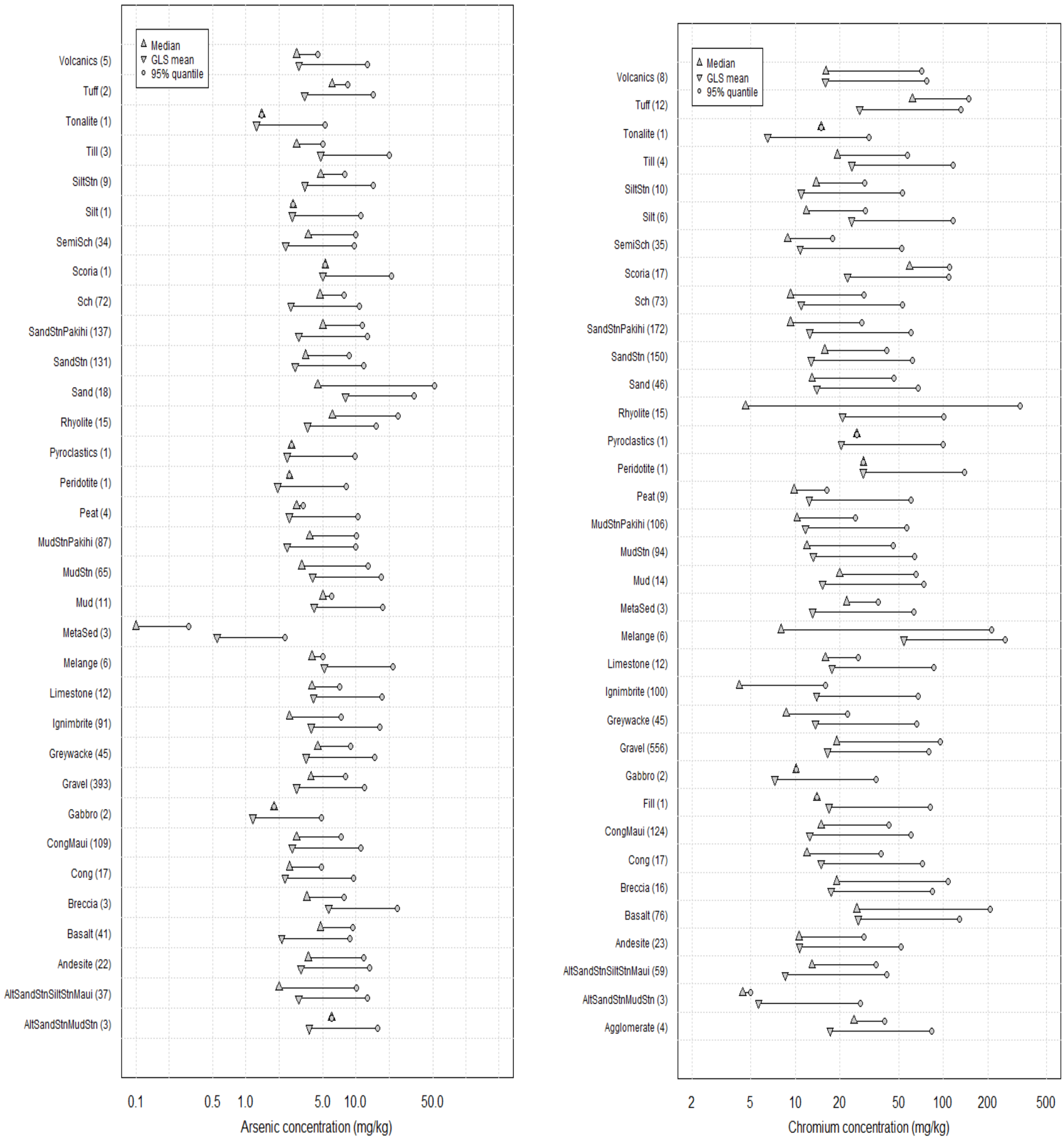

Figure 11 Comparison of predictions for 50% (triangle) and 95th percentile (circle) estimates (without a correction for spatial autocorrelation, upper line) and the GLS mean estimates and 95th quantile estimate (with a correction for spatial autocorrelation) for arsenic and chromium. The number of elements for each of the Chemical4 factor levels is indicated on the vertical axis of the plots.

Figures 12–15 show the estimated effective median and 5th and 95th percentile of predicted concentrations distribution for the individual trace elements in the different *Chemical4* subgroups. These predicted concentration ranges provide a first-order estimation; however, estimates from groupings with fewer samples should be considered as indicative only. There is no guidance on what constitutes a minimum number from a statistical analysis perspective. It is noted that a minimum of 30 samples is recommended to be collected to characterise background concentrations of a given pedo-geological area (ISO 2011), although this is based on the conventional statistical analysis of the data as opposed to developing predictive relationships. Comparison of the predicted median with the median and spread of data for samples underlying the different subgroups suggested that around 8–10 is a reasonable minimum number, although this may be dependent on the degree of spatial clustering of those data points. A good example is the melange subgroup, for which the median of the actual data points for the different elements is typically much lower than the predicted median, taking into account spatial autocorrelation. In this subgroup, three of the six data points are located close together (within 500 m). This illustrates the potential influence of spatial clustering on estimates produced without accounting for this and/or the potential for an exaggeration of the spatial correlation effect at low sample numbers. To provide a conservative approach, a minimum of 30 samples is used as the basis for determining a reliable estimate of the concentration distribution for a given sub-group. However, as discussed below, there may be additional factors that influence the concentrations within a given sub-group. The fit of the predicted concentration distribution compared to the underlying data is shown for As and Cr gravel sub-groups are shown as examples (Figure 16).

The predicted concentration ranges are similar for many of the subgroups for a given element, suggesting that these subgroups could be merged to provide a simpler delineation of background concentration ranges if required. The analysis also reveals subgroups for which higher or lower concentrations might be anticipated, for example As having higher concentrations in sand and mud and lower concentrations in basalt and semischist. In the case of sand, further analysis of the data underlying this subgroup revealed that higher predicted concentrations of As are driven by the presence of three data points in the Waikato region with As concentrations of 18, 21.9 and 51 mg/kg and one data point in Canterbury with an As concentration of 13 mg/kg (Figure 17). The data points from Waikato are located within the mineralised area, suggesting it may be relevant to delineate mineralised areas, and develop different background concentrations for areas that intersect with mineralisation zones. In contrast to sand, the data points underlying the mud subgroup show a relatively even spread of concentrations across different regions, suggesting the predicted concentration values are more likely to be representative of soils within the mud subgroup (although the number of samples in this group is low). Similarly, while samples underlying the basalt subgroup are dominated by samples from Waikato and Auckland, this subgroup also includes samples from a number of regions and thus predicted concentrations are considered more representative of concentrations across New Zealand.

Predicted concentrations for Cd indicate higher concentrations in limestone and rhyolite although sample numbers are low in both groups. Nonetheless, the samples underlying the limestone subgroup include samples from Waikato, Wellington, Canterbury and Auckland regions, and samples underlying the rhyolite subgroup are located in Waikato and Bay of Plenty, suggesting that the predicted concentrations, or at least the presence of higher concentrations, are representative of soils within these subgroups. In contrast to these higher concentrations, lower concentrations of Cd appear to occur in the schist subgroup, for which all samples are located in Otago. Similarly, the schist and semischist subgroups typically have lower Cu concentrations, with most samples underlying these subgroups located in Otago, and some samples located in Marlborough. The conglomerate subgroup also has lower Cu concentrations, with most samples underlying this subgroup located in Auckland.

For Cr, Ni, Zn, and to a lesser extent, Cu, the volcanic subgroups basalt, tuff, and scoria show the highest predicted concentrations, although the number of samples in some of the tuff and scoria subgroups is low. This is consistent with findings from the Auckland Regional Council (ARC 2001) that Cu Cr, Ni and Zn concentrations in volcanic soils were higher than those in non-volcanic soils. However, it should be noted that all samples from the tuff and scoria subgroups, and the basalt subgroup for Zn were located in Auckland, and thus may be only representative of those rock groups in the Auckland region. In contrast, basalt samples for Cu, Cr and Ni were more widely spread and included samples from Auckland, Waikato, Northland, Canterbury and Marlborough, and thus may be considered more representative of concentrations in that subgroup across New Zealand.

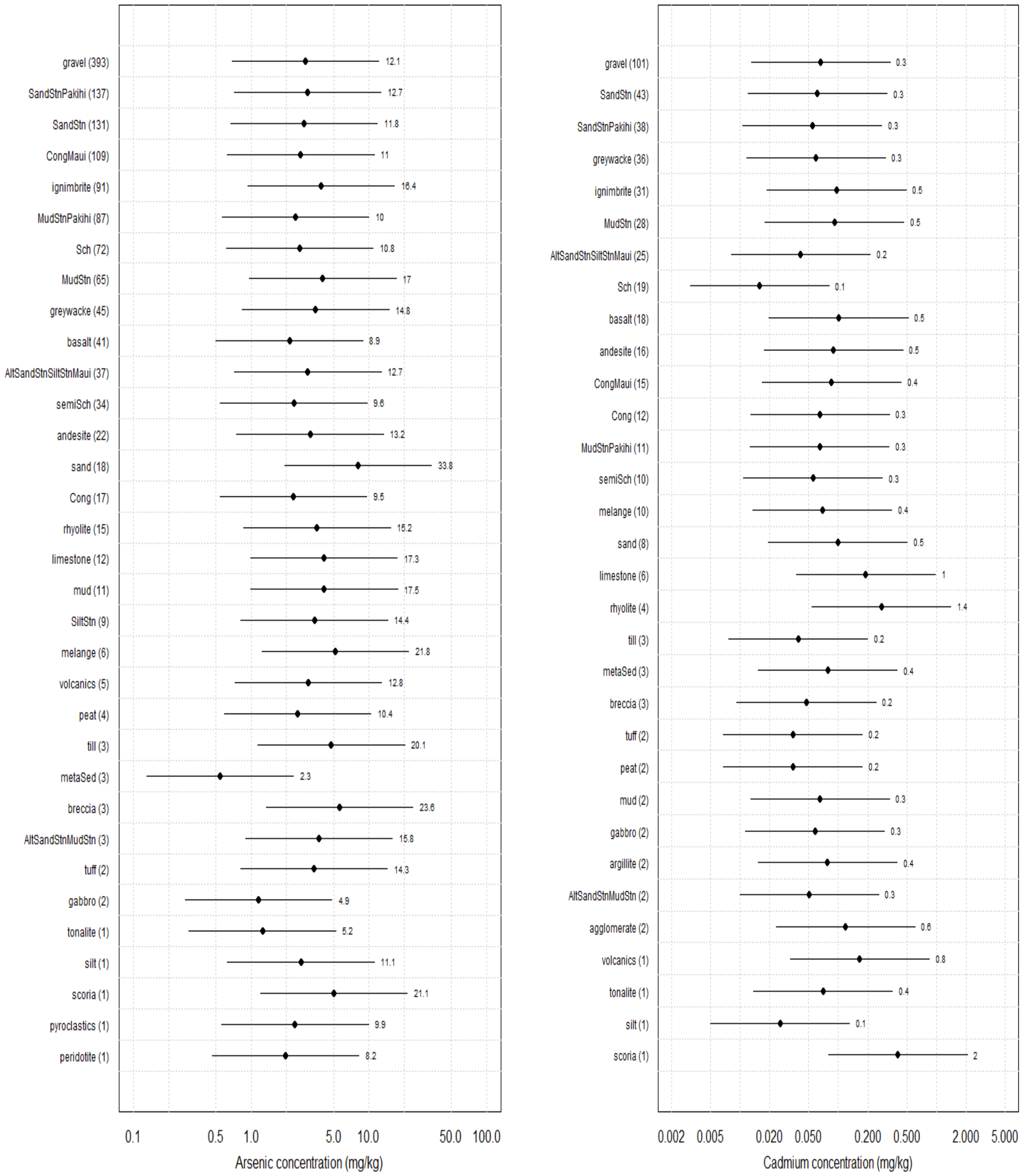

Figure 12 Effective median, 5th and 95th percentile estimates of arsenic and cadmium concentrations for individual Chemical4 subgroups. Subgroups with less than 30 samples (number of samples given in brackets) are considered less reliable and below 10 samples unreliable.

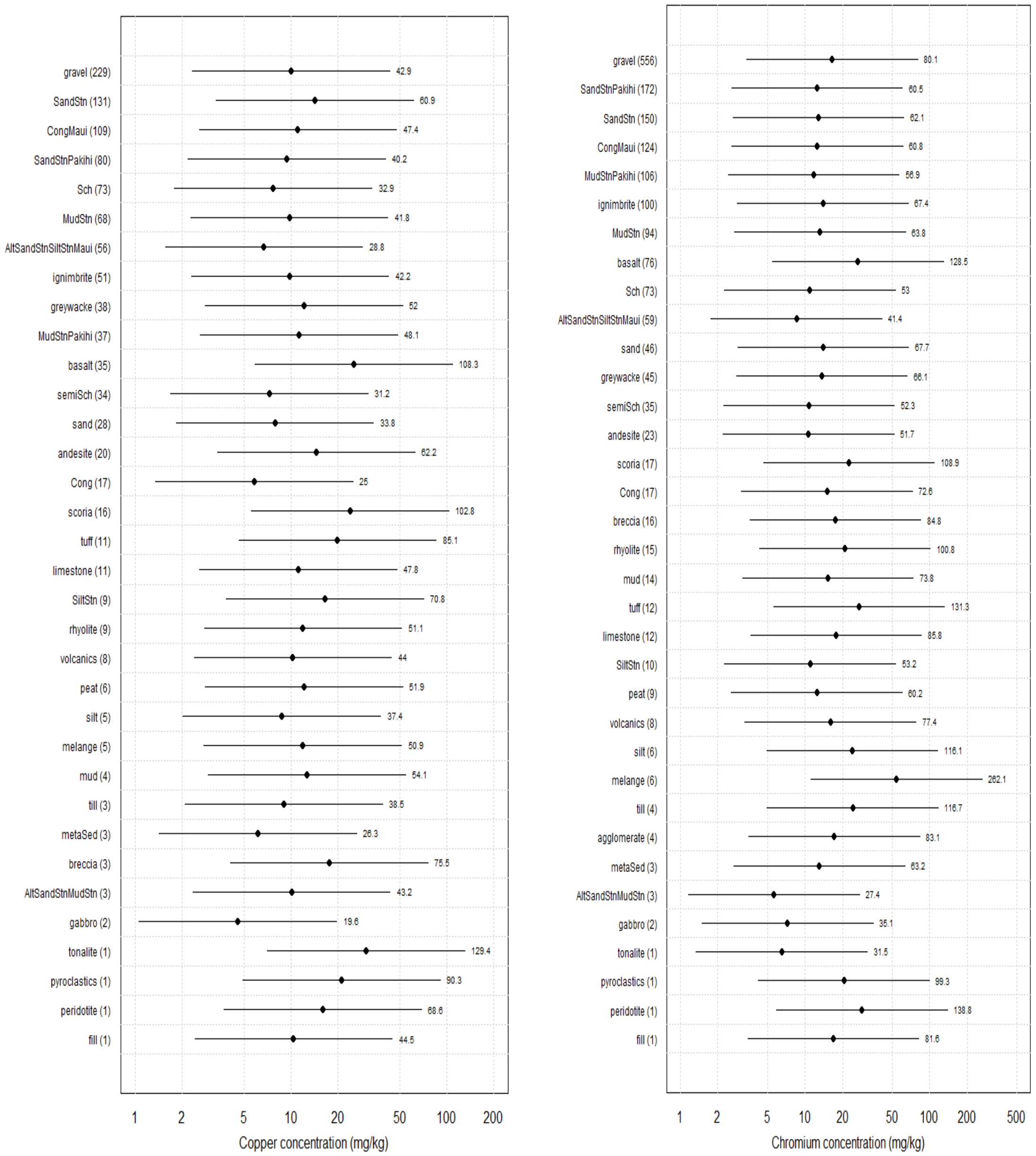

Figure 13 Effective median, 5th and 95th percentile estimates of copper and chromium concentrations for individual Chemical4 subgroups. Subgroups with less than 30 samples (number of samples given in brackets) are considered less reliable and below 10 samples, unreliable.

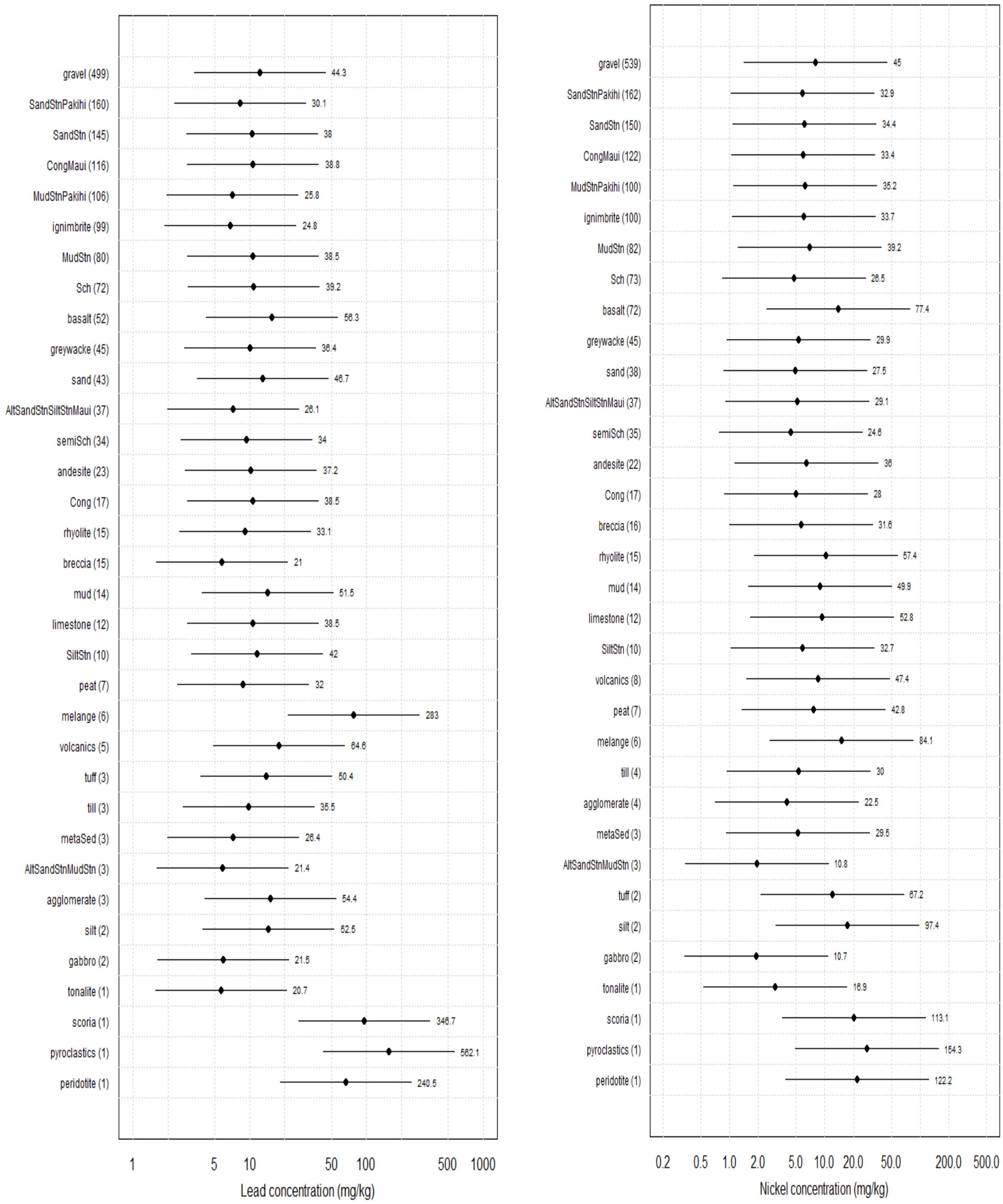

Figure 14 Effective median, 5th and 95th percentile estimates of lead and nickel concentrations for individual Chemical4 subgroups. Subgroups with less than 30 samples (number of samples given in brackets) are considered less reliable and below 10 samples, unreliable.

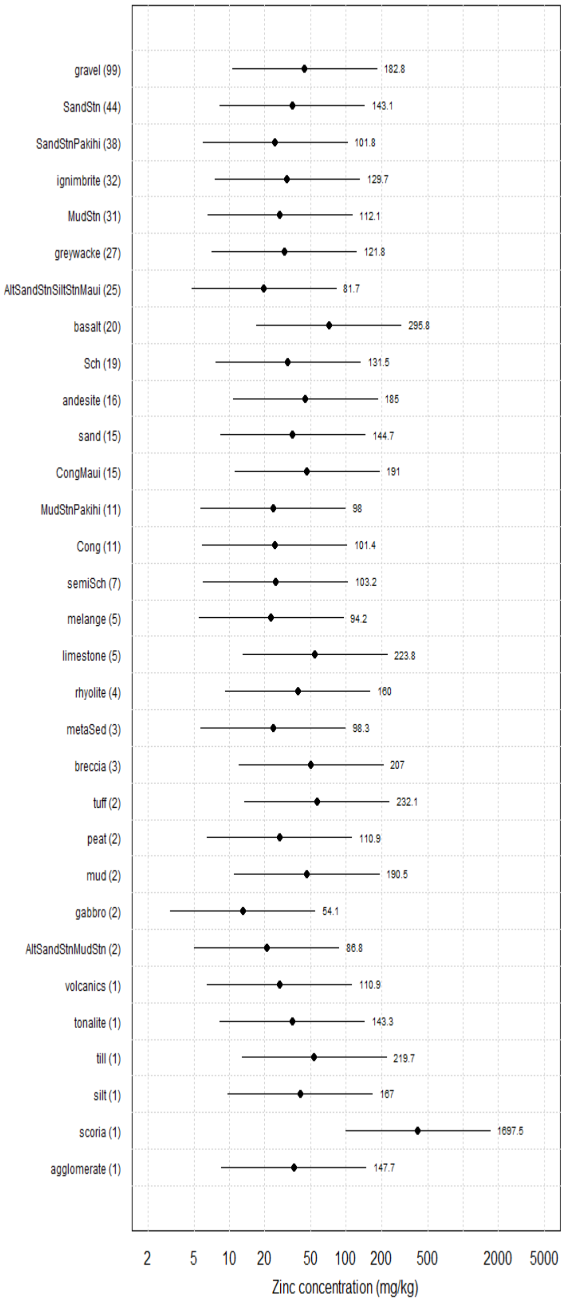

Figure 15 Effective median, 5th and 95th percentile estimates of zinc concentrations for individual Chemical4 subgroups. Subgroups with less than 30 samples (number of samples given in brackets) are considered less reliable, and below 10 samples unreliable.

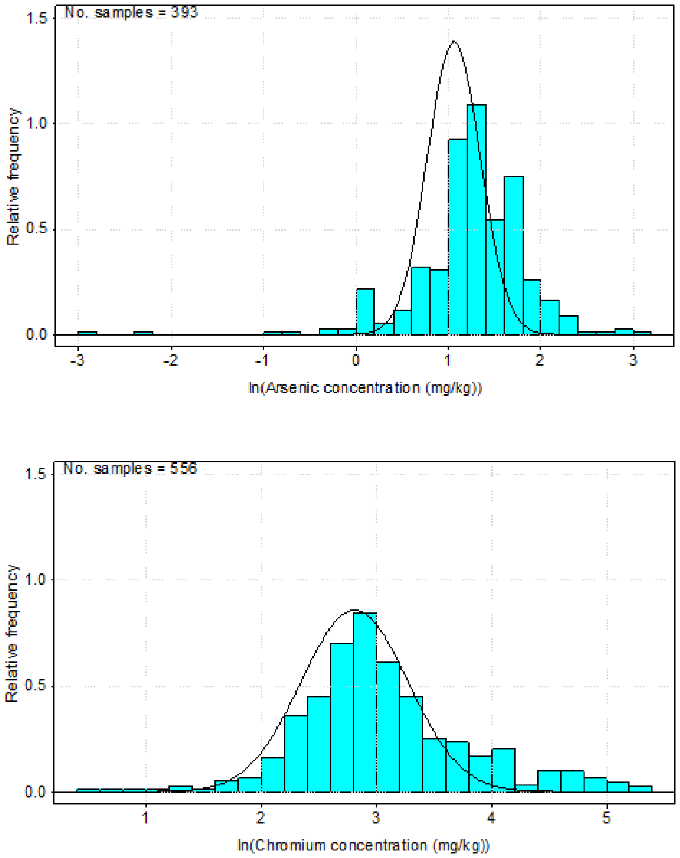

Figure 16 Comparison of predicted concentration distribution (line) of As and Cr gravel sub-groups with underlying data (histogram). The predicted concentration distribution for As appears to be influenced by a small number of low concentration estimates. The fit for Cr is generally good, although refer to the text for discussion on the elevated concentrations in the underlying dataset.

4.5 Comparison with exploration data

Comparison of the distribution of samples with elevated trace element concentrations in rocks or soils captured either in the Crown Minerals database or the PETLAB database, with the distribution of the regional council soil sampling locations indicates that some soil sampling has been undertaken in regions where concentrations might be naturally elevated (Figures 17-22). Further analysis of these data points may provide more insight into the extent of any elevated concentrations in the soil, but has not been undertaken here. There is limited minerals exploration or research data for Cr and Ni; these mineralisation maps also identify the Dun Mountain terrane, and the serpentinite rock group that are known to be elevated in Cr and/or Ni.

There is a reasonably large area in the Wellington region that is identified as having elevated concentrations of As and Cu in soil (Figures 17 & 18). However, this is a function of the sample points with elevated As and Cu being located in a relatively large polygon. Further, for this polygon there were c. 603 samples that had been analysed for As and Cu in soil, with all of them located at the southern end of the polygon (Red Rocks outcrop on the Wellington coast; Roser & Grapes 1990). Only 12 of those samples had elevated As (>20 mg/kg) and 12 had elevated Cu (>60 mg/kg). This suggests that the actual area of naturally elevated concentrations is likely to be much smaller than indicated.

Erosion of mineralised areas may result in elevated concentrations in surficial depositional groups (gravels) downstream of these areas. Higher gravel Cr concentrations (>60 mg/kg) were evident in gravels from Marlborough and Tasman (and one sample from Southland), although the location of data points underlying the gravel subgroup were widely distributed across New Zealand (Figure 23). The elevated concentration in Southland has been attributed to the eroding Dun Mountain Ultramafic Group rocks at West Dome (Martin et al. 2015; Turnbull et al. 2015) and erosion of the Dun Mountain terrane in Marlborough may have influenced gravel Cr concentrations in Marlborough and Tasman. Elevated concentrations in Tasman gravel area had previously been noted (Cavanagh 2015b). Despite the relatively low predicted Cu concentrations overall, there is a reasonable spread of concentrations in the samples underlying this group with three data points of >80 mg/kg all located in Southland. Similarly, the higher concentration data points in the sandstone subgroup were located in Southland. Figure 18 shows that areas of Cu mineralisation do occur in Southland. Further investigation is required to better establish the relationship of soil samples from within sedimentary rock groups and erosion zones of mineralised areas. However, it may be challenging to attribute variation in concentrations to geological causes using QMAP as the QMAP seamless GIS geological unit polygon layers are too broadly based, geomorphically controlled and regional in extent to differentiate lithological differences in alluvium. Alluvium by its nature varies from boulders to coarse gravel to sand to silt on a metre scale with depth, but these beds can have horizontal extents of kilometres in some cases. Furthermore, the clast components from multiple rock type sources are typically thoroughly intermixed. More detailed geological maps are unlikely to characterise these differences. In rare cases, catchments have very distinctive rock units in their upper reaches and the chemical signature of those units is recorded in alluvium downstream, as noted above for elevated Cr in Oreti River alluvium. Elevated As may also be observed in sedimentary samples due to erosion of Otago schist, which can have high As concentrations (Craw et al. 2003).

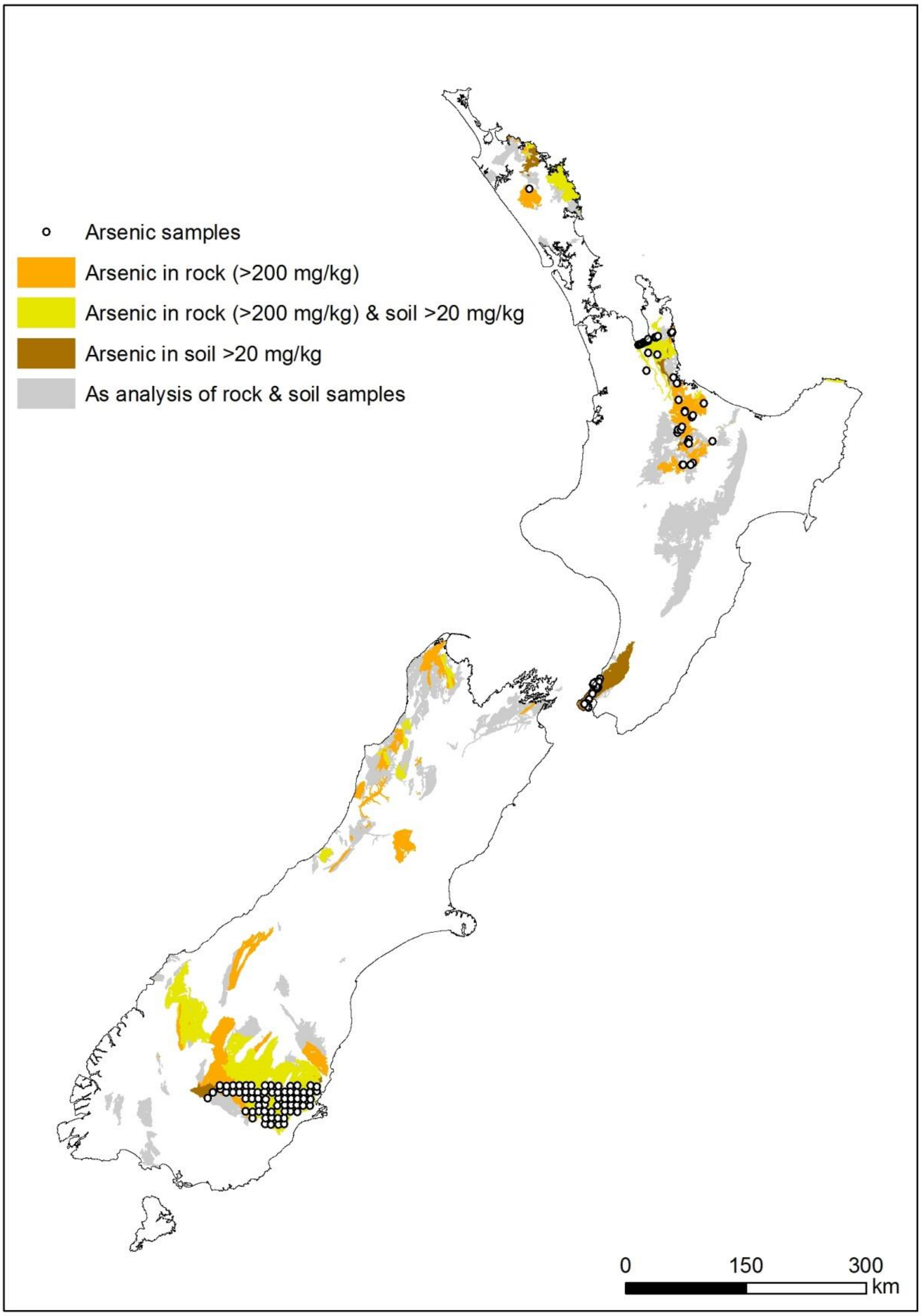

Figure 17 Identification of arsenic-mineralised areas from rock and soil samples from the Crown Minerals database, and rock samples from GNS Science's PETLAB database applied to the surrounding QMAP map unit polygon. This potentially highlights areas of elevated background soil concentration. Dots show the location of samples in the current dataset that have been collected from within the identified mineralised areas.

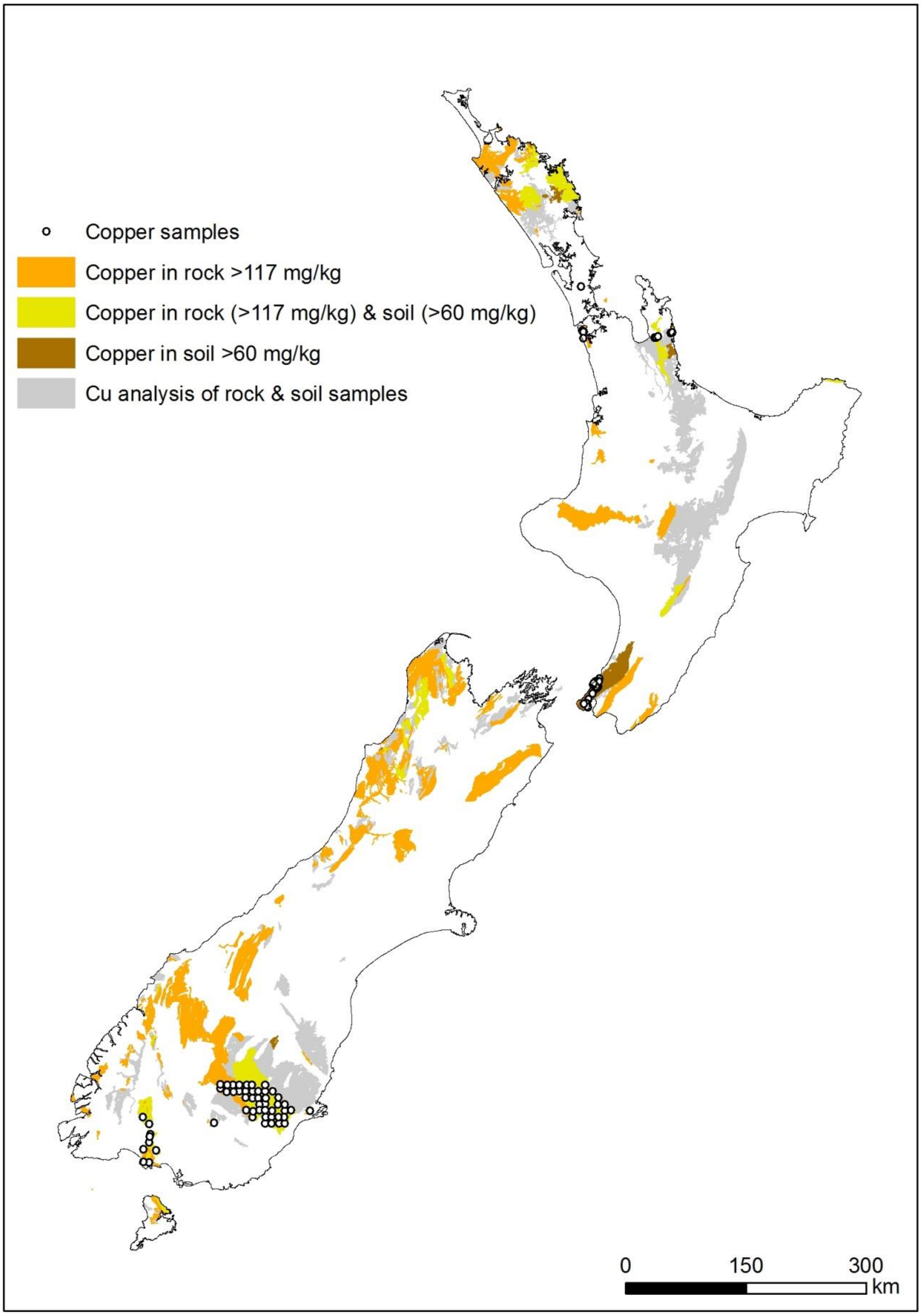

Figure 18 Identification of copper-mineralised areas from rock and soil samples from the Crown Minerals database, and rock samples from GNS Science's PETLAB database applied to the surrounding QMAP map unit polygon. This potentially highlights areas of elevated background soil concentration. Dots show the location of samples in the current dataset that have been collected from within the identified mineralised areas.

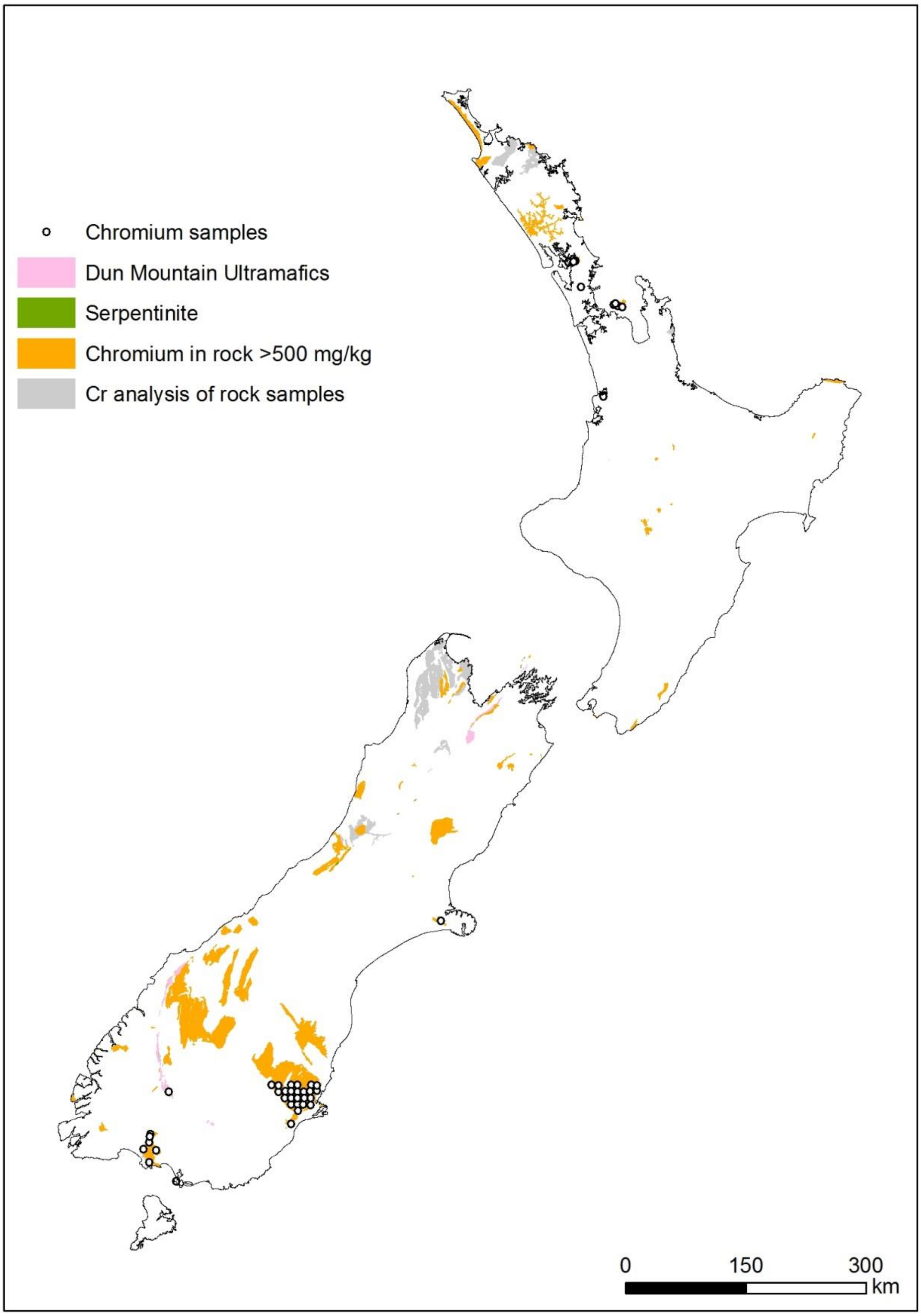

Figure 19 Identification of chromium-mineralised areas (note: area of serpentinite is very small) from rock and soil samples from the Crown Minerals database, and rock samples from GNS Science's PETLAB database applied to the surrounding QMAP map unit polygon. This potentially highlights areas of elevated background soil concentration. Dots show the location of samples in the current dataset that have been collected from within the identified mineralised areas.

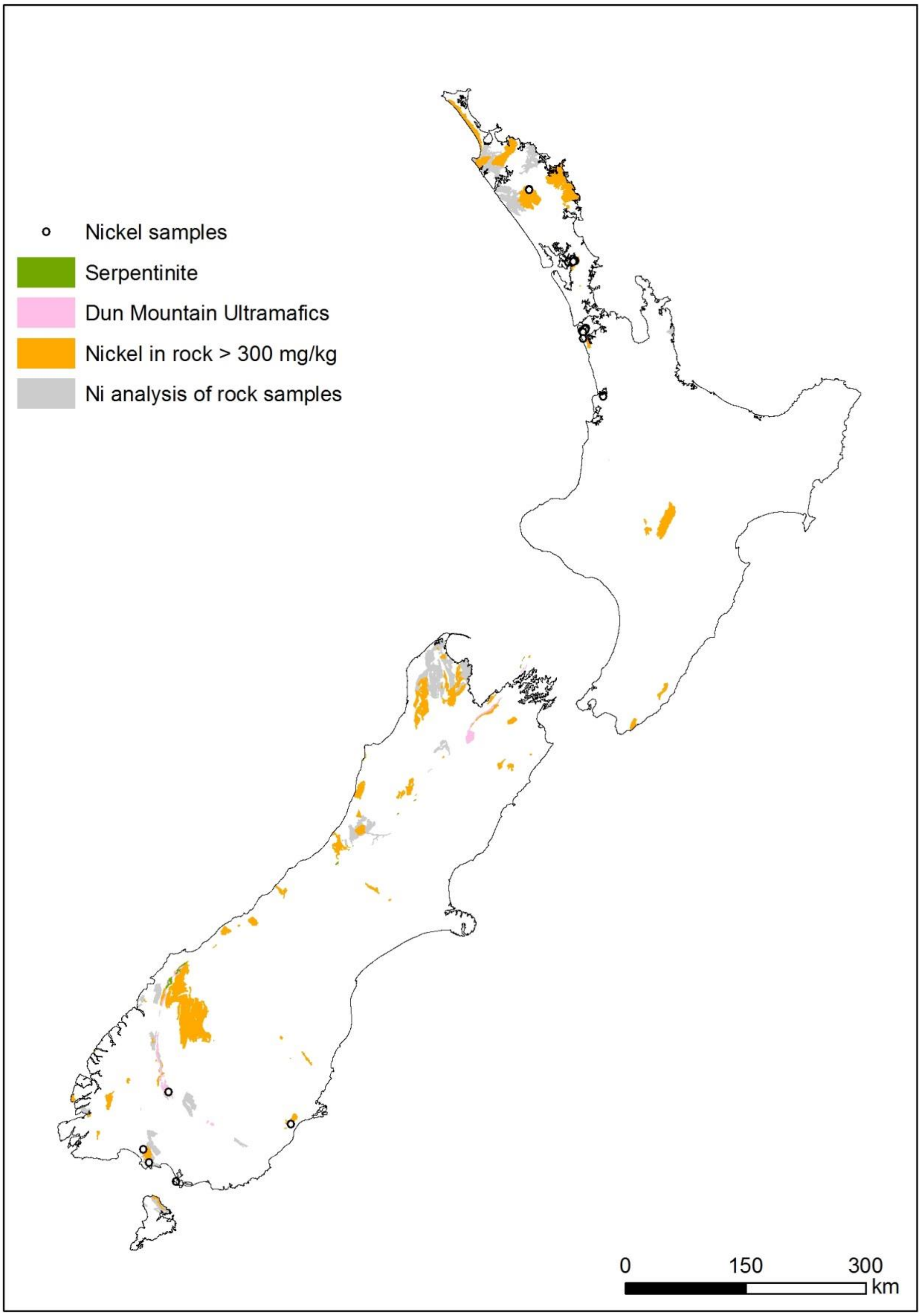

Figure 20 Identification of nickel-mineralised areas (note: area of serpentinite is very small) from rock and soil samples from the Crown Minerals database, and rock samples from GNS Science's PETLAB database applied to the surrounding QMAP map unit polygon. This potentially highlights areas of elevated background soil concentration. Dots show the location of samples in the current dataset that have been collected from within the identified mineralised areas.

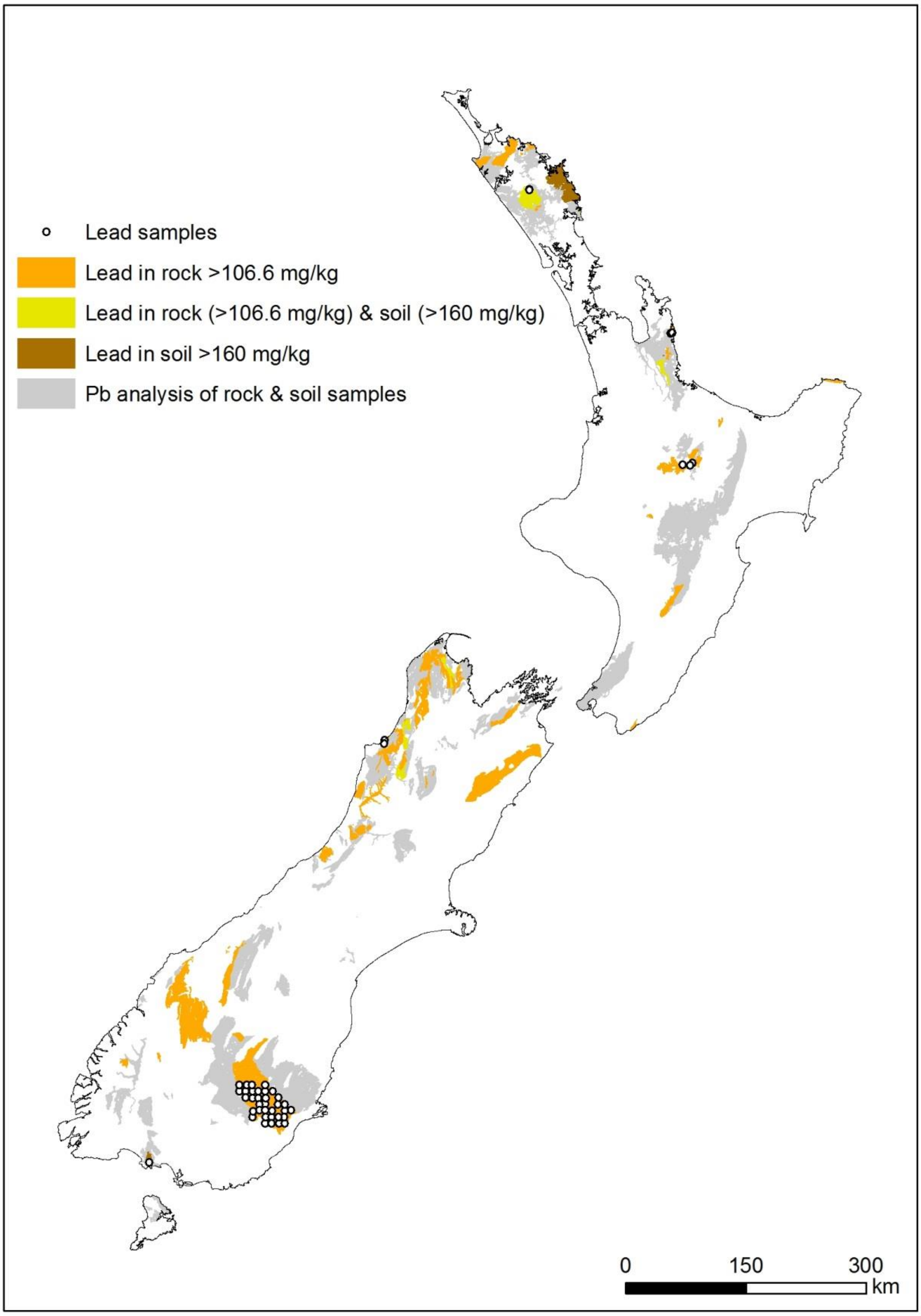

Figure 21 Identification of lead-mineralised areas from rock and soil samples from the Crown Minerals database, and rock samples from GNS Science's PETLAB database applied to the surrounding QMAP map unit polygon. This potentially highlights areas of elevated background soil concentration. Dots show the location of samples in the current dataset that have been collected from within the identified mineralised areas.

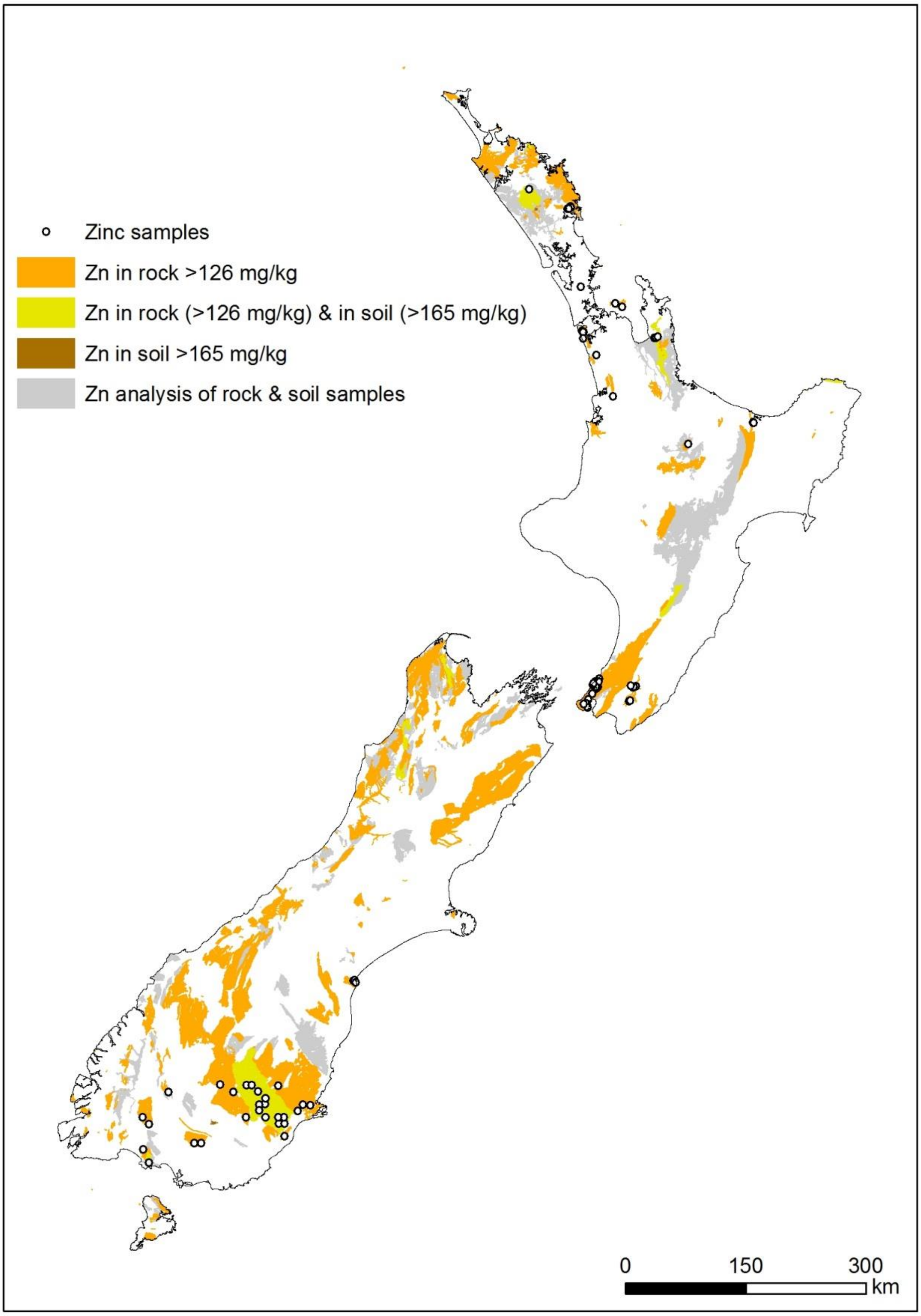

Figure 22 Identification of zinc-mineralised areas from rock and soil samples from the Crown Minerals database, and rock samples from GNS Science's PETLAB database applied to the surrounding QMAP map unit polygon. This potentially highlights areas of elevated background soil concentration. Dots show the location of samples in the current dataset that have been collected from within the identified mineralised areas.

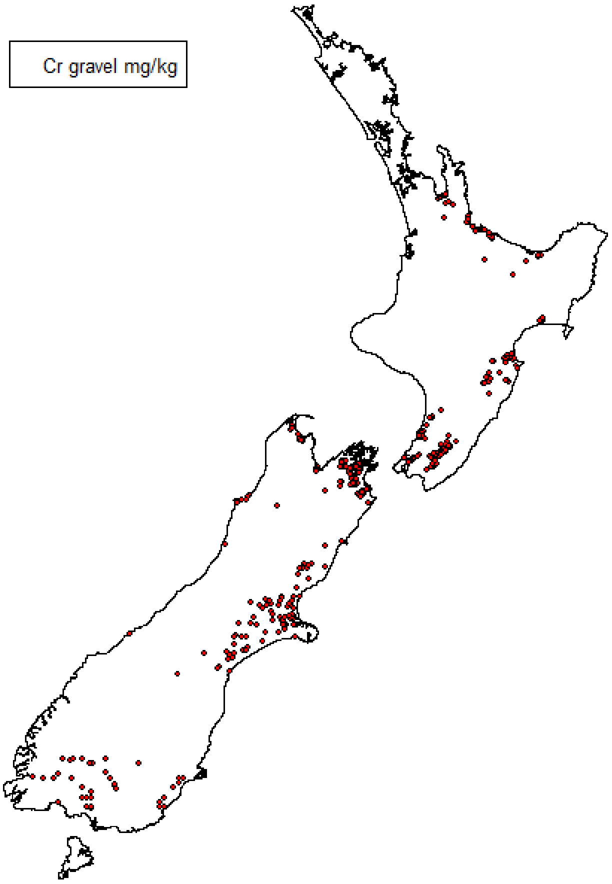

Figure 23 Sample locations within the gravel subgroup for chromium (Cr).

4.6 Discussion

4.6.1 Preliminary analyses

The use of different spatial databases in the data analysis provided an assessment of the utility of those databases for determining background concentrations. While the *rock-type-of-fines* parameter from S-Map shows promise as a potential explanatory variable for determining background concentrations, the limited coverage of S-Map prevented the further use of this database. Similarly, despite national coverage of the Fundamental Soils Layer, the classification of ‘town’ in *Top rock* (used to indicate urban areas) reduced the number of samples that could be used in subsequent analysis, and thus prevented further use of this parameter and database. Most data were retained using the different QMAP parameters, and thus QMAP was the primary database used.

Underlying geology is generally regarded as the major contributor to the geochemical signals in soils and surficial material. The rocks are composed of minerals and these have defined chemical compositions, most commonly as silicates, and many minerals naturally contain the trace elements of interest in this study. The geochemical baseline survey of southern New Zealand established a first order correlation between many element concentrations and the composition of the underlying geology (Martin et al. 2015; Turnbull et al. 2015). Geological maps depict areas of similar rock types emplaced over a discrete period of time and these are referred to as geological units. Each geological unit has a degree of lithological homogeneity placed on it, in some cases embodied in a formal stratigraphic name. Logically, each unit also should have a signature geochemical composition, based on the proportion and chemical composition of its constituent minerals, and commonly reflected in their overlying soils.

Analysis of the regional council data found a number of different parameters appeared to provide the best fit for the regression models, although a rock group-based classification (*Chemical4)* was the best fit for many trace elements. In contrast, the more geochemistry derived groupings (*Chemical1–3, 5*), and in particular *Chemical5*, provided the best fit for the models developed from the GNS Science Southland-Otago region dataset.

The regional council dataset used in this study is strongly oversampled in areas of young sedimentary deposits and rocks, typically in low-lying populated or intensively farmed areas. For example, 63% of samples with As concentration measurements occur in areas of Pakihi Supergroup, the Pliocene-Quaternary alluvium-dominated high level stratigraphic unit. In terms of rock type, 81% of regional council samples are categorised as ‘clastic sediment’ according to their coincident QMAP geological unit. The oversampling likely contributes to a relatively weak correlation between measured trace element values and their geochemical lithological substrate. By way of contrast, the grid-based southern South Island geochemical baseline survey data show a stronger correlation to their underlying geology, as is evidenced by geochemistry derived groupings providing the best fit for the models developed. This is attributed to the geological (and therefore chemical) diversity of the Southland area and the less biased sampling.

Land-use effects were evident for most trace elements in the regional council dataset, with the effects observed differing for the individual trace elements. The identification of a significant land-use effect on Cd for all agricultural land uses provided confirmation that the method of analysis was sensitive enough to detect land-use effects where they occurred. The significant differences noted for Cr and Ni are not considered attributable to land use, but rather the location of the land use. Similarly, the significantly lower concentrations of most trace elements for forestry is attributed to location as opposed to the land use. There was some evidence for land-use effects in the GNS Science Southland-Otago dataset, although these differences were small and the low number of samples in the background land-use class reduced the robustness of estimates produced for this class. Thus, land-use classes were aggregated for subsequent analysis of the GNS Science Southland-Otago dataset.

Comparison of the predicted background concentrations using the models developed for the individual datasets indicated that similar results were obtained, and there was no bias in the estimates developed from each dataset. Thus, to provide a fuller dataset for the final analyses, the two individual datasets were combined, along with the land uses that were not significantly different from background.

4.6.2 Final analyses

The aggregation of data from selected land uses and of the two datasets resulted in spatial correlation effects being evident in the data. This necessitated a slight change in the analysis to account for this effect. However, the general modelling approach remained the same, and the rock group-based parameter *Chemical4* was found to provide the best fit – likely driven by the amount of data from the regional council dataset. The effective median and 95th quantile estimates of the background concentration for the different *Chemical4* classes are shown in Tables 14 and 15.

As noted in section 4.4, estimates for *Chemical4* sub-groups with n <30 are considered less reliable. Estimates for sub-groups with n >30, appear to provide reasonable predictions of background concentration when considering the underlying data, and provide insight into selected factors influencing higher or lower concentrations of a given trace element. Further, it is noted that the predicted concentration range for a number of different subgroups is similar, suggesting these groups could be combined to provide a simpler presentation of background concentrations across New Zealand. However, it is also noted that for some subgroups, the underlying data is localised to a particular area. In some cases, this is due to the distribution of that subgroup; in other cases it is simply a factor of where samples have been collected, and raises questions about the representativeness of the predicted concentrations.

Using this approach, predicted background concentrations are available for a significant area of New Zealand. Table 16 lists the area covered by each *Chemical4* subgroup, and the subgroups for which background concentration estimates were produced. Using a criterion of accepting only concentrations for subgroups with sample numbers greater than 30, a significant area of New Zealand is covered (Figure 32), with the largest gaps existing in Fiordland and the north-west coast of the South Island. Granite, followed by diorite, are the groups with the greatest area for which no data is available, and should be a target for further data collection. Similarly, areas for which predicted background concentrations were developed but no or limited data (n < 30) are available, are targets for further sampling and analysis to confirm concentrations fall within the predicted concentrations or to refine predicted concentrations.

Further analysis of the surficial depositional subgroups (e.g. gravel, sand, mud) is warranted to determine the extent to which erosion of mineralised rocks is contributing to elevated concentrations in these soils. Mineralised areas have been delineated using data from the Crown Minerals database, the GNS Science PETLAB database, and for Cr and Ni, knowledge of specific lithologies that are elevated in those trace elements. These delineated areas are not considered to be a definitive map of mineralised zones in New Zealand, but rather provides best estimates for locations that may have greater concentrations than predicted for the given rock subgroup in that location and to identify the extent to which erosion of those mineralised areas is contributing to elevated concentrations in soils within downstream surficial depositional subgroups. The latter is unlikely to be achieved simply using QMAP data as the QMAP seamless GIS geological unit polygon layers are too broadly based, geomorphically controlled and regional in extent to differentiate lithological differences in alluvium. Rather, an assessment of the topography and erosional processes for the region in question needs to be made along with additional sampling to confirm the likely extent of elevated concentrations.

Additional sampling should also be undertaken in locations that fall within the mineralised zones to more adequately confirm the likely background concentrations in these regions. A minimum of 30 samples is recommended to characterise background concentrations of a given pedo-geological area (ISO 2011), although this depends on the area under consideration. It may be appropriate to analyse 30 samples for a regionally based characterisation, and a smaller number for a specific site investigation. In this instance, 7 to 10 samples may be appropriate, as used by some US EPA jurisdictions (Diamond et al. 2009). In this situation the intent is to identify whether the site in question contains naturally elevated concentrations of a trace element, rather than determining the background concentration per se.

Table 13 Predicted background concentrations (median and 95th quantile estimates) for arsenic, cadmium and copper in each of the Chemical4 factor levels. n = number of samples. Estimated concentrations for sub-groups with n <30 are considered less reliable and for n <10, unreliable.

Arsenic				Cadmium				Copper			
Chemical4 Factor	n	median	95%	Chemical4 Factor	n	median	95%	Chemical4 Factor	n	median	95%
gravel	393	2.88	12.06	gravel	101	0.066	0.34	gravel	229	10.00	42.85
SandStnPakihi	137	3.03	12.67	SandStn	43	0.061	0.31	SandStn	131	14.19	60.85
SandStn	131	2.81	11.77	SandStnPakihi	38	0.054	0.28	CongMaui	109	11.05	47.36
CongMaui	109	2.64	11.04	greywacke	36	0.059	0.30	SandStnPakihi	80	9.37	40.17
ignimbrite	91	3.91	16.38	ignimbrite	31	0.096	0.49	Sch	73	7.69	32.95
MudStnPakihi	87	2.38	9.97	MudStn	28	0.091	0.46	MudStn	68	9.76	41.83
Sch	72	2.58	10.80	AltSandStnSiltStnMaui	25	0.041	0.21	AltSandStnSiltStnMaui	56	6.71	28.77
MudStn	65	4.05	16.95	Sch	19	0.016	0.08	ignimbrite	51	9.83	42.16
greywacke	45	3.53	14.76	basalt	18	0.101	0.51	greywacke	38	12.14	52.03
basalt	41	2.12	8.87	andesite	16	0.089	0.45	MudStnPakihi	37	11.23	48.14
AltSandStnSiltStnMaui	37	3.03	12.67	CongMaui	15	0.085	0.43	basalt	35	25.27	108.3
semiSch	34	2.30	9.63	Cong	12	0.065	0.33	semiSch	34	7.28	31.19
andesite	22	3.16	13.24	MudStnPakihi	11	0.065	0.33	sand	28	7.88	33.78
sand	18	8.07	33.77	melange	10	0.069	0.35	andesite	20	14.50	62.17
Cong	17	2.28	9.54	semiSch	10	0.055	0.28	Cong	17	5.82	24.95
rhyolite	15	3.63	15.19	sand	8	0.099	0.50	scoria	16	23.98	102.8
limestone	12	4.14	17.32	limestone	6	0.19	0.97	limestone	11	11.14	47.77
mud	11	4.17	17.47	rhyolite	4	0.27	1.40	tuff	11	19.84	85.05
SiltStn	9	3.45	14.42	breccia	3	0.047	0.24	rhyolite	9	11.92	51.12
melange	6	5.20	21.75	metaSed	3	0.078	0.40	SiltStn	9	16.52	70.82

Arsenic				Cadmium				Copper			
Chemical4 Factor	n	median	95%	Chemical4 Factor	n	median	95%	Chemical4 Factor	n	median	95%
volcanics	5	3.05	12.75	till	3	0.039	0.20	volcanics	8	10.26	43.98
peat	4	2.49	10.42	agglomerate	2	0.12	0.60	peat	6	12.10	51.89
AltSandStnMudStn	3	3.78	15.82	AltSandStnMudStn	2	0.051	0.26	melange	5	11.88	50.94
breccia	3	5.65	23.64	argillite	2	0.078	0.40	silt	5	8.73	37.42
metaSed	3	0.55	2.28	gabbro	2	0.058	0.30	mud	4	12.63	54.13
till	3	4.79	20.06	mud	2	0.065	0.33	AltSandStnMudStn	3	10.07	43.18
gabbro	2	1.16	4.86	peat	2	0.034	0.18	breccia	3	17.61	75.52
tuff	2	3.42	14.32	tuff	2	0.034	0.18	metaSed	3	6.13	26.29
peridotite	1	1.95	8.18	scoria	1	0.402	2.05	till	3	8.98	38.49
pyroclastics	1	2.36	9.89	silt	1	0.026	0.13	gabbro	2	4.56	19.57
scoria	1	5.03	21.08	tonalite	1	0.07	0.36	fill	1	10.37	44.45
silt	1	2.65	11.08	volcanics	1	0.17	0.84	peridotite	1	15.99	68.55
tonalite	1	1.25	5.23					pyroclastics	1	21.06	90.29
								tonalite	1	30.19	129.4

Table 14 Predicted background concentrations (median and 95th quantile estimates) for chromium and lead in each of the Chemical4 factor levels for which data is available. n = number of samples. Estimated concentrations for sub-groups with n <30 are considered less reliable and for n <10, unreliable

	Chromium				Lead		
Chemical4 Factor	n	median	95%	Chemical4 Factor	n	median	95%
gravel	556	16.56	80.15	gravel	499	12.20	44.34
SandStnPakihi	172	12.50	60.50	SandStnPakihi	160	8.27	30.08
SandStn	150	12.83	62.07	SandStn	145	10.44	37.96
CongMaui	124	12.57	60.82	CongMaui	116	10.67	38.80
MudStnPakihi	106	11.76	56.88	MudStnPakihi	106	7.11	25.83
ignimbrite	100	13.92	67.35	ignimbrite	99	6.82	24.79
MudStn	94	13.19	63.83	MudStn	80	10.60	38.55
basalt	76	26.56	128.5	Sch	72	10.79	39.23
Sch	73	10.95	53.00	basalt	52	15.50	56.34
AltSandStn/SiltStnMaui	59	8.56	41.39	greywacke	45	10.02	36.43
sand	46	13.98	67.65	sand	43	12.85	46.71
greywacke	45	13.66	66.08	AltSandStn/SiltStnMaui	37	7.18	26.10
semiSch	35	10.80	52.26	semiSch	34	9.35	34.01
andesite	23	10.68	51.67	andesite	23	10.24	37.22
Cong	17	15.01	72.62	Cong	17	10.60	38.52
scoria	17	22.51	108.92	breccia	15	5.78	21.02
breccia	16	17.53	84.80	rhyolite	15	9.10	33.09
rhyolite	15	20.84	100.84	mud	14	14.15	51.45
mud	14	15.26	73.83	limestone	12	10.59	38.49
limestone	12	17.74	85.84	SiltStn	10	11.56	42.01
tuff	12	27.14	131.3	peat	7	8.79	31.97
SiltStn	10	11.00	53.21	melange	6	77.84	283.0
peat	9	12.45	60.24	volcanics	5	17.76	64.56
volcanics	8	16.00	77.40	agglomerate	3	14.96	54.39
melange	6	54.17	262.1	AltSandStnMudStn	3	5.88	21.37
silt	6	23.99	116.1	metaSed	3	7.26	26.40
agglomerate	4	17.18	83.15	till	3	9.76	35.47
till	4	24.11	116.7	tuff	3	13.86	50.39
AltSandStnMudStn	3	5.66	27.38	gabbro	2	5.92	21.51
metaSed	3	13.06	63.20	silt	2	14.45	52.54
gabbro	2	7.26	35.10	peridotite	1	66.16	240.5
fill	1	16.87	81.62	pyroclastics	1	154.62	562.1
peridotite	1	28.68	138.8	scoria	1	95.38	346.7
pyroclastics	1	20.51	99.26	tonalite	1	5.69	20.68
tonalite	1	6.51	31.50				

Table 15 Predicted background concentrations (median and 95th quantile estimates) for nickel and zinc in each of the Chemical4 factor levels for which data is available. n = number of samples. Estimated concentrations for sub-groups with n <30 are considered less reliable and for n <10, unreliable

	Nickel				Zinc		
Chemical4 Factor	n	median	95%	Chemical4 Factor	n	median	95%
gravel	539	7.98	44.96	gravel	99	44.06	182.8
SandStnPakihi	162	5.83	32.88	SandStn	44	34.50	143.1
SandStn	150	6.10	34.38	SandStnPakihi	38	24.53	101.8
CongMaui	122	5.93	33.42	ignimbrite	32	31.25	129.7
ignimbrite	100	5.99	33.75	MudStn	31	27.02	112.1
MudStnPakihi	100	6.24	35.15	greywacke	27	29.35	121.8
MudStn	82	6.96	39.21	AltSandStn/SiltStnMaui	25	19.68	81.66
Sch	73	4.71	26.52	basalt	20	71.29	295.8
basalt	72	13.74	77.43	Sch	19	31.70	131.5
greywacke	45	5.30	29.86	andesite	16	44.59	185.0
sand	38	4.88	27.49	CongMaui	15	46.03	191.0
AltSandStn/SiltStnMaui	37	5.16	29.07	sand	15	34.86	144.7
semiSch	35	4.36	24.58	Cong	11	24.43	101.4
andesite	22	6.38	35.98	MudStnPakihi	11	23.61	97.97
Cong	17	4.97	28.02	semiSch	7	24.86	103.2
breccia	16	5.61	31.60	limestone	5	53.93	223.8
rhyolite	15	10.19	57.44	melange	5	22.71	94.24
mud	14	8.85	49.90	rhyolite	4	38.55	160.0
limestone	12	9.36	52.78	breccia	3	49.88	207.0
SiltStn	10	5.81	32.74	metaSed	3	23.69	98.29
volcanics	8	8.42	47.45	AltSandStnMudStn	2	20.91	86.77
peat	7	7.60	42.83	gabbro	2	13.03	54.05
melange	6	14.92	84.10	mud	2	45.92	190.5
agglomerate	4	3.99	22.49	peat	2	26.73	110.9
till	4	5.33	30.02	tuff	2	55.93	232.1
AltSandStnMudStn	3	1.92	10.83	agglomerate	1	35.60	147.7
metaSed	3	5.24	29.55	scoria	1	409.09	1697.5
gabbro	2	1.90	10.69	silt	1	40.26	167.0
silt	2	17.29	97.44	till	1	52.95	219.7
tuff	2	11.92	67.18	tonalite	1	34.54	143.3
peridotite	1	21.67	122.2	volcanics	1	26.74	110.9
pyroclastics	1	27.38	154.3				
scoria	1	20.06	113.1				
tonalite	1	3.01	16.95				

Table 16 Summary of the area of individual Chemical4 subgroups and elements for which background concentration estimates are available

Chemical4 group	Area (km^2)	Elements captured
gravel	43717	As, Cu, Cd, Cr, Pb, Ni, Zn
sandstone	42199	As, Cu, Cd, Cr, Pb, Ni, Zn
schist	19747	As, Cu, Cd, Cr, Pb, Ni, Zn
mudstone	19313	As, Cu, Cd, Cr, Pb, Ni, Zn
greywacke	18513	As, Cu, Cd, Cr, Pb, Ni, Zn
sandstone Pakihi	14965	As, Cu, Cd, Cr, Pb, Ni, Zn
ignimbrite	12044	As, Cu, Cd, Cr, Pb, Ni, Zn
semischist	11505	As, Cu, Cd, Cr, Pb, Ni, Zn
conglomerate Maui	9494	As, Cu, Cd, Cr, Pb, Ni, Zn
mudstone Pakihi	8242	As, Cu, Cd, Cr, Pb, Ni, Zn
granite	6096	
limestone	5556	As, Cu, Cd, Cr, Pb, Ni, Zn
basalt	5298	As, Cu, Cd, Cr, Pb, Ni, Zn
alternating sandstone/mudstone	4241	As, Cu, Cd, Cr, Pb, Ni, Zn
till	3951	As, Cu, Cd, Cr, Pb, Ni, Zn
breccia	3727	As, Cu, Cd, Cr, Pb, Ni, Zn
diorite	3392	
andesite	3087	As, Cu, Cd, Cr, Pb, Ni, Zn
melange	2571	As, Cu, Cd, Cr, Pb, Ni, Zn
metasediment	2300	As, Cu, Cd, Cr, Pb, Ni, Zn
rhyolite	2271	As, Cu, Cd, Cr, Pb, Ni, Zn
conglomerate	1889	As, Cu, Cd, Cr, Pb, Ni, Zn
sand	1690	As, Cu, Cd, Cr, Pb, Ni, Zn
granodiorite	1642	
orthogneiss	1351	
alternating sandstone/mudstone Maui	1338	
agglomerate	1217	Cd, Cr, Pb, Ni, Zn
argillite	1131	Cd
alternating sandstone/siltstone Maui	1125	As, Cu, Cd, Cr, Pb, Ni, Zn
siltstone	904	As, Cu, Cr, Pb, Ni, Zn
peat	894	As, Cu, Cd, Cr, Pb, Ni, Zn
paragneiss	815	
gabbro	762	As, Cu, Cd, Cr, Pb, Ni, Zn
gneiss	655	
greenschist	576	

Chemical4 group	Area (km^2)	Elements captured
tonalite	459	As, Cu, Cd, Cr, Pb, Ni, Zn
loess	401	
siltstone Pakihi	382	
tuff	343	As, Cu, Cd, Cr, Pb, Ni, Zn
silt	322	As, Cu, Cd, Cr, Pb, Ni, Zn
mylonite	296	
peridotite	271	As, Cu, Cr, Pb, Ni
volcanics	268	As, Cu, Cd, Cr, Pb, Ni, Zn
mud	266	As, Cu, Cd, Cr, Pb, Ni, Zn
metavolcanic	213	
fill	206	Cu, Cr
amphibolite	178	
quartzite	138	
dacite	133	
monzodiorite	116	
unknown	89.0	
monzonite	49.6	
chert	49.0	
serpentinite	40.0	
pyroclastics	20.0	As, Cu, Cr, Pb, Ni
marble	16.9	
syenite	15.8	
pyroxenite	15.4	
hornfels	14.3	
metaplutonic	13.7	
scoria	11.6	As, Cu, Cd, Cr, Pb, Ni, Zn
migmatite	9.42	
cataclasite	8.29	
dolerite	5.64	
hornblendite	4.99	
hypabyssal intrusive	4.33	
metavolcanics	4.22	
shale	1.96	
gabbronorite	1.13	
lignite	0.16	

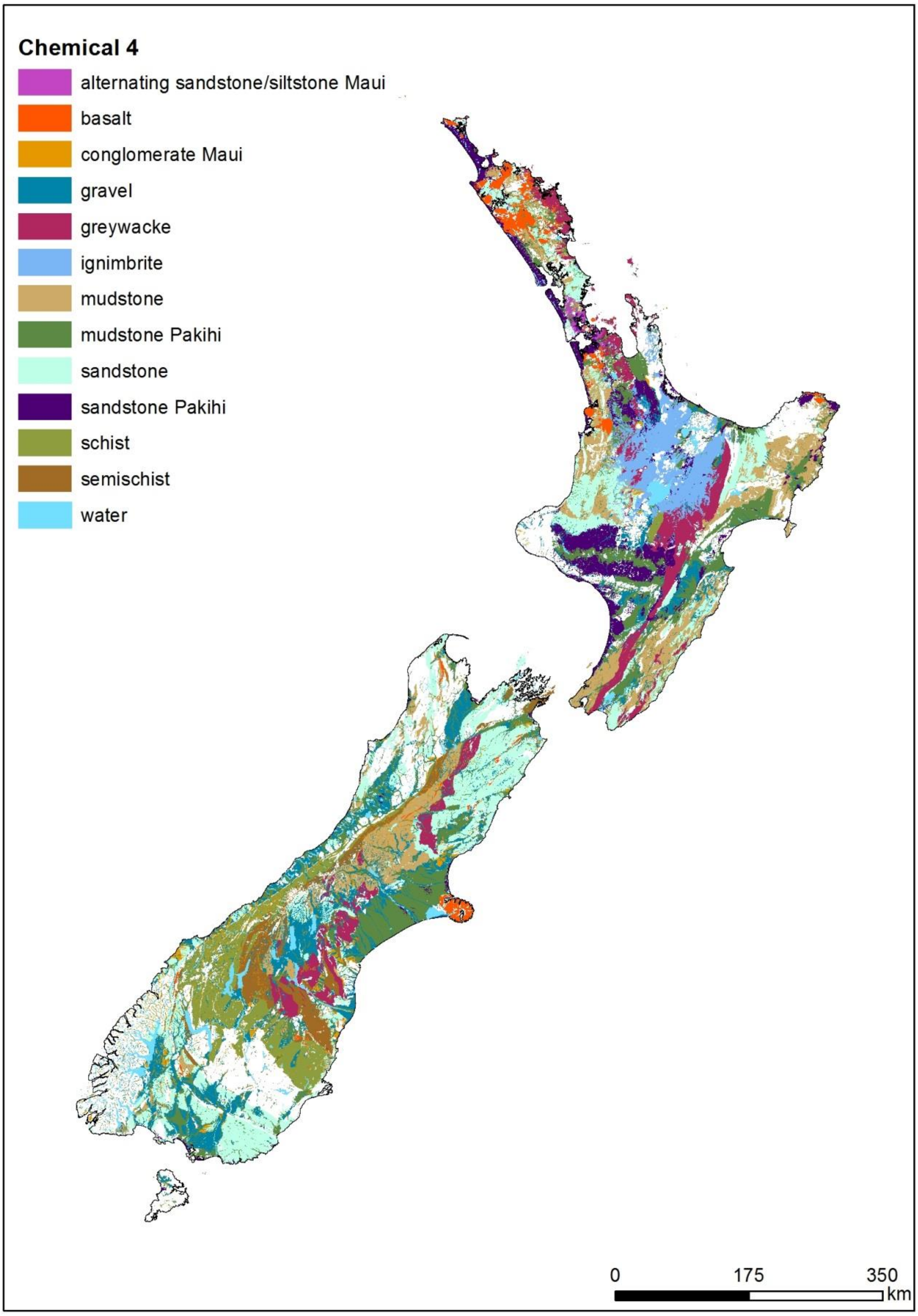

Figure 24 Areas for which predicted concentration ranges are typically available for Chemical4 subgroups (with n >30) from the QMAP geological map GIS dataset. Areas for which no data is available are shown in white.

4.7 Use of predicted background concentration data

It is recognised that that this hard copy presentation of the predicted background concentrations across New Zealand is challenging to use, as it is difficult to identify what rock grouping, and thus what background concentrations apply at a given location.

Options to simplify use include aggregating classes with similar concentrations to create larger domains, thus making it easier to identify the expected background concentration range for a given location (i.e. there would be 3–5 different 'domains'). Alternatively, the detailed information could be made accessible via the internet. This could be achieved through a downloadable dataset or an interactive online map. Such a system can be delivered through the LRIS portal (https://lris.scinfo.org.nz/) or the Our Environment website (http://ourenvironment.scinfo.org.nz/home).

The LRIS portal is primarily designed for GIS data discovery and download for GIS professionals. Uploading data layers into the portal is very quick and easy, but while data available via the LRIS portal can be queried (click on a polygon and determine attributes), the user interface is not designed to provide a full data browsing experience. In particular, there is minimal map visualisation capability. The effort involved in preparing and making a data layer available in the Our Environment website is much greater; however, Our Environment is specifically designed to deliver a good web browsing experience with good location and address searching, colour background maps, good quality data visualisation and reporting tools, all of which are optimised to deliver exactly the type of data browsing service required for this dataset.

4.8 Application of background concentration data

The most typical use for determining background concentration is to define an upper limit of the concentration that includes a high proportion of the data and is likely to exclude the very high results that would be associated with point source contamination. The upper confidence limit (UCL) for the 95th percentile is probably the most widely used threshold for determining upper limits for background concentrations (e.g. NREPC 2004; Cave et al. 2012), although the 99th percentile is also used if the dataset is sufficiently large (Diamond et al. 2009), with the observation that there is little difference between the UCLs of the 95th and 99th percentiles (Diamond et al. 2009). There are some statistical tools available to calculate the UCL of the 95th percentile, such as ProUCL developed by US Environment Protection Agency (US EPA 2013). Cave et al. (2012) also provide the code used for the statistical package R to determine the UCL of 95th percentiles. For contaminated land investigations, the upper 95th confidence limit of the mean may also be used for comparison of concentrations for a site under investigation with background concentrations. In this case, the upper 95th confidence limit of the mean of the background concentrations is the point of comparison. This approach is recommended by US EPA (2002), with the upper confidence limit (UCL) for the 95th percentile (in this case termed an upper tolerance limit, UTL) recommended to indicate whether a single sample is likely to be an outlier in the dataset (i.e. whether there is a hot spot, even if the site average is the same as the background concentration).

5 Urban soils and organic analytes

5.1 Introduction

In urban areas, it is inevitable that there will be background contamination arising from diffuse sources. Internationally, ambient background concentration is referred to in contaminated land guidance or regulations (e.g. BMU 1999; Cicchella et al. 2005; FMfE 2007; Diamond et al. 2009; DEFRA 2012). In New Zealand, while background concentrations are naturally occurring concentrations only, it would be unreasonable to expect a given site owner to remediate below ambient concentrations in urban areas. Of particular interest in the urban environment are polycyclic aromatic hydrocarbon (PAH) concentrations as these are derived from a number of diffuse anthropogenic combustion sources (e.g. vehicles, domestic woodburners). Similarly, lead may also be elevated due to the historical use of leaded petrol. In this section, we focus on PAH concentrations only.

Some regional councils have surveyed concentrations of contaminants in some urban areas (Table 17). For example, studies in Christchurch yielded an estimate for the upper ambient background concentrations for benzo(a)pyrene (BaP) of 0.595 mg/kg by using the upper confidence limit of the mean (Tonkin & Taylor 2007b). These studies have appropriately targeted parks, reserves, schools or other areas expected to have been minimally disturbed. However, it should be noted that some international studies have found that large park areas may be relatively unimpacted from urban diffuse pollution sources, and suggest concentrations measured in these locations may be more representative of natural concentrations (BGS 2011).

The organochlorine pesticide dichlorodiphenyltrichloroethane (DDT) was widely used in pastoral agriculture and horticulture in the 1950s–60s and while such uses largely ceased by the mid-1970s (Buckland et al. 1998), residues (primarily pp-dichlorodiphenyldichloroethylene, pp-DDE) still persist in agricultural soils (e.g. Boul 1995; Buckland et al. 1998; Gaw et al. 2006, numerous contaminated land site investigation reports). This historical, widespread use of DDT has resulted in the ubiquitous presence of DDT residues in soil that, given there are no naturally occurring sources of these residues, can trigger requirements under the National Environmental Standard for assessing and managing contaminants in soil for the protection of human health. This ubiquitous distribution of DDT residues may be considered as the ambient concentration of the residues.

5.2 Methods

5.2.1 Polycyclic aromatic hydrocarbons.

For the current study, soil type is assumed to not influence PAH concentrations (although higher organic soils may have retained more). Data were compiled and assigned to urban, provincial and rural categories. The urban category includes samples collected from major urban centres, (e.g Hamilton, Christchurch, Wellington); provincial towns includes samples collected from smaller regional centres, while rural includes samples from areas surrounded predominantly by agricultural land. Benzo(a)pyrene and BaP-equivalents (BaP-eq), determined using the toxicity equivalence factors provided by the Ministry for the Environment (MfE 2011c), were used in subsequent analysis as these are the common measures used to assess PAH contamination in soils.

The statistical package R was used for descriptive statistics including determination of 95% upper confidence limits of the 95th percentile using bootstrapping technique.

Table 17 Summary of studies to determine background concentrations of polycyclic aromatic hydrocarbons

City/town	Number of samples	Source
Christchurch	22	Tonkin and Taylor 2006, 2007a, b
Hamilton	5	Waikato Regional Council unpublished
Waikato provincial towns (10 towns)	4–5 per town	Waikato Regional Council unpublished
Wellington region (urban provincial, rural)	40	URS (2003)
Bay of Plenty region (agricultural land)	42	Bay of Plenty Regional Council unpublished

5.2.2 DDT residues

This study compiled existing data from the sources identified in Table 18. Due to the lack of availability of data for individual sites, this data is provided as a simple summary of concentration ranges found in different land uses on a regional basis. This summary builds on an earlier summary (PDP 2007).

Table 18 Summary of studies to determine concentrations of DDT residues in New Zealand

Region	Number of samples	Source
Auckland	43	ARC 2002
Canterbury	Collation of existing data	PDP 2007
Canterbury	Collation of contaminated site investigations	Environment Canterbury unpublished data
Auckland, Waikato and Tasman	28 (Auckland), 35 (Waikato), 20 (Tasman)	Gaw et al. 2006
Across New Zealand	35 total; in urban and provincial centres (23) and indigenous forest and grassland (12)	Buckland et al. (1998)
Bay of Plenty	128	SEM 2005
	42	Bay of Plenty Regional council unpublished (2014 data)

5.3 Results

5.3.1 Polycyclic aromatic hydrocarbons

Benzo(a)pyrene concentrations in the urban, provincial and rural categories are shown in Figure 25. As can be seen in Figure 25a, there is quite a spread in the BaP concentrations for the provincial category. Further analysis of the data highlights two towns (Thames and Waihi) that have remarkably high BaP concentrations (Figure 26). These results appear to be exceptional and were excluded from subsequent determination of background concentrations. Further investigation is warranted to determine the cause for these higher concentrations. Similarly, there was a single high point in the urban dataset that was removed from some further data analysis. Benzo(a)pyrene concentrations from samples collected from the centre of parks in urban and provincial towns were used in the primary data analysis. However, the concentration of PAHs will vary across a park, largely in relation to the proximity to a roadway. Typically concentrations will be higher closer to the road, including under trees lining a park (Figure 27).

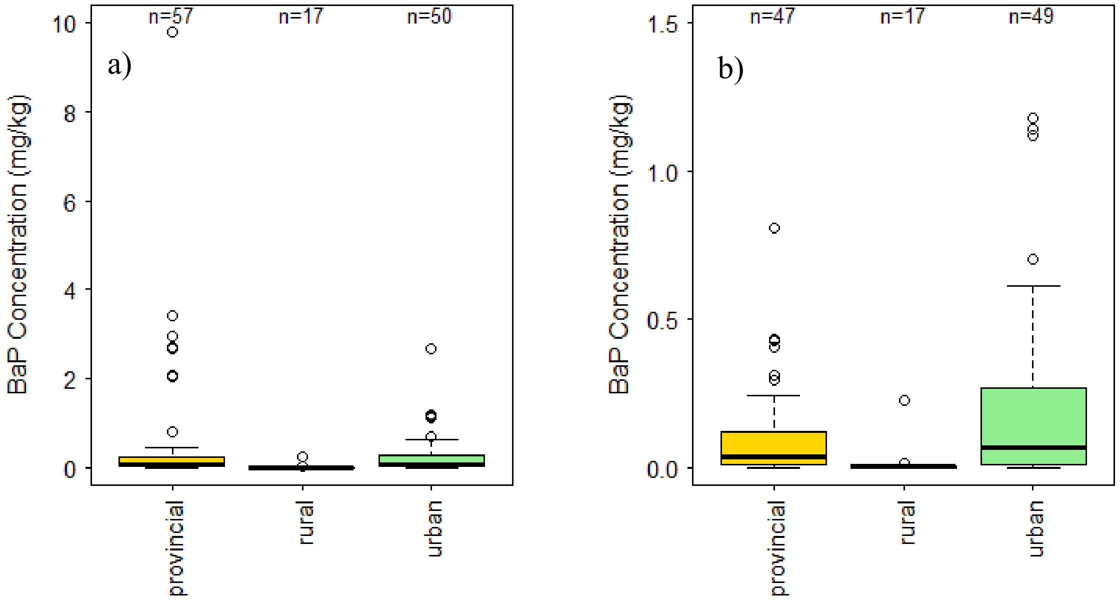

Figure 25 Benzo(a)Pyrene (BaP) concentrations in provincial, rural and urban sampling locations, a) all data, b) excluding two provincial towns with high BaP concentrations (Tha and Wai- see Figure 34) and urban outliers. Number of samples (n) for each location is shown at the top of the graph. Note that scale of vertical axes are different.

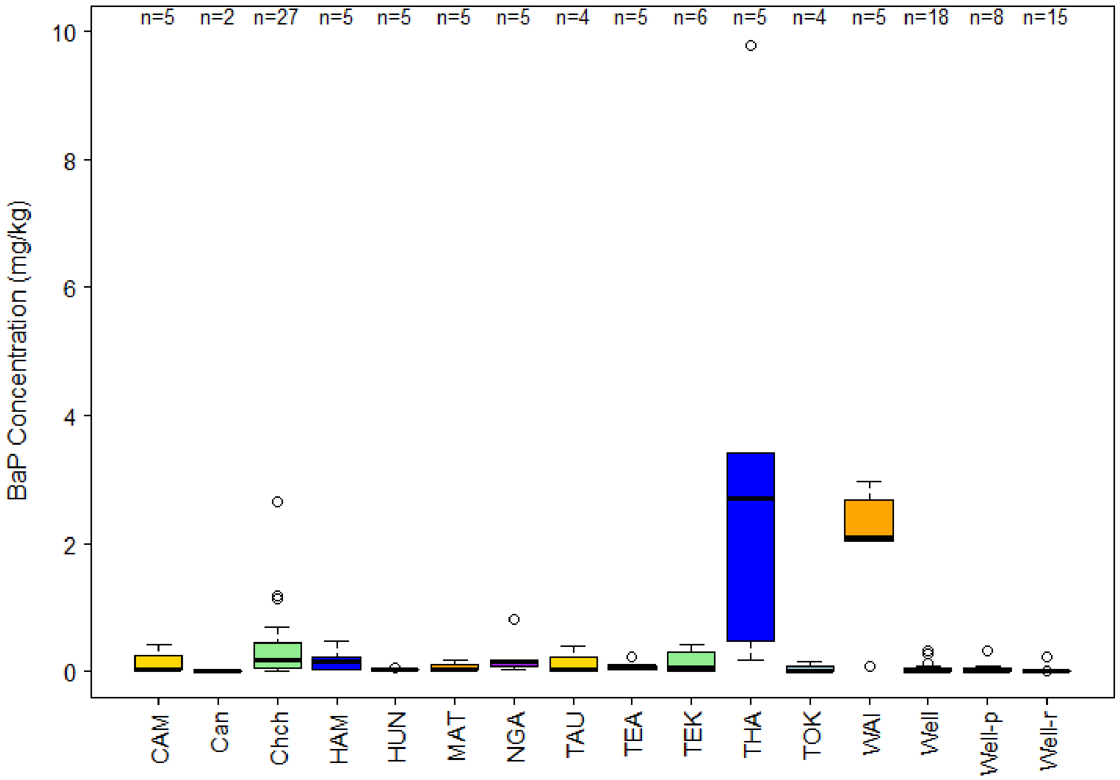

Figure 26 Benzo(a)Pyrene (BaP) concentrations in samples collected from individual provincial towns and Hamilton (Ham) in the Waikato region, urban Christchurch (Chch) and rural Canterbury (Can), and urban (Well), provincial (Well-p) and rural (Well-r) locations in the Wellington region. Number of samples (n) for each location is shown at the top of the graph.

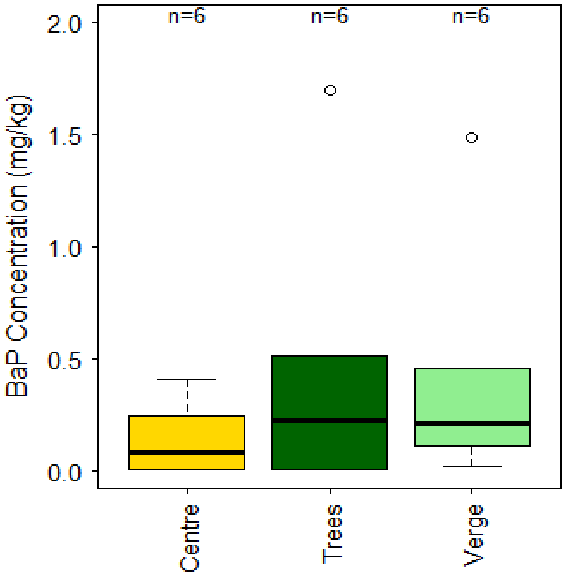

Figure 27 Benzo(a)Pyrene (BaP) concentrations in park locations, from the centre of the park, under trees lining the park and at the verge (close to roadside).

The statistical description of BaP and Bap-eq concentrations determined for urban, provincial and rural New Zealand are shown in Table 19.

Table 19 Summary of Benzo(a)Pyrene (BaP) and BaP-equivalent (BaP-eq) concentrations (mg/kg) determined for urban, provincial and rural categories across New Zealand

Analyte	Concentration statistic	Urban (n = 49)	Provincial (n = 47)	Rural (n = 17)	Rural 2[1] (n = 59)
BaP	Mean	0.2	0.102	0.0163	0.019
	Median	0.066	0.034	0.003	0.015
	Range	0.0004-1.18	0.0004-0.807	0.0004-0.225	0.0004-0.225
	95th	0.95	0.422	0.056	0.036
	99th	1.16	0.634	0.191	0.121
	95thUCL	1.16	0.693	0.225	0.063
BaP-eq	Mean	0.389	0.16	0.026	NA
	Median	0.1089	0.052	0.0055	
	Range	0.001-4.36	0.002-1.22	0.001-0.355	
	95th	1.43	0.64	0.09	
	99th	1.82	0.96	0.302	
	95thUCL	1.83	1.05	0.3551	

[1] Rural 2 includes data from 42 agricultural and background sites in the Bay of Plenty region, for which PAH concentrations, including BaP, were all below the detection limit and included at half the detection limit of 0.05 mg/kg.

Combining the data from Hamilton, Christchurch and Wellington provides provisional upper limits (based on the 95th UCL of the 95th percentile) for BaP and BaP-eq in urban soils of 1.2 mg/kg or 1.8 mg/kg, respectively. These are higher than background levels for BaP and BaP-eq (0.595 mg/kg and 0.922 mg/kg) previously determined for Christchurch (Tonkin & Taylor 2007). This is largely due to the difference in the selection of the metric used to define background with the 95th UCL of the *mean* concentration used by Tonkin & Taylor (2007). It is also noted that PAH concentrations between cities varied, with Wellington having generally lower concentrations than Christchurch or Hamilton, and further sampling of urban areas is required to establish more robust estimates of ambient urban BaP concentrations. As a comparison with international data, the concentrations were markedly below 'normal background concentrations' (defined as the upper 95th confidence limit of the 95th percentile concentration) of BaP in urban environments in UK of 3.6 mg/kg (Johnson et al. 2012).

The provisional upper limits for ambient BaP and BaP-eq concentrations in provincial towns are 0.7 mg/kg and 1.1 mg/kg, respectively. Two provincial towns appeared to have markedly elevated BaP concentrations and were excluded from the determination of these upper limits.

There were limited data for PAH concentrations in rural environments, and provisional upper limits for ambient BaP and BaP-eq concentations are 0.06 and 0.09 mg/kg, respectively. In this case, the upper limit is based on the 95th percentile of the data. This is based on data mainly from the Waikato and Wellington regions, with two sites from Canterbury. PAH analyses were also undertaken on 42 rural soils in the Bay of Plenty region, although all were below detection limits. If this data is included (with the concentration set at half the detection limit, as is conventionally done) it significantly boosts the dataset and changes the statistical descriptors of the BaP concentrations (Table 19).

Further sampling and analysis is required in different environments to develop more robust estimates of ambient background concentrations of PAHs and BaP.

5.3.2 DDT residues

A summary of ∑DDT and pp-DDE, the primary DDT residue, concentrations from various studies are shown in Table 20. These studies provide an indication of the variability and range in concentrations found in different agricultural land uses in different regions. The sample numbers are typically low, and it remains unclear whether the difference between land uses in a given region truly reflect a land use difference. Some general observations can be made: higher maximum concentrations are found in orchard or vineyards while lower concentrations are found in market gardens and pastoral land, and concentrations in the Bay of Plenty region are lower than other regions for a given land use. However, a systematic approach to collating data from individual sites is required to provide more robust estimates of concentrations in the different land uses and to determine any regional differences.

Some studies undertook analysis of DDT residues at control (background) sites, with DDT residues occasionally being detected (SEM 2005, Gaw et al. 2006). The study of Buckland et al. (1998) also provides data for indigenous forestry and grassland sites. Much lower detection limits were used in this study, resulting in detection of DDT residues at concentrations below detection limits typically used in more recent studies. Buckland et al. (1998) also provide some of the only accessible data on DDT residues in urban environments, which show that Christchurch appears to have higher concentrations than other towns and cities (Table 21).

Table 20 Summary of DDT and pp-DDE concentrations (mg/kg) in different land uses in different regions compiled from different studies. N = number of samples

Land use	Region		∑DDT concentration (mg/kg)				pp-DDE concentration (mg/kg)			
		N	Min	Max	Med	Mean	Min	Max	Med	Mean
Orchard	Auckland[1]	12	<0.005	24.2	4.71	2.23	<0.005	6.7	1.26	1.69
	Waikato[1]	7	0.73	34.5	3.22	8.39	0.274	12.9	1.97	3.83
	Tasman[1]	5	1.49	7.14	3.09	3.66	0.831	4.27	1.49	2.04
	Bay of Plenty (2005)[2]	35	<0.01	1.87	0.01	0.08	<0.01	3.46	0.03	0.187
	Bay of Plenty (2014)[3]	5	<0.03	0.015	0.015	0.015	<0.005	0.012	0.0025	0.0044
	Hawkes Bay[4]		0.02	15.3						
	Canterbury[4]		<0.03- 24.1	24.1						
	Canterbury[5]	41		30.5	0.11	1.1				
Vineyard	Auckland[1]	5	0.26	2.84	1.1	1.36	0.41	1.3	0.54	0.708
	Waikato[1]	7	<0.03	1.26	0.06	0.23	<0.003	0.457	0.032	0.092
	Hawkes Bay[4]	5	0.02	0.35						
	Canterbury[4]		0.225	10.09						
	Canterbury[5]	5		1.72	0.23	0.3				
Market garden	Auckland[1]	8	0.09	0.91	0.19	0.32	0.039	0.53	0.08	0.145
	Waikato[1]	7	0.04	1.68	0.39	0.63	0.016	1.11	0.165	0.296
	Tasman[1]	5	0.07	1.16	0.12	0.32	0.036	0.574	0.061	0.171
	Bay of Plenty (2005)[2]	14	<0.02	0.095	0.025	0.032143	<0.01	0.85	0.015	0.081
	Hawkes Bay[4]	6	<0.01	0.12						
	Canterbury[4]		0.003	0.74						
	Canterbury[5]	88		2.66	0.21	0.34				

Land use	Region		∑DDT concentration (mg/kg)				pp-DDE concentration (mg/kg)			
		N	Min	Max	Med	Mean	Min	Max	Med	Mean
Pasture	Auckland[1]	3	<0.03	<0.03	<0.03	<0.03	<0.005	0.009	<0.005	<0.005
	Waikato[1]	7	<0.03	0.75	0.23	0.08	<0.005	0.223	0.05	0.096
	Tasman[1]	5	<0.03	1.3	0.49	0.11	<0.005	0.638	0.049	0.242
	Bay of Plenty (2005)[2]	24	<0.02	1.19	0.025	0.127708	<0.01	0.98	0.005	0.090208
	Bay of Plenty (2014)[3]	27	<0.03	1.69	0.03	0.223148		1.3	0.027	0.141667
Crop	Bay of Plenty (2005)[2]	14	<0.02	0.095	0.025	0.032143	<0.01	0.06	0.005	0.011071
	Bay of Plenty (2014)[3]	5	<0.03	0.06	0.015	0.024	<0.0005	0.042	0.0025	0.0119
Background	Waikato[1]	7		0.05				0.015		
	Bay of Plenty (2005)[2]	25		<0.03				0.11		
	Bay of Plenty (2014)[3]	5		<0.03				0.006		
Sports turf	Canterbury[5]			4	0.26	0.604				

[1] Data from Gaw et al. (2006); 2 Data from SEM (2005); 3 Unpublished data from Bay of Plenty Regional Council; 4 Data from PDP 2007; 5 Data from contaminated land investigation reports compiled by Environment Canterbury.

Table 21 Summary of ∑DDT and pp-DDE concentrations (mg/kg) from Buckland et al. (1998). N = number of samples

Location	N	∑DDT concentration (mg/kg)				DDE concentration mg/kg)			
		Min	Max	Med	Mean	Min	Max	Med	Mean
Christchurch	6	0.233	0.853	0.347	0.431	0.119	0.469	0.19	0.23
Auckland	9	0.002	0.070	0.024	0.032	<0.001	0.038	0.004	0.013
Provincial	8	0.001	0.245	0.024	0.053	<0.001	0.087	0.009	0.023
Grassland	5	0.001	0.004	0.002	0.002	<0.001	0.002	0.001	0.001
Forest	7	0.000	0.005	0.002	0.003	<0.001	0.003	0.001	0.001

While some general comment can be made on the apparent difference in concentrations related to land use, concentrations will ultimately be dependent on historical pesticide usage and may be highly variable. While the concentrations for a given land use could be better characterised with further sampling and analysis, the relevance of doing so is perhaps debateable. Whether some action should be taken in response to the DDT residues at a site is ultimately depends on whether any effect might occur. In this context, ecological considerations are the most relevant, including potential bioaccumulation in the food chain.

6 Database requirements

In determining background concentrations of trace elements based on existing data for this project, substantial effort has been placed in collating trace element data from regional councils across New Zealand. Much of this information is collected from sites used for State of the Environment monitoring, for which additional soil quality parameters (pH, mineralisable N, total C, total N, Olsen P, bulk density, macroporosity) are collected – and which informs the soil quality indicators website, SINDI (https://sindi.landcareresearch.co.nz/). Partly to ensure the ongoing utility of such collated material and partly recognising that additional information is available for the sites for which trace element data is available, consideration was given to how trace element data could be captured in a database, and subsequently used. It is also recognised that there are various ongoing conversations about the capture of soils data in databases, including within the *Environmental Monitoring and Reporting* project, as well as in the context of a National Soils database. These conversations are ‘bigger’ than the current project, and the purpose of capturing information within this project is to provide an input to these larger, and ongoing conversations. Specifically, this section is intended to inform considerations of what and how trace element (or other contaminant) data should be captured, and to highlight that processes for the compilation of trace element data from disparate sources have been developed, and have been used to compile current trace elements data from regional council sources. Further, the recommendations outlined in this report draw on the experience of compiling that data to ensure that subsequent data input is easier, and that some data outputs are identified.

6.1 Data capture process

As discussed in section 3.2.1, the primary data sources were the results of State of the Environment soil quality monitoring undertaken by regional councils and the results of specific studies to determine background concentrations in different regions. The latter sampling was typically undertaken to inform contaminated land management. Both sets of results were provided as excel sheets by various regional councils, in varying formats. Data from these spreadsheets was compiled using the programme 'R' (R Core Team 2015) to read in the data (See section 3.2.1 for more details). Such a process provides transparency in any transformations of the original data to provide collated data and/or provides flexibility in how data is collated. However, the efficiency of data capture would be greatly enhanced by the provision of data in a consistent format, the consistent use, or at least identification of, the spatial co-ordinate system and, in particular, consistency in land-use categorisation. Further discussion on land-use categorisation is provided below.

6.2 Potential application of information

The purpose of capturing data in a database is to maximise the use of such data by linking with data from other sources and enabling national assessment of regionally collected data. From a soil quality perspective, the key use of trace element data is to assess trace element concentrations in relation to land use, and to assess change over time to determine whether some action is required to avoid negative effects on the environment. National assessment of regionally collected data or collation of data from disparate studies can enable trends or changes that are not otherwise apparent to be identified.

From a contaminated land management perspective, a database provides a means to capture information relevant for determining background soil concentrations. A database capturing contaminated land investigations could provide a national picture of the likelihood of contamination of certain land-use activities, to streamline the management of contaminated land and enable a focus on the assessment and management of higher risk sites.

Soil parameters, specifically clay content and type, organic matter (or total carbon) content, cation exchange capacity and pH, can influence the toxicity of contaminants (through their influence on bioavailability). Some of these parameters are assessed at sites for which trace element data is available, and linking soil quality data to trace element data for the same sites and times provides an opportunity for enhanced analysis of the data, and is relevant to both soil quality and contaminated land management.

The desired outputs and use of data in a database are critical to informing data input, and to streamlining that process, so these need to be clearly established from the outset. It may be that two different databases are developed, with a soil quality database containing only a subset of information from the contaminated land investigations, specifically that data relating to background soil concentrations and agricultural land.

6.3 Analytes for consideration

From a soil quality perspective, the standard ‘contamination’ suite of heavy metals (As, Cu, Cd, Cr, Hg, Ni, Pb and Zn) is useful to provide measures on soil quality and background or ambient distribution of potentially toxic elements. The organic analytes, PAHs and DDTs (sum and individual isomers) are also useful measures to assess soil quality in relation to ambient distribution (e.g. PAHs for urban ambient environments, DDTs for agricultural settings) as well as potential toxicity. However, currently broad-scale surveys of organic contaminants in urban or rural environments are limited.

Analyses of essential elements including (B, Co, Cu, I, Fe, Mn, Mo, Se and Zn) provide a measure of soil quality in relation to plant and livestock nutrition, and are useful for identifying where deficiencies may occur.

6.4 Land-use categorisation for describing trace element concentrations and database capture – soil quality

There is currently considerable discussion within the research community regarding land-use classification for a range of purposes. Historically, land-use classifications have been developed for different purposes and have led to classifications that are not suitable for other uses. Ideally, a classification system can be used at different levels of detail to provide the desired information for different purposes in a consistent fashion across New Zealand. Specifically, a soil quality classification system should enable greater integration and interpretation of data collected for different purposes (e.g. trace element concentration and soil biodiversity). While on-the-ground information is necessary to provide detailed information and ground-truthing for satellite imagery, use of spatial databases and satellite imagery provides a means for describing land use in locations for which on-the-ground information is not available and/or for providing assessment on a broader regional or national scale.

Land use is used in regional council State Of The Environment soil quality monitoring to classify sites from which soil samples are collected, although classifications are not used consistently across regions. The land-use classification shown in Table 2 is based on experiences obtained in assessing cadmium concentrations in NZ soils (Cavanagh 2014), and determining background concentrations described earlier. This classification has included consideration of classification systems used in the Landcover database, regional council soil quality monitoring, MfE carbon monitoring (LUCAS), and New Zealand Standard Industrial Output Categories (NZSIOC) used by Statistics NZ. The classification is primarily focussed on delineating agricultural land uses that have similar ‘trace element’ applications, (e.g. fertiliser application, types of pesticides and veterinary chemicals used).

Further differentiation of dairy and drystock systems is potentially useful, particularly for delineating between intensive (high input) and extensive (low input) systems, although further work is required to ascertain how this can be most effectively obtained in a consistent manner. The classification of farm classes used by Beef and Lamb NZ (http://beeflambnz.com/farm-classes/) may be useful to delineate intensive (high input) and extensive (low input) pastoral systems and key pastoral system ‘types’ (e.g. mixed cropping), particularly if such information is available across the country. Table 23 provides a description of the farm classes, a grouping that is likely to be relevant from a trace element perspective, and the anticipated Land Cover Database (LCDB) classification (the use of LCDB information could provide a means for national assessment). It should be noted that 'intensive' and 'extensive' are often used to describe the areal extent of a farming system, as opposed to inputs, and there is potentially an overlap between a high input (intensive) system and large area (extensive) being farmed.

Land-use history for a given piece of land will be a significant factor influencing concentrations at a given point of time, and may be more significant than the current land use. However, current land use is most often used for grouping results of soil analyses (e.g. Cavanagh 2014; regional council surveys) as sufficient information to provide alternative classifications based on land use history is not typically available. This is a source of error in interpreting the influence of land use on trace element concentrations. It is also noted that classification of agricultural land for soil quality data collected in Canterbury is slightly different to that in other regions as it includes classifications based on the duration of cropping or pastoral use (Lawrence-Smith & Tregurtha 2013). At times this classification is challenging to neatly categorise land use according to Table 2, as sometimes the land use classification does not distinguish between pastoral use and cropping use.

Table 22 Description of farm classes used by Beef and Lamb NZ, potential grouping as intensive or extensive pastoral systems, and anticipated LCDB cover class

Farm class	Description	Grouping	LCDB cover class
1. South Island high country	Extensive run country at high altitude carrying fine wool sheep, with wool as the main source of revenue. Located mainly in Marlborough, Canterbury and Otago.	Extensive pasture	Low-productivity grassland
2. South Island hill country	Mainly mid-micron wool sheep, mostly carrying between two and seven stock units per hectare. Three-quarters of the stock units wintered are sheep and one-quarter beef cattle.		
3. North Island hard hill country	Steep hill country or low fertility soils with most farms carrying 6–10 stock units per hectare. While some stock are finished, a significant proportion are sold in store condition.	Extensive pasture	High-productivity grassland?
4. North Island hill country	Easier hill country or higher fertility soils than Class 3. Mostly carrying between 7 and 13 stock units per hectare. A high proportion of sale stock sold is in forward store or prime condition.	Intensive pasture	High-productivity grassland
5. North Island intensive finishing farms	Easy-contour farmland with the potential for high production. Mostly carrying between 8 and 15 stock units per hectare. A high proportion of stock is sent to slaughter and replacements are often bought in.	Intensive pasture	High-productivity grassland
6. South Island finishing-breeding farms	A more extensive type of finishing farm, also encompassing some irrigation units and frequently with some cash cropping. Carrying capacity ranges from 6 to 11 stock units per hectare on dryland farms and over 12 stock units per hectare on irrigated units. Mainly in Canterbury and Otago. This is the dominant farm class in the South Island.	Intensive pasture	High-productivity grassland/Cropping?
7. South Island intensive finishing farms	High-producing grassland farms carrying about 10–14 stock units per hectare, with some cash crop. Located mainly in Southland, South and West Otago.	Intensive pasture	High-productivity grassland
8. South Island mixed cropping and finishing farms	Located mainly on the Canterbury Plains. A high proportion of their revenue is derived from grain and small seed production as well as stock finishing.	Mixed cropping livestock	High-productivity grassland/Cropping

Finally, consideration of the farming system within which a given site is located is also useful for describing trace element concentrations – particularly for mixed cropping systems, which may include grain crops, fodder crops and pasture (drystock, dairy) on a rotational basis. Trace element concentrations in these systems are likely to be different to a site that has remained in pasture, and also leads to misclassification of the land use. For example, a site on which grain crops are present could be classified as an arable site, although may be more appropriately classified as drystock or dairy (or ideally a mixed cropping system); a site on which kale is grown as a fodder crop could be classified as a horticultural site although is likely more appropriately classified as drystock or dairy. These 'misclassifications' appear to relate primarily to drystock or dairy systems, and a suggested farm system classification is given in Table 23. Further discussion is required as to whether a minimum proportion of the farm or frequency of other crops needs to be specified to distinguish between pastoral and mixed cropping, or between different mixed cropping farms.

Table 23 Potential farm system categorisation for describing trace element concentrations on drystock and dairy farms1

Farm system	Description
Drystock – pastoral	Pastoral farm system.
Drystock – mixed cropping	Pasture and feed crops (including dairy support) grown on rotational basis – is there a need to consider frequency of crop rotation/pasture renewal or proportion of farm used for other crops to further delineate?
Dairy – pastoral	Pastoral farm system
Dairy – mixed cropping	Pasture and feed crops grown on rotational basis – is there a need to consider frequency of crop rotation/pasture renewal or proportion of farm used for other crops to further delineate?

[1] It is anticipated that by considering farm system, sites previously identified as pasture would be able to be assigned to drystock or dairy and thus improve land use classification

6.5 Land-use categorisation for describing trace element concentrations and database capture – contaminated land management

The previous land-use categorisation is useful in the context of determining background concentrations for use in contaminated land investigations. In addition, some specific studies have determined regional background concentrations, and there is also potential to use information captured in contaminated land investigations to further contribute to regional background concentrations, *if the data could be captured in a systematic manner*. An example is determining ambient concentrations of DDTs in agricultural land.

Similarly, there is wider application in providing an empirical evidence base for the likelihood of contamination arising from specific activities and duration. This is particularly relevant in the context of the subdivision of ex-agricultural land for residential purposes, particularly for land that falls into the following categories on the Hazardous Activities and Industry list (MfE 2011d):

- A10. Persistent pesticide bulk storage *or use* including sport turfs, market gardens, orchards, glass houses or spray sheds
- I. Any other land that has been subject to the intentional or accidental release of a hazardous substance in sufficient quantity that it could be a risk to human health or the environment.

For category I, the application of superphosphate is a key activity that triggers investigation as it causes increased Cd concentrations, which in ex-agricultural land can exceed the Soil Contaminant Standard for rural residential land use of 0.8 mg/kg (MfE 2011a, Cavanagh 2014). However, for many sites investigations triggered solely for this reason may pose an unnecessary cost for the landowner and some suggestions have been put forward on how to identify land for which Cd investigation is relevant (Ferry et al. 2014). A database to capture investigations undertaken to date and subsequent data analysis would help to substantiate when Cd investigation is appropriate. A similar situation arises for sites captured under category A10. Specifically, the historical subdivision of orchard land can lead to a significant number of sites being potentially contaminated, although there may actually be no or low concentrations of contamination on site. Currently, the only means to remove the label of 'potentially contaminated' is to undertake a site investigation, which poses a cost for the landowner. Once again, a database to capture investigations undertaken to date and subsequent data analysis would help to substantiate when it is relevant to undertake investigations.

6.6 Soils data to interpret trace element concentrations

Soil parameters, specifically clay content and type, organic matter (or total carbon) content, cation exchange capacity (CEC) and pH, provide useful additional measures that can help describe the toxicity (through their influence on bioavailability) of trace elements in relation to soil properties. Specifically, linear multiple regressions have been developed to describe the influence of soil properties on toxicity or plant uptake (e.g. EC 2008, SCEW 2013). Of these parameters, pH and total C are routinely measured as part of State of the Environment monitoring, although CEC and clay content and type are not. Thus, linking soil quality data to trace element data captured for the same sites and time provides an opportunity for enhanced analysis of the data to provide an assessment of risk. However, it should be noted that CEC and clay content and type are not routinely measured, thus the analysis is limited. While CEC is often considered to be correlated with total C, further analysis is required to substantiate this relationship in New Zealand soils.

6.7 Key considerations for database development

6.7.1 Data collation and ongoing maintenance

There are effectively two data streams that can inform a database that captures data on trace elements (and other contaminants) in soil – soil quality monitoring and contaminated land investigations.

The collation of soil quality monitoring data is comparatively easy, as it typically comes from one source only, that is regional council soil quality monitoring groups (although it may be undertaken for different purposes, e.g. State of the Environment Reporting, specific investigations on soil quality), and thus just requires the presentation of data in a consistent format (see below) to enable input into a database.

To systematically capture the data from contaminated land investigations is not an insignificant task as it is typically received in a report format (cf. data tables) from multiple parties (e.g. consultants) by multiple parties (territorial authorities and regional councils). Thus consideration needs to be given to how the data provided in the reports can be readily input to a database or whether additional data (i.e. data tables) are required. Further, systems need to be developed at both territorial authority and regional council level to ensure that the data are being captured in a systematic manner. In the first instance, the capture of the site location, and the fact that an investigation report exists, would be helpful. Additional studies, such as to inform local background concentrations, may also be available. It is anticipated that these will usually be regional council driven, and thus it would be comparatively easy to ensure data consistency and format to facilitate input into a database.

Funding for the construction, data input and ongoing maintenance of a database is a critical consideration for the success of any developed database. Consideration of how data from multiple sources (i.e. different regional councils) is input to a (central) database or is used (no central database) is an ongoing conversation in some of the other related projects (see section 6.8). Ongoing data input is particularly important for contaminated land investigations, for which remedial activities may have occurred on a given site, changing the concentrations initially reported.

6.7.2 Inputs

To facilitate the capture of trace element (or other contaminant) data in a database, information should be provided in a consistent format with the following items captured:

- Unique sample identifier
- Date of sample collection
- Location (GPS)
- Spatial co-ordinate system used
- Soil sample depth

- Sample collection method
- Soil type (can also be obtained through interrogation of spatial databases, although some error exists in relation to the scale used in the spatial databases. Misclassification may also arise if person is not sufficiently experienced in soil classification)
- Land use at point of sample collection (using land-use classes defined in previous section)
- Farm system (if drystock or dairy)
- Laboratory analytical methods used for all analytes, and laboratory name
- Analyte concentration
- Detection limits for all analytes.

In addition to the above, information on land-use history gathered from contaminated land investigations, in particular the timing and duration (in years) of activities such as persistent pesticide application and phosphate fertiliser application, would ideally be captured where available.

Of the above, consistent use of agreed land-use, and farm-system categories is critical to effectively use the data.

The analytes captured will depend on the purpose of investigation; however, the standard ‘contamination’ suite of heavy metals (As, Cu, Cd, Cr, Ni, Pb and Zn) is a minimum suite to provide measures on soil quality and background or ambient distribution of potentially toxic elements. Mercury may also be a useful addition to the contaminant suite for some regions. From a broader perspective, essential elements, including B, Co, Cu, I, Fe, Mn, Mo, Se and Zn, would be useful to capture to provide an assessment of potential deficiencies. Finally, the inclusion of the ubiquitous organic contaminants, PAHs and DDTs would also be useful to determine ambient distribution (e.g. PAHs for urban ambient environments, DDTs for ambient concentrations in agricultural settings) as well as potential toxicity.

6.7.3 Outputs and data analysis

The minimum output generated from a database should be trace element concentration data for land-use classes presented at regional and national levels. It should also be possible to assess changes in concentration over time on a regional and national level. Ultimately, it should be possible to interrogate the database to attain additional data collected by other sources – in a way that does not contravene any privacy or usage conditions relating to their collection. Other outputs could include ‘risk maps’ to identify areas where, due to the soil properties and/or soil contaminant concentration, there is a high risk of negative effects on the environment, food quality, or human health.

Systematic capture of contaminated land investigations would enable the development of an empirical database to provide evidence of the likelihood of contamination from certain land-use activities, and would streamline investigations to target higher risk sites.

The desired outputs and use of data in a database are critical to informing data input, and to streamlining that process, so these need to be clearly established from the outset. It may be that two different databases are developed, with only a subset of information from the contaminated land investigations (i.e. data that relates to background soil concentrations and agricultural land) being brought into a soil quality database.

6.8 Related projects

As noted earlier, there are ongoing projects for which the capture and use of trace element and other contaminant data in a database are relevant. These are:

Environmental Monitoring and Reporting (EMAR) (Land Monitoring Forum/soil quality driven) – Discussion in this document relates to two of the three themes being considered in the EMAR project – land cover & use, and soil properties & processes. The development of Eco-SGVs also helps to inform the third theme – soil movement & protection – if protection includes protection of soil quality.

Data consistency project (contaminated land management driven) led by Environment Canterbury – to facilitate consistent classification of sites in council contaminated land registers, and high quality of information informing classification.

National soils database – Landcare Research/National Land Resource Centre

7 Summary and Conclusions

7.1 Background concentrations – trace elements

The use of different spatial databases in the data analysis provided an assessment of the utility of those databases for the determining background concentrations. QMAP was the primary database used, as it provided potential explanatory variables for the greatest number of sampling location. While the *rock-type-of-fines* parameter from S-Map shows promise as a potential explanatory variable for determining background concentrations, the limited coverage of S-Map prevented the further use of this database. Similarly, despite national coverage of the Fundamental Soils Layer, the classification of ‘town’ in *Toprock* (used to indicate urban areas) reduced the number of samples that could be used in subsequent analysis, and thus prevented further use of this parameter and database. The most data were retained using the different QMAP parameters, and thus QMAP was the primary database used.

Land-use effects were evident for most trace elements in the regional council dataset, with the effects observed differing for the individual trace elements. The significant differences noted for Cr and Ni are not considered attributable to anthropogenic land use, but rather the location of the land use. Similarly, the significantly lower concentrations of most trace elements for forestry are attributed to location as opposed to the land use. There was some evidence for land-use effects in the GNS Science Southland-Otago dataset, although these differences were small and the low number of samples in the background land-use class reduced the robustness of estimates produced for this class. Thus, land-use classes were aggregated for subsequent analysis of the GNS Science Southland-Otago dataset.

Preliminary analysis revealed that there was no bias in the estimates developed from each individual dataset. Thus, for the final analysis, the two individual datasets were combined to provide a fuller dataset, and the land uses that were not significantly different from background were also combined.

The aggregation of data from selected land uses and of the two datasets resulted in spatial correlation effects being evident in the data. This necessitated a slight change in the analysis to account for this effect. However, the general modelling approach remained the same, and the rock group-based parameter *Chemical 4* was found to provide the best fit. The effective median, 5th and 95th quantile estimates of the background concentration for the different *Chemical4* classes were developed for all factors for which data was available. Estimates for factors with n <30 are considered less reliable and for n <10 unreliable.

The predicted concentration ranges for a number of different subgroups are similar, suggesting these groups could be combined to provide a simpler presentation of predicted background concentrations across New Zealand. However, for some subgroups, the underlying data is localised to a particular area. In some cases, this is due to the distribution of that subgroup, in other cases it is simply a factor of where samples have been collected, and raises questions about the representativeness of the predicted concentrations.

Reliable predicted background concentration estimates are available for a significant amount of New Zealand, with the largest gaps existing in Fiordland and the north-west coast of the South Island. Granite, followed by diorite are the rock group*s* with the greatest area for which no data is available, and should be a target for further data collection. Similarly, areas for which predicted background concentrations were developed but no or limited data (n < 30) are available, are targets for further sampling and analysis to confirm concentrations fall within the predicted concentrations or to refine predicted concentrations.

Further analysis of the surficial depositional subgroups (e.g. gravel, sand, mud) is warranted to determine the extent to which erosion of mineralised rocks is contributing to elevated concentrations in soils within those sedimentary subgroups. Mineralised areas have been delineated using data from the Crown Minerals database, the GNS Science PETLAB database, and for Cr and Ni, knowledge of specific lithologies that are elevated in those trace elements. This is not considered to be a definitive map of mineralised zones in New Zealand, but rather identifies locations that may have greater concentrations than those predicted for the given rock subgroup in that location and identifies the extent to which erosion of those mineralised areas may contributing to elevated concentrations in soils within downstream sedimentary subgroups. Topography and erosional processes for the region in question need to be assessed along with additional sampling to confirm the likely extent of elevated concentrations.

Additional sampling should also be undertaken in locations that fall within the mineralised zones to more adequately confirm the likely background concentrations in these regions. A minimum of 30 samples is recommended to characterise background concentrations of a given pedo-geological area (ISO 2012), although this depends on the area under consideration. It may be appropriate to analyse 30 samples for a regionally based characterisation, and a smaller number for a specific site investigation. In this instance, 7 to 10 samples may be appropriate, as used by some US EPA jurisdictions (Diamond et al. 2009). In this situation the intent is to identify whether the site in question contains naturally elevated concentrations of a trace element, rather than determining the background concentration per se.

7.2 Ambient concentrations – PAH and DDT residues

Collation and analysis of existing PAH data from three studies showed BaP concentrations were generally higher in urban areas compared to provincial towns or rural areas. Samples were typically collected from recreational parks, which are likely to have lower concentrations compared to samples collected closer to roads. Two provincial towns showed remarkably high BaP concentrations. Provisional estimates of the upper limit of ambient concentrations of BaP and BaP-eq were developed for urban (city) areas, provincial towns and rural areas. However, further sampling and analysis is required to develop more robust estimates of background concentrations for PAHs.

Existing data on DDT residue concentration in soil was collated; however, data on individual sites was typically lacking and no substantive analysis could be undertaken. DDT residues may be present on land that is not known to have had DDT applied, although the concentrations are typically low, and often below the detection limit typically used (0.005 mg/kg). DDT residues may also be found in the urban environment – sometimes at concentrations higher than agricultural land, although there is no recent data. While some general comment can be made on the apparent difference in concentrations related to land use, whether action is ultimately required depends on whether any effect might occur. In this context, ecological considerations are the most relevant, including potential bioaccumulation in the food chain.

7.3 Database development

We have drawn on our experience in compiling current trace element data from regional council sources for this project to highlight considerations of what and how trace element (or other contaminant) data could be captured. This is intended to make subsequent data input easier, and to identify some data outputs. Consideration is also given to the use and collation of data within contaminated land investigation reports, to assist in contaminated land management. There are various ongoing conversations about the capture of soils data in databases, including within the *Environmental Monitoring and Reporting* project, as well as in the context of a National Soils database. These conversations are ‘bigger’ than the current project, thus the purpose of capturing information within this project is to provide an input to these larger, and ongoing conversations.

8 Recommendations

Further work is required to provide the data contained in this report in a manner that enables identification of predicted background concentrations of a specific location. The detailed information could be made accessible via the internet to enable the background concentration of the location of interest to be specifically determined. This could be achieved through a downloadable dataset or an interactive online map. Such a system could be delivered through the LRIS portal (https://lris.scinfo.org.nz/) or the Our Environment website (http://ourenvironment.scinfo.org.nz/home).

Additional sampling and analysis is required to refine the predicted background concentrations for trace elements:

- Further analysis of the sedimentary subgroups (e.g. gravel, sand, mud) is warranted to determine the extent to which erosion of mineralised rocks is contributing to elevated concentrations in soils within those sedimentary subgroups.
- Further data should be collected from areas underlain by granite and diorite as these groups comprise with the largest area for which no data is available.
- Areas for which predicted background concentrations were developed but no or limited data (n < 30) were available should be further sampled and analysed to confirm concentrations fall within the predicted concentrations or to refine predicted concentrations.
- Additional sampling and analysis should also be undertaken in locations that fall within the mineralised zones to more adequately confirm the likely background concentrations in these areas.

Additional systematic sampling and analysis is required to more robustly determine ambient concentrations of PAHs in urban areas. It is considered that the concentrations of DDT residues are highly dependent on historical usage and remain variable, thus ecological soil guideline values are expected to provide a more useful point of comparison to determine whether any action should be undertaken.

9 References

Ander EL, Cave MR, Johnson CC, Palumbo-Roe B 2011. Normal background concentrations of contaminants in the soils of England. Available data and data exploration. British Geological Survey Commissioned Report, CR/11/145. 124 p.

ARC (Auckland Regional Council) 2001. Background concentrations of inorganic elements in soils from the Auckland Region. Technical Publication No. 153. Auckland, Auckland Regional Council.

ARC (Auckland Regional Council) 2002. Pesticide Residues in Horticultural soils in the Auckland Region. Working report no 96. Auckland, Auckland Regional Council.

BGS (British Geological Survey) 2011. London Earth: Cd in surface soils. G-Base geochmemical map. Nottingham, UK, Keyworth.

Bain LJ Engelhardt M 1992. Introduction to probability and mathematical statistics. Pacific Grove, CA, Duxbury.

BMU 1999. Federal Soil Protection and Contaminated Land Ordinance. Germany Federal Ministry for the Environment (Bundesministerium für Umwelt, BMU).

Boul HL 1995. DDT residues in the environment—A review with a New Zealand perspective. New Zealand Journal of Agriculture 38, 257–277.

Buckland S, Ellis H, Salter R 1998. Ambient concentrations of selected organochlorines in soils. Ministry for the Environment, Wellington NZ.

Burnham KP, Anderson DR 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd edn. New York, Springer-Verlag.

Cavanagh JE, 2013. Determining background soil concentrations of contaminants for managing land. Envirolink Advice Grant 1251-MLDC83 for Marlborough District Council. Lincoln, Manaaki Whenua, Landcare Research.

Cavanagh J 2014. Status of cadmium in New Zealand soils: 2014. Report for the Fertiliser Association of New Zealand and Ministry for Primary Industries. Lincoln, Manaaki Whenua, Landcare Research.

Cavanagh J 2015a. Workshop discussion document. Envirolink Advice Grant: 1555-TSDC110.

Cavanagh J 2015b. Background concentrations of trace elements and options for managing soil quality in the Tasman and Nelson Districts. Envirolink Advice Grant: 1555-TSDC110.

Cave MR, Johnson CC, Ander EL, Palumbo-Roe B 2012. Methodology for the determination of normal background contaminant concentrations in English soils. Nottingham, UK, British Geological Survey (CR/12/003). 41 p.

Cicchella D, De Vivo B, Lima A 2005. Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial areas of Napoli (Italy). Geochemistry: Exploration, Environment, Analysis 5: 29–40.

Craw D, Falconer D, Youngson JH 2003. Environmental arsenopyrite stability and dissolution theory, experiment, and field observations. Chemical Geology 199:71–82.

Crown Minerals 2009. Crown Minerals Open File Metallic Minerals Geochemical Database. Mineral Report Series MR4510. Wellington, Ministry for Economic Development.

Crommentuijn T, Polder MD, Van de Plassche EJ 1997. Maximum permissible concentrations and negligible concentrations for metals, taking background concentrations into account. Report no. 601501 001. Bilthoven, National Institute of Public Health and the Environment (RIVM).

DEFRA (Department for Environment, Food and Rural Affairs) 2012. Environmental Protection Act 1990: Part 2A Contaminated Land Statutory Guidance. April 2012. UK, HM Government. Available online at: http://www.defra.gov.uk/environment/quality/land/ (accessed March 2013).

Diamond D, Baskin D, Brown D, Lund L, Najita J, Javandel I 2009. Analysis of background distributions of metals in the soil at Lawrence Berkeley National Laboratory. Report prepared for U. S. Department of Energy under Contract DE-AC02005CH11231.

Diez M, Simón M, Martin F, Dorronsoro C, Garcia I, Van Gestel CAM 2009. Ambient trace element background concentrations in soils and their use in risk assessment. Science of the Total Environment 407: 4622–4632.

EC (European Commission) 2008. European Union Risk Assessment Report Zinc metal. Brussels, European Commission Joint Research Centre.

Ferry J, Windross L, Hillman C 2014. Superphosphate application to pastoral land – is it really a HAIL activity? Paper presented at WasteMINZ Conference 21-23 October 2014, Wellington.

FMfE (Finnish Ministry for the Environment) 2007. Government decree on the assessment of soil contamination and remediation needs. 1 March 2007. Decree 214/2007. Ministry for the Environment. [Unofficial translation].

Gaw 2003. Report for Auckland…..

Gaw SK, Wilkins AL, Kim ND, Palmer GT and Robinson P. 2006. Trace element and Sigma DDT concentrations in horticultural soils from the Tasman, Waikato and Auckland regions of New Zealand. Science of the Total Environment 355: 31-47.

Heron DW (custodian) 2014. Geological map of New Zealand 1:250,000. GNS Science geological map 1. Lower Hutt, New Zealand, GNS Science.

Hill RB, Sparling GP 2009. Soil quality monitoring. In: Land Monitoring Forum. Land and soil monitoring: a guide for state owned enterprise and regional council reporting. Hamilton, Land Monitoring Forum. Pp. 27–88.

ISO (International Standards Organisation) 2011. Soil quality – Guidance on the determination of background values (ISO 19258:2011-09)..

Jarva J, Tarvainen T, Reinikainen J, Eklund M 2010. TAPIR — Finnish national geochemical baseline database. Science of the Total Environment 408: 4385–4395.

Johnson CC, Anders EL, Cave MR, Palumbo-roe B 2012. Normal background concentrations (NBCs) of contaminants in English soils: Final project report. Nottingham, UK, British Geological Survey (CR/12/035). 40 p.

Johnson CC, Demetriades A 2011. Urban geochemical mapping studies. In: Johnson C, Demetriades A, Locutrura J, Ottesan RT eds. Mapping the chemical environment of urban areas. John Wiley.

Koenker R 2005. Quantile regression. Cambridge, Cambridge University Press.

Lado R, Hengl L, Reuter HI 2008. Heavy metals in European soils: a geostatistical analysis of the FOREGS Geochemical database. Geoderma 148: 189–199.

Lawrence-Smith EJ, Tregurtha CS 2013. Regional environmental monitoring programme for soil quality 2012–13: Arable & pastoral project – Final report. Plant & Food Milestone No. 51731. Plant & Food Research Contract No. 29222. SPTS No. 9007.

Lawley 2009. The Soil-Parent Material Database (SPM-v4): A User Guide. Nottingham, UK, British Geological Survey (OR/08/034). 45pp.

Lilburne LR, Webb TH, Hewitt, AE, Lynn IH, de Pauw B 2012. S-map database manual. Lincoln, Landcare Research.

MAF (Ministry of Agriculture and Forestry) 2011. Cadmium and New Zealand agriculture and horticulture: a strategy for long term risk management. Wellington, Ministry of Agriculture and Forestry.

MfE (Ministry for the Environment) 1997. Health and environmental guidelines for selected timber treatment chemicals. Wellington, Ministry for the Environment.

MfE 2011a. Methodology for deriving standards for contaminants in soil to protect human health. Wellington, Ministry for the Environment.

MfE 2011b. Contaminated Land Management Guidelines No. 4: Classification and Information Management Protocols. Wellington, Ministry for the Environment.

MfE 2011c. Toxicological intake value for priority contaminants. Wellington, Ministry for the Environment.

MfE 2011d. Hazardous Activities and Industry List. Wellington, Ministry for the Environment.

Martin AP, Turnbull RE, Rattenbury MS, Baisden WT, Christie AB, Cohen DR, Hoogewerff J, Rogers KM 2015. Geochemical atlas of southern New Zealand. GNS Science Report 2015/26. Lower Hutt, GNS Science,

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

McDowell RW, Taylor MD, Stevenson BA 2013. Natural background and anthropogenic contributions of cadmium to New Zealand soils. Agriculture, Ecosystems and Environment 165: 80–87.

Mortimer N, Rattenbury MS, King PR, Bland KJ, Barrell DJA, Bache F, Begg JG, Campbell HJ, Cox SC, Crampton JS, Edbrooke SW, Forsyth PJ, Johnston MR, Jongens R, Lee J, Leonard GS, Raine JI, Skinner DNB, Timm C, Townsend DB, Tulloch AJ, Turnbull IM, Turnbull RE 2014. High-level stratigraphic scheme for New Zealand rocks. New Zealand Journal of Geology and Geophysics 57: 402–419.

NREPC (Natural Resources and Environmental Protection Cabinet) 2004. Kentucky guidance for ambient background assessment. Kentucky, USA, Natural Resources and Environmental Protection Cabinet.

NZWWA (New Zealand Water and Wastes Association) 2003. Guidelines for the safe application of biosolids to land in New Zealand. Wellington, New Zealand Water and Wastes Association.

Oehlert, GW 1992. A note on the delta method. The American Statistician 46(1): 27–29.

PDP (Pattle Delamore Partners) 2007. Contamination of horticultural land in Canterbury – A scoping study.Christchurch, Pattle Delamore Partners Ltd.

Pinehero and Bates 200

R Core Team (2015). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing. http://www.R-project.org/.

Rattenbury MS, Isaac MJ 2012. The QMAP 1:250 000 Geological Map of New Zealand project. New Zealand Journal of Geology and Geophysics 55: 393–405.

Rattenbury MS, Martin AP, Turnbull, RE, Christie AB 2014. Sampling methodology for a regional multi-element geochemical baseline survey. GNS Science Report 2014/62.

Reimann C 2005. Geochemical mapping: technique or art? Geochemistry: Exploration, Environment, Analysis 5: 359–370.

Reimann C, Garrett RG. 2005. Geochemical background — concept and reality. Science of the Total Environment 350: 12–27.

Roser BP, Grapes RH 1990. Geochernistry of a metabasalt-chert - coloured-argillite - turbidite association at Red Rocks, Wellington. New Zealand Journal of Geology and Geophysics 33: 181-192.

SEM (Solutions in Environmental Management) 2005. Background Levels of Agrichemical Residues in Bay Of Plenty Soils. Report prepared for Bay of Plenty Regional Council.

SCEW (Standing Council on Environment and Water) 2013. National environment protection (assessment of site contamination) Measure Schedule B5b. Methodology to derive ecological investigation levels in contaminated soils. Former COAG (Council of Australian Governments) Standing Council on Environment and Water. http://www.scew.gov.au/archive/site-contamination/asc-nepm.html (accessed January 2013).

Sheppard SC, Grant CA, Sheppard MI, de Jong R, Long J 2009. Risk indicator for agricultural inputs of trace elements to Canadian soils. Journal of Environmental Quality 38: 919–932.

Tonkin and Taylor 2006. Background concentrations of selected trace elements in Canterbury soils. Environment Canterbury Report No. R07/1.

Tonkin and Taylor Ltd 2007a. Background concentrations of selected trace elements in Canterbury soils Addendum 1: Additional samples and Timaru specific background levels. Report R07/1/2. Report for Environment Canterbury.

Tonkin and Taylor Ltd 2007b. Background concentration of PAH compounds in Christchurch urban soils. Report R07/19. Report for Environment Canterbury.

Turnbull RE, Martin AP, Rattenbury MS, Christie AB 2015. Regional soil geochemical baseline data for mineral exploration, southern New Zealand. p. 477-486 In: Sterk, R. (convener) Australasian Institute of Mining and Metallurgy 48th New Zealand Branch Annual Conference on exploration, mining and New Zealand's mineral resources, Australasian Institute of Mining and Metallurgy, New Zealand Branch.

URS 2003. Determination of common pollutant background soil concentrations for the Wellington Region. Report prepared for Greater Wellington Regional Council.

US EPA (United States Environmental Protection Agency) 2002. Guidance for comparing background and chemical concentrations in soil for CERCLA sites. EPA 540-R-01-003, OSWER 9285.7-41. US, United States Environmental Protection Agency.

US EPA (United States Environmental Protection Agency) 2013. ProUCL Version 5.0.00 User Guide. Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations. EPA/600/R-07/041. US, United States Environmental Protection Agency.

Appendix 1 – Pedo-geological parameters

Rock-type of fines – S-Map

This attribute-field describes the rock class of underlying rock (if rock occurs within 100 cm) or of stones occurring within the soil profile within a depth of 0–100 cm. Twenty-two rock classes have been identified to record rock type (Table 23). Further details of the classification are provided in Lilburne et al. 2012.

Table 24 Grouping of rock-type of fines for the family-level classification for soils

Rock type	Examples	Rock-class groups	Rock-class-group code
Cb	Carbonaceous rock (coal)	Carbonaceous	Carb
Sq	Soft quartzitic sediments	Soft quartzitic sedimentary rocks	SQrtz
Sm	Mudstone	Soft sedimentary sandstones and mudstones	SSed
Ss	Sandstone		
Hq	Quartzite	Hard quartzitic sedimentary rocks	HQrtz
Hm	Mudstone	Hard sedimentary sandstones and mudstones (greywacke)	HSed
Hs	Sandstone		
Tm	Tuffaceous mudstone	Hard tuffaceous sedimentary rocks (tuffaceous greywacke)	Tuf
Ts	Tuffaceous sandstone		
Sc	Schist	Schist	Sch
Ls	Soft calcareous	Soft calcareous rocks	SCalc
Li	Limestone	Hard calcareous rocks	HCalc
Ma	Marble		
Rh	Rhyolite	Siliceous igneous rocks	Sil-Ig
Gr	Granite and Gneiss		
Ig	Ignimbrite		
An	Andesite	Intermediate igneous rocks	Int-Ig
Di	Diorite		
Ba	Basalt	Basic igneous rocks	Bas-Ig
Ga	Gabbro		
Fe	Iron sands[a]		
Ul	Ultrabasic	Ultrabasic rocks	Ubas

[a] a sand containing particles of iron ore (titanomagnetite)

Fundamental Soils Layer – Top rock

‘Toprock’ is a parameter that describes the near surface lithology (rock type) and was used to investigate variation in trace element composition in different soils. Table 24 shows the grouping of individual Toprock classifications provided in the FSL into: ultramafics, igneous, surficial, weak sedimentary rock, strong sedimentary rock, and metamorphic rocks.

Table 25 Grouping of the individual Toprock classifications used in the Fundamental Soils Layers used in this report

Igneous		Ultramafics		Surficial soil types		Sedimentary rocks-weak		Sed rocks strong		Metamorphic	
Tb	Pyroclastics (ash & lapilli)	Um	Ultramafics	Al	Alluvium, colluvium, glacial drift	Ms	Mudstone	Ar	Argillite	St1	Semi-schist
In	Ancient volcanoes, minor intrusives (dikes & sills)			Lo	Loess	Fy	Interbedded sandstone & mudstone	Gw	Greywacke	Gs	Gneiss
Vo	Lavas			Wb	Windblown sand	Ss	Sandstone	Ls	Limestone	St2	Schist
Gn	Plutonics			Pt	Peat	Cw	Conglomerate	Hs	Sandstone	Ma	Marble
Ng	Ngauruhoe ash			Gr	Gravels	Us	Unconsolidated to moderately consolidated clays, silts, sands, tephra & breccias	Cg	Conglomerate		
Ta	Tarawera ash and lapilli			Cl	Coarse slope deposits	Mm	Mudstone or fine siltstone — massive	Ac	Argillite — crushed		
Lp	Pumiceous lapilli			Gl	Glacial till	Mj	Mudstone or fine siltstone — jointed	Gw	Greywacke		
Tp	Taupo & Kaharoa breccia & volcanic alluvium			Uf	Unconsolidated clays and silts	Sm	Sandstone or coarse siltstone — massive	Ar	Argillite		
Ft	Breccias older than Taupo breccia			Us	Unconsolidated sands and gravel	Cw	Weakly consolidated conglomerate	Li	Limestone		
Vu 1	‘Soft’ volcanic rocks					Mb	Mudstone or fine	Cg	Conglomerate &		

Igneous		Ultramafics	Surficial soil types	Sedimentary rocks-weak		Sed rocks strong		Metamorphic
					siltstone — banded		breccia	
Gn	Crystalline intrusive rocks, Plutonics			Me	Mudstone — bentonitic	Si	Indurated sandstone	
Rm	Rotomahana mud			Sb	Sandstone or coarse siltstone — banded			
Sc 1, Sc	Scoria			Mx	Sheared mixed lithologies			
Kt	Kaharoa & Taupo ashes			Mf	Frittered mudstone			
Mo	Ashes older than Taupo pumice			Me	Bentonitic mudstone			
La	Lahar deposits			Sm	Massive sandstone			
Vo	Lavas, ignimbrite & other ‘hard’ volcanic rocks			Mx	Sheared mixed lithologies			
Mo	Ashes older than Taupo ash			Ac	Crushed argillite association of rocks			
Ft	Quaternary breccias older than Taupo breccia							
La	Lahar deposits							
Vu	Extremely weak altered volcanic							
Tb	Indurated fine grained pyroclastics							
Vb	Indurated volcanic breccias							

QMAP

The QMAP seamless GIS geological units polygon layer has many attribute fields that describe the composition, age and stratigraphic affiliation of each unit. For this study the layer's lithological fields *Rock Group* and *Rock Class* have been analysed, along with five derivative fields have been constructed based on the high level stratigraphic affiliation, age and lithological composition of each stratigraphic unit.

Chemical1 has the most detailed stratigraphic affiliations (43 categories) based on Mortimer et al. (2014). Some group-level units are distinguished, for example, the Maitai Group, Dun Mountain Ultramafics Group and Livingstone Volcanics Group of the Dun Mountain-Maitai Terrane. Also included are plutonic suites defined on geochemical and age criteria and differentiating schist components in the Torlesse Composite Terrane. *Chemical2* is more generalised and has 24 primarily stratigraphic categories based on basement terranes, volcanic supersuites and sedimentary components of the post Late Cretaceous cover supergroups based on Mortimer et al. (2014). *Chemical3* is even more generalised, with additional grouping on a geographic and tectonostratigraphic basis (9 categories). *Chemical5* is based on *Chemical1* but subdivides the Miocene and younger sedimentary rocks and sediments (Maui and Pakihi supergroups; Mortimer et al. (2014)) on a five-part regional basis (Northland, Auckland (including Coromandel Peninsula), central North Island, East Coast (both North and South islands) and West Coast (South Island including southern New Zealand), forming 51 categories. These broadly based regions have contributed differing igneous rock types into the younger sedimentary Maui and Pakihi supergroups.

Chemical4 is a variant on *Rock-Group* but subdivides the younger sedimentary rocks and sediments (Maui and Pakihi 'supergroups') from older 'clastic sediments' on a high-level stratigraphic basis (Maui Supergroup, Pakihi Supergroup, other) forming 68 categories. The underlying rationale for this subdivision is that New Zealand's youngest rocks and sediments have had significant influx of volcanic-derived material, particularly in the North Island, with their distinctive geochemical signatures from plate boundary volcanism. *Chemical4b* is a variant on the existing 19-category Rock Class where the broad 'clastic sediment' field has been subdivided on a three-fold high level stratigraphy-age basis (i.e. subdivides the younger sedimentary rocks and sediments (Maui and Pakihi 'supergroups') from older 'clastic sediments') to create 14 categories.

A full listing of all the subgroups within the different parameters is shown in Tables 25 and 26.

Table 26 Listing of the individual subgroups within Chemcal1-3

Chemical1	Chemical2	Chemical3
Brook Street Terrane	Brook Street Terrane	Austral Superprovince central terrane
Buller Terrane	Buller Terrane	Austral Superprovince eastern terranes
Caples Terrane	Caples Terrane	Austral Superprovince intrusive rocks
Darran Suite	Drumduan Terrane	Austral Superprovince western terranes
Drumduan Terrane	Dun Mountain-Maitai Terrane	Ruaumoko Volcanic Province Horomaka Supersuite
Dun Mountain-Maitai Terrane	Eastern Province undifferentiated	Ruaumoko Volcanic Province Te Raupua Supersuite
Dun Mountain Ultramafics Group	Haerenga Supergroup	Ruaumoko Volcanic Province Whakaari Supersuite
Eastern Province undifferentiated	Horomaka Supersuite	Zealandia Megasequence older cover
Ferrar Suite	Maui Supergroup	Zealandia Megasequence younger cover
Foulwind Suite	Momotu Supergroup	
Haerenga Supergroup	Murihiku Terrane	
Horomaka Supersuite	Pakihi Supergroup	
Jaquiery Suite	Takaka terrane	
Karamea Suite	Takaka Terrane	
Livingstone Volcanics Group	Te Raupua Supersuite	
Longwood Suite	Torlesse Composite Terrane	
Maitai Group	Tuhua Intrusives	
Maui Supergroup	Waipapa Composite Terrane	
Median Batholith undifferentiated	Waka Supergroup	
Mistake Suite	Western Province undifferentiated	
Momotu Supergroup	Whakaari Supersuite	
Murihiku Terrane	Zealandia Megasequence undifferentiated older cover	
Otanomomo Complex	Zealandia Megasequence undifferentiated younger cover	
Pakihi Supergroup		
Paringa Suite		
Rahu Suite		
Ridge Suite		

Chemical1	Chemical2	Chemical3
Separation Point Suite		
Takaka Terrane		
Tarpaulin Suite		
Te Raupua Supersuite		
Tobin Suite		
Torlesse Composite Terrane		
Torlesse Composite Terrane schist		
Torlesse Composite Terrane semischist		
Tuhua Intrusives undifferentiated		
Waipapa Composite Terrane		
Waka Supergroup		
Western Province undifferentiated		
Whakaari Supersuite		
Willsher Group		
Zealandia Megasequence undifferentiated older cover		
Zealandia Megasequence undifferentiated younger cover		

Table 27 Listing of the individual subgroups within Chemcial4-5, Rock Group and Rock Class

Rock Group	Chemical4	Rock Class	Chemical4b	Chemical5
Agglomerate	Agglomerate	Chemical sediment	Chemical sediment	Brook Street Terrane
Alternating sandstone/mudstone	Alternating sandstone/mudstone	Clastic sediment	Clastic sediment	Buller Terrane
Alternating sandstone/ siltstone	Alternating sandstone/mudstone maui	Felsic extrusive	Clastic sediment maui	Caples Terrane
Amphibolite	Alternating sandstone/siltstone maui	Felsic intrusive	Clastic sediment pakihi	Darran Suite
Andesite	Amphibolite	Gneiss	Felsic extrusive	Drumduan Terrane
Argillite	Andesite	Intermediate extrusive	Felsic intrusive	Dun Mountain Ultramafics Group
Basalt	Argillite	Intermediate intrusive	Gneiss	Eastern Province undifferentiated
Breccia	Basalt	Mafic extrusive	Intermediate extrusive	Ferrar Suite

Rock Group	Chemical4	Rock Class	Chemical4b	Chemical5
Cataclasite	Breccia	Mafic intrusive	Intermediate intrusive	Foulwind Suite
Chert	Cataclasite	Schist	Mafic extrusive	Haerenga Supergroup
Conglomerate	Chert	Tectonic	Mafic intrusive	Horomaka Supersuite
Dacite	Conglomerate	Ultramafic intrusive	Schist	Jaquiery Suite
Diorite	Conglomerate maui		Tectonic	Karamea Suite
Dolerite	Dacite		Ultramafic intrusive	Livingstone Volcanics Group
Fill	Diorite			Longwood Suite
Gabbro	Dolerite			Maitai Group
Gabbronorite	Fill			Maui Supergroup Auckland
Gneiss	Gabbro			Maui Supergroup central North Island
Granite	Gabbronorite			Maui Supergroup East Coast
Granodiorite	Gneiss			Maui Supergroup Northland
Gravel	Granite			Maui Supergroup West Coast
Greenschist	Granodiorite			Median Batholith undifferentiated
Greywacke	Gravel			Mistake Suite
Hornblendite	Greenschist			Momotu Supergroup
Hornfels	Greywacke			Murihiku Terrane
Hypabyssal intrusive	Hornblendite			Otanomomo Complex
Ignimbrite	Hornfels			Pakihi Supergroup Auckland
Lignite	Hypabyssal intrusive			Pakihi Supergroup central North Island
Limestone	Ignimbrite			Pakihi Supergroup East Coast
Loess	Lignite			Pakihi Supergroup Northland
Marble	Limestone			Pakihi Supergroup West Coast
Melange	Loess			Paringa Suite
Metaplutonic	Marble			Rahu Suite
Metasediment	Melange			Ridge Suite
Metavolcanic	Metaplutonic			Separation Point Suite
Metavolcanics	Metasediment			Takaka terrane
Migmatite	Metavolcanic			Takaka Terrane
Monzodiorite	Metavolcanics			Tarpaulin Suite

Rock Group	Chemical4	Rock Class	Chemical4b	Chemical5
Monzonite	Migmatite			Te Raupua Supersuite
Mud	Monzodiorite			Tobin Suite
Mudstone	Monzonite			Torlesse Composite Terrane
Mylonite	Mud			Torlesse Composite Terrane schist
Orthogneiss	Mudstone			Torlesse Composite Terrane semischist
Paragneiss	Mudstone pakihi			Tuhua Intrusives undifferentiated
Peat	Mylonite			Waipapa Composite Terrane
Peridotite	Orthogneiss			Waka Supergroup
Pyroclastics	Paragneiss			water
Pyroxenite	Peat			Western Province undifferentiated
Quartzite	Peridotite			Whakaari Supersuite
Rhyolite	Pyroclastics			Willsher Group
Sand	Pyroxenite			Zealandia Megasequence undifferentiated older cov*
Sandstone	Quartzite			Zealandia Megasequence undifferentiated younger c*
Schist	Rhyolite			
Scoria	Sand			
Semischist	Sandstone			
Serpentinite	Sandstone Pakihi			
Shale	Schist			
Silt	Scoria			
Siltstone	Semischist			
Syenite	Serpentinite			
Till	Shale			
Tonalite	Silt			
Tuff	Siltstone			
Unknown	Siltstone Pakihi			
Volcanics	Syenite			
	Till			
	Tonalite			
	Tuff			
	Unknown			
	Volcanics			

Appendix 2 – Statistical analysis results

Model comparison for regional council data

The following tables provide the statistical results of comparing the fit of different linear regression models for the background land-use class based on mean (Tables 27–33) and 95th percentile concentrations (Tables 34–40).

For arsenic, the model involving Soil Order is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 28 Statistical results of comparison of the mean concentration models for arsenic

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.6639	2633.8735	16%
Rock Class Fines	0.6648	Not applicable	7.6%
Top Rock	0.7169	Not applicable	3%
Rock Class	0.7199	2829.3699	1.7%
Rock Group	0.6978	2783.3909	6.4%
Chemical1	0.6868	2726.6876	9.9%
Chemical2	0.6872	2724.0567	10%
Chemical3	0.7098	2791.2816	4.5%
Chemical4	0.6903	2761.8087	8.2%
Chemical5	0.6773	2705.1899	11.9%
Chemical4b	0.7077	2785.7221	5%

For copper, the model involving Soil Order is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 29 Statistical results of comparison of the mean concentration models for copper

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.6953	3159.785	28.9%
Rock Class Fines	0.6526	Not applicable	17.2%
Top Rock	0.7529	Not applicable	17.7%
Rock Class	0.7372	3320.3014	20.4%
Rock Group	0.7073	3234.2169	25.8%
Chemical1	0.7277	3302.1201	21.9%
Chemical2	0.7296	3305.6528	21.6%
Chemical3	0.7384	3323.1475	20.1%
Chemical4	0.7034	3224.0357	26.4%
Chemical5	0.708	3234.9891	25.7%
Chemical4b	0.7351	3311.8976	20.8%

For cadmium, the model involving Soil Order is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 30 Statistical results of comparison of the mean concentration models for cadmium

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.622	3578.9479	54.4%
Rock Class Fines	0.5603	Not applicable	50.9%
Top Rock	0.6411	Not applicable	51.2%
Rock Class	0.683	3916.6839	45.2%
Rock Group	0.6293	3650.5381	53%
Chemical1	0.6726	3877.5415	46.6%
Chemical2	0.6738	3880.4707	46.5%
Chemical3	0.6785	3890.0145	46%
Chemical4	0.6262	3638.1796	53.4%
Chemical5	0.6255	3620.0939	53.6%
Chemical4b	0.6859	3932.716	44.7%

For chromium, the model involving Chemical4 is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 31 Statistical results of comparison of the mean concentration models for chromium

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.7546	3400.8175	30.7%
Rock Class Fines	0.6291	Not applicable	43.7%
Top Rock	0.8039	Not applicable	22.6%
Rock Class	0.7618	3417.0962	29.6%
Rock Group	0.7017	3210.7612	39.5%
Chemical1	0.7562	3415.2592	30.2%
Chemical2	0.7578	3417.5035	30%
Chemical3	0.7684	3440.236	28.4%
Chemical4	0.69	3167.3034	41.4%
Chemical5	0.6915	3165.4647	41.3%
Chemical4b	0.7576	3400.5814	30.4%

For nickel, the model involving Rock Group is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 32 Statistical results of comparison of the mean concentration models for nickel

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.8812	3855.6331	29.5%
Rock Class Fines	0.6713	Not applicable	43%
Top Rock	0.8726	Not applicable	23.3%
Rock Class	0.9175	3962.5761	23.9%
Rock Group	0.7881	3550.5363	43.1%
Chemical1	0.9035	3937.3743	25.7%
Chemical2	0.9049	3937.8288	25.5%
Chemical3	0.9213	3972.5539	23.3%
Chemical4	0.7817	3532.7756	43.9%
Chemical5	0.8137	3642.7166	39.4%
Chemical4b	0.9028	3914.8568	26.3%

For lead, the model involving Soil Order is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 33 Statistical results of comparison of the mean concentration models for lead

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.6437	2951.9209	26.7%
Rock Class Fines	0.626	Not applicable	24.5%
Top Rock	0.6905	Not applicable	10.3%
Rock Class	0.6735	3074.161	20.1%
Rock Group	0.6501	3005.0947	24.6%
Chemical1	0.662	3043.2532	22.3%
Chemical2	0.6627	3042.0787	22.2%
Chemical3	0.6736	3072.5654	20.1%
Chemical4	0.6479	3001.2206	25%
Chemical5	0.6521	3012.5967	24.2%
Chemical4b	0.6664	3042.9756	21.8%

For zinc, the model involving Chemical5 is favoured using the AIC criterion, while the RockClassFines model is favoured using RMS error.

Table 34 Statistical results of comparison of the mean concentration models for zinc

Parent rock factor	RMS error	AIC	R-squared
SoilOrder	0.6228	2901.4102	24.5%
RockClassFines	0.5254	Not applicable	24.3%
TopRock	0.6359	Not applicable	11.7%
RockClass	0.6596	3062.6592	15.7%
RockGroup	0.6103	2864.2061	26.9%
Chemical1	0.6375	2979.7008	20.7%
Chemical2	0.638	2978.2321	20.7%
Chemical3	0.6548	3038.3893	17%
Chemical4	0.6098	2867.8236	26.9%
Chemical5	0.6067	2844.366	27.8%
Chemical4b	0.6453	2996.1892	19.3%

For arsenic, the 95th percentile model involving Chemical5 is favoured using the AIC criterion.

Table 35 Statistical results of comparison of the 95th percentile concentration models for arsenic

Parent rock factor	AIC
SoilOrder	3549.8766
RockClassFines	Not applicable
TopRock	Not applicable
RockClass	3636.3594
RockGroup	3468.2612
Chemical1	3547.8791
Chemical2	3548.637
Chemical3	3611.7567
Chemical4	3443.956
Chemical5	3412.8528
Chemical4b	3633.969

For copper, the model involving SoilOrder is favoured using the AIC criterion.

Table 36 Statistical results of comparison of the 95th percentile concentration models for copper

Parent rock factor	AIC
SoilOrder	3982.2016
RockClCusFines	Not applicable
TopRock	Not applicable
RockClCus	4203.7586
RockGroup	4035.692
Chemical1	4194.0086
Chemical2	4192.6674
Chemical3	4250.2034
Chemical4	4007.3542
Chemical5	4029.4595
Chemical4b	4193.512

For cadmium, the model involving Chemical4 is favoured using the AIC criterion.

Table 37 Statistical results of comparison of the 95th percentile concentration models for cadmium

Parent rock factor	AIC
SoilOrder	4625.991
RockClassFines	Not applicable
TopRock	Not applicable
RockClass	5184.2423
RockGroup	4575.1493
Chemical1	5137.9023
Chemical2	5147.503
Chemical3	5205.3399
Chemical4	4542.943
Chemical5	4551.8389
Chemical4b	5222.4311

For chromium, the model involving Chemical4 is favoured using the AIC criterion, while the RockClassFines model is favoured using RMS error.

Table 38 Statistical results of comparison of the 95th percentile concentration models for chromium

Parent rock factor	AIC
SoilOrder	4569.4747
RockClassFines	Not applicable
TopRock	Not applicable
RockClass	4654.3246
RockGroup	4418.3407
Chemical1	4697.169
Chemical2	4725.3605
Chemical3	4784.0151
Chemical4	4389.9332
Chemical5	4384.0161
Chemical4b	4658.0086

For nickel, the model involving Chemical5 is favoured using the AIC criterion.

Table 39 Statistical results of comparison of the 95th percentile concentration models for nickel

Parent rock factor	AIC
SoilOrder	4976.0225
RockClassFines	Not applicable
TopRock	Not applicable
RockClass	5071.7321
RockGroup	4762.9164
Chemical1	5078.88
Chemical2	5090.391
Chemical3	5115.6934
Chemical4	4762.8263
Chemical5	4683.5432
Chemical4b	5026.1862

For lead, the model involving SoilOrder is favoured using the AIC criterion.

Table 40 Statistical results of comparison of the 95th percentile concentration models for lead

Parent rock factor	AIC
SoilOrder	3957.793
RockClassFines	Not applicable
TopRock	Not applicable
RockClass	4151.0375
RockGroup	4087.7754
Chemical1	4075.4922
Chemical2	4085.0789
Chemical3	4179.6973
Chemical4	4055.4107
Chemical5	4054.6005
Chemical4b	4129.1517

For zinc, the model involving Chemical4 is favoured using the AIC criterion.

Table 41 Statistical results of comparison of the 95th percentile concentration models for zinc

Parent rock factor	AIC
SoilOrder	3472.1288
RockClassFines	Not applicable
TopRock	Not applicable
RockClass	3586.5845
RockGroup	3417.839
Chemical1	3503.331
Chemical2	3506.0207
Chemical3	3568.4727
Chemical4	3380.476
Chemical5	3436.0477
Chemical4b	3535.799

Southland/Otago data

The following tables provide the results of the statistical fit of different linear regression models based on mean (Tables 41–47) and 95th percentile concentrations (Tables 48–54) for the GNS Science Southland Otago dataset based on aggregated land use (Table 6).

For arsenic, the model involving Chemical5 is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 42 Statistical fit of mean concentration models for arsenic

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.6443	711.7207	19.5%
Rock Class Fines	0.4606	Not applicable	22.1%
Top Rock	0.6535	Not applicable	18.2%
Rock Class	0.6379	704.7191	21.1%
Rock Group	0.5646	636.0041	36.7%
Chemical1	0.5688	635.1643	36.3%
Chemical2	0.5776	641.7747	34.7%
Chemical3	0.5957	651.2741	31.8%
Chemical4	0.5643	637.7037	36.5%
Chemical5	0.5478	613.0032	40.6%
Chemical4b	0.6325	694.8788	22.9%

For copper, the model involving Chemical5 is favoured using the AIC criterion, and using RMS error.

Table 43 Statistical fit of mean concentration models for copper

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.6942	763.4	7.4%
Rock Class Fines	0.6407	Not applicable	2.6%
Top Rock	0.6796	Not applicable	12.3%
Rock Class	0.6815	750.5862	10.7%
Rock Group	0.6607	745.0665	14%
Chemical1	0.6529	730.8654	16.8%
Chemical2	0.6541	728.1652	17%
Chemical3	0.6771	740.103	12.6%
Chemical4	0.6592	745.5064	14.1%
Chemical5	0.6401	721.1398	19.5%
Chemical4b	0.6757	740.6888	12.7%

For cadmium, the model involving Chemical2 is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 44 Statistical fit of mean concentration models for cadmium

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.8253	883.4833	19.7%
Rock Class Fines	0.6765	Not applicable	18.3%
Top Rock	0.7539	Not applicable	33.8%
Rock Class	0.7404	808.158	35.4%
Rock Group	0.7229	807.5565	36.9%
Chemical1	0.7292	807.5506	36.4%
Chemical2	0.7304	804.7025	36.5%
Chemical3	0.7493	810.4171	34.4%
Chemical4	0.7225	809.1784	36.7%
Chemical5	0.7225	805.1307	37.1%
Chemical4b	0.7508	813.8395	33.9%

For chromium, the model involving Chemical5 is favoured using the AIC criterion, and using RMS error.

Table 45 Statistical fit of mean concentration models for chromium

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.5284	574.0697	7.5%
Rock Class Fines	0.4258	Not applicable	15.2%
Top Rock	0.4735	Not applicable	26.6%
Rock Class	0.4621	480.9661	29.3%
Rock Group	0.4483	476.0183	31.8%
Chemical1	0.4297	440.5917	37.9%
Chemical2	0.4464	463.0721	33.4%
Chemical3	0.475	494.1065	26%
Chemical4	0.4483	478.0074	31.6%
Chemical5	0.4193	427.5888	40.5%
Chemical4b	0.4616	476.2521	29.9%

For nickel, the model involving Chemical5 is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 46 Statistical fit of mean concentration models for nickel

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.6297	695.7315	13.3%
Rock Class Fines	0.5295	Not applicable	13.7%
Top Rock	0.6021	Not applicable	21.7%
Rock Class	0.5979	659.8386	21.8%
Rock Group	0.568	640.201	27.7%
Chemical1	0.575	642.7116	26.6%
Chemical2	0.5965	664.1542	21.5%
Chemical3	0.6186	677.4202	17.1%
Chemical4	0.568	642.1772	27.5%
Chemical5	0.5658	635.5559	28.5%
Chemical4b	0.5918	648.6263	23.9%

For lead, the model involving Chemical4 is favoured using the AIC criterion, while the Rock Group model is favoured using RMS error.

Table 47 Statistical fit of mean concentration models for lead

Parent rock factor	RMS error	AIC	R-squared
Soil Order	0.4976	532.3788	12.6%
Rock Class Fines	0.5051	Not applicable	26.1%
Top Rock	0.5005	Not applicable	12.6%
Rock Class	0.4739	498.4244	20.7%
Rock Group	0.463	498.2883	22.5%
Chemical1	0.4839	522.9665	16.1%
Chemical2	0.4878	524.5831	15.2%
Chemical3	0.4965	524.8001	13.8%
Chemical4	0.4612	497.6974	22.8%
Chemical5	0.4719	509.6222	19.7%
Chemical4b	0.4931	522.0031	14.7%

For zinc, the model involving Chemical2 is favoured using the AIC criterion, while the Rock Class Fines model is favoured using RMS error.

Table 48 Statistical fit of mean concentration models for zinc

Parent rock factor	RMS error	AIC	R-squared
SoilOrder	0.5605	615.0218	12.7%
Rock Class Fines	0.433	Not applicable	15.6%
Top Rock	0.5372	Not applicable	20.7%
Rock Class	0.5409	590.2019	18.7%
Rock Group	0.5317	594.3127	19.5%
Chemical1	0.5359	593.7945	19%
Chemical2	0.5359	589.8108	19.4%
Chemical3	0.547	592.0376	17.6%
Chemical4	0.5301	594.2448	19.7%
Chemical5	0.5341	595.5	19%
Chemical4b	0.5448	591.213	18%

For arsenic, the 95th percentile model involving Chemical5 is favoured using the AIC criterion.

Table 49 Statistical fit of 95th percentile concentration models for arsenic

Parent rock factor	AIC
Soil Order	760.0498
Rock Class Fines	Not applicable
Top Rock	755.7164
Rock Class	748.1906
Rock Group	728.9314
Chemical1	710.4233
Chemical2	713.7975
Chemical3	722.3097
Chemical4	724.8559
Chemical5	700.9905
Chemical4b	740.2585

For copper, the model involving Chemical5 is favoured using the AIC criterion.

Table 50 Statistical fit of 95th percentile concentration models for copper

Parent rock factor	AIC
Soil Order	1030.6084
Rock Class Fines	Not applicable
Top Rock	995.9044
Rock Class	1008.1949
Rock Group	978.4162
Chemical1	928.9913
Chemical2	928.6587
Chemical3	990.8796
Chemical4	977.829
Chemical5	869.5063
Chemical4b	996.2037

For cadmium, the model involving Chemical5 is favoured using the AIC criterion.

Table 51 Statistical fit of 95th percentile concentration models for cadmium

Parent rock factor	AIC
Soil Order	1039.8001
Rock Class Fines	Not applicable
Top Rock	1025.9382
Rock Class	1018.1336
Rock Group	993.5839
Chemical1	1005.0041
Chemical2	1001.8134
Chemical3	1032.8451
Chemical4	991.8686
Chemical5	986.6333
Chemical4b	1008.455

For chromium, the model involving Chemical1 is favoured using the AIC criterion.

Table 52 Statistical fit of 95th percentile concentration models for chromium

Parent rock factor	AIC
Soil Order	727.2264
Rock Class Fines	Not applicable
Top Rock	740.0958
Rock Class	690.3981
Rock Group	656.9394
Chemical1	636.706
Chemical2	659.0456
Chemical3	757.2595
Chemical4	658.9111
Chemical5	638.3649
Chemical4b	666.8518

For nickel, the model involving Chemical1 is favoured using the AIC criterion.

Table 53 Statistical fit of 95th percentile concentration models for nickel

Parent rock factor	AIC
Soil Order	912.7923
Rock Class Fines	Not applicable
Top Rock	899.2899
Rock Class	882.5807
Rock Group	834.9942
Chemical1	807.2459
Chemical2	829.942
Chemical3	940.4506
Chemical4	834.9547
Chemical5	809.0456
Chemical4b	859.8164

For lead, the model involving Rock Class is favoured using the AIC criterion.

Table 54 Statistical fit of 95th percentile concentration models for lead

Parent rock factor	AIC
Soil Order	903.8579
Rock Class Fines	Not applicable
Top Rock	806.6298
Rock Class	696.3887
Rock Group	746.9752
Chemical1	764.0149
Chemical2	765.0284
Chemical3	777.3952
Chemical4	744.3027
Chemical5	753.3687
Chemical4b	798.6017

For zinc, the model involving Rock Class is favoured using the AIC criterion.

Table 55 Statistical fit of 95th percentile concentration models for zinc

Parent rock factor	AIC
Soil Order	695.3756
Rock Class Fines	Not applicable
Top Rock	680.3573
Rock Class	656.6421
Rock Group	673.3178
Chemical1	672.9889
Chemical2	670.2305
Chemical3	686.4623
Chemical4	669.7863
Chemical5	668.8198
Chemical4b	672.8745

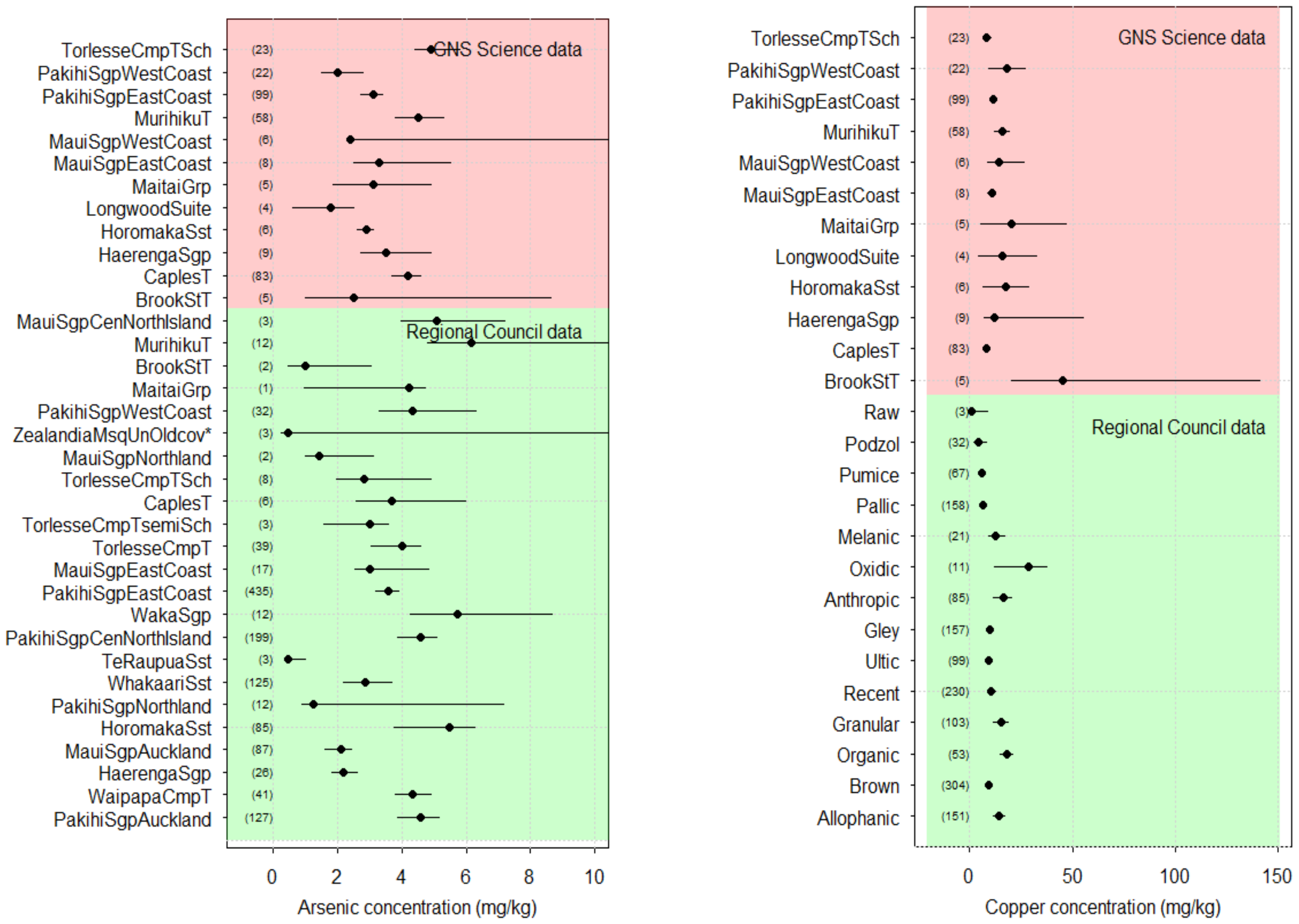

Figure 28 Comparison of predicted median and 95th% confidence interval (black line) background concentrations of arsenic (As) and copper (Cu) (estimated median concentration (black circle)) using the favoured 95th quantile models for each of the regional council data (Chemical5 for As, Soil Order for Cu) and the Southland-Otago (GNS Science) data (Chemical5 for As and Cu. Refer to Appendix 1 for the specific parameters used to develop estimates). The number of samples in each of the levels for the given explanatory variable is shown at the left of the plot area.

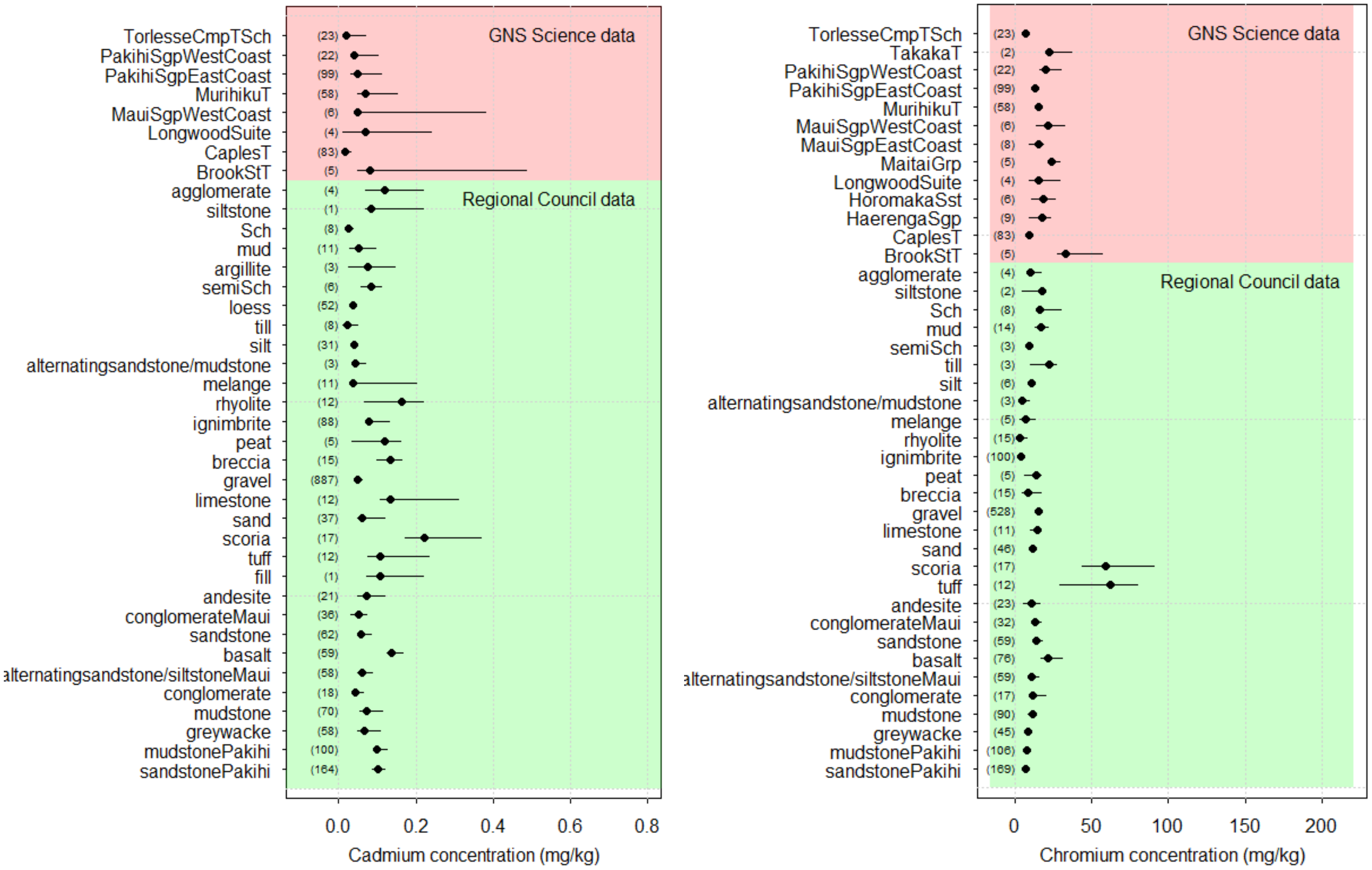

Figure 29 Comparison of predicted background concentrations of cadmium (Cd) and chromium (Cr) (estimated median concentration (black circle) and estimated 95% confidence interval (black line)) using the favoured 95th quantile models for each of the regional council data (Chemical4 for Cd and Cr) and the Southland-Otago (GNS Science) data (Chemical5 for Cd and Chemical1 for Cr). The number of samples in each of the levels for the given explanatory variable is shown at the left of the plot area.

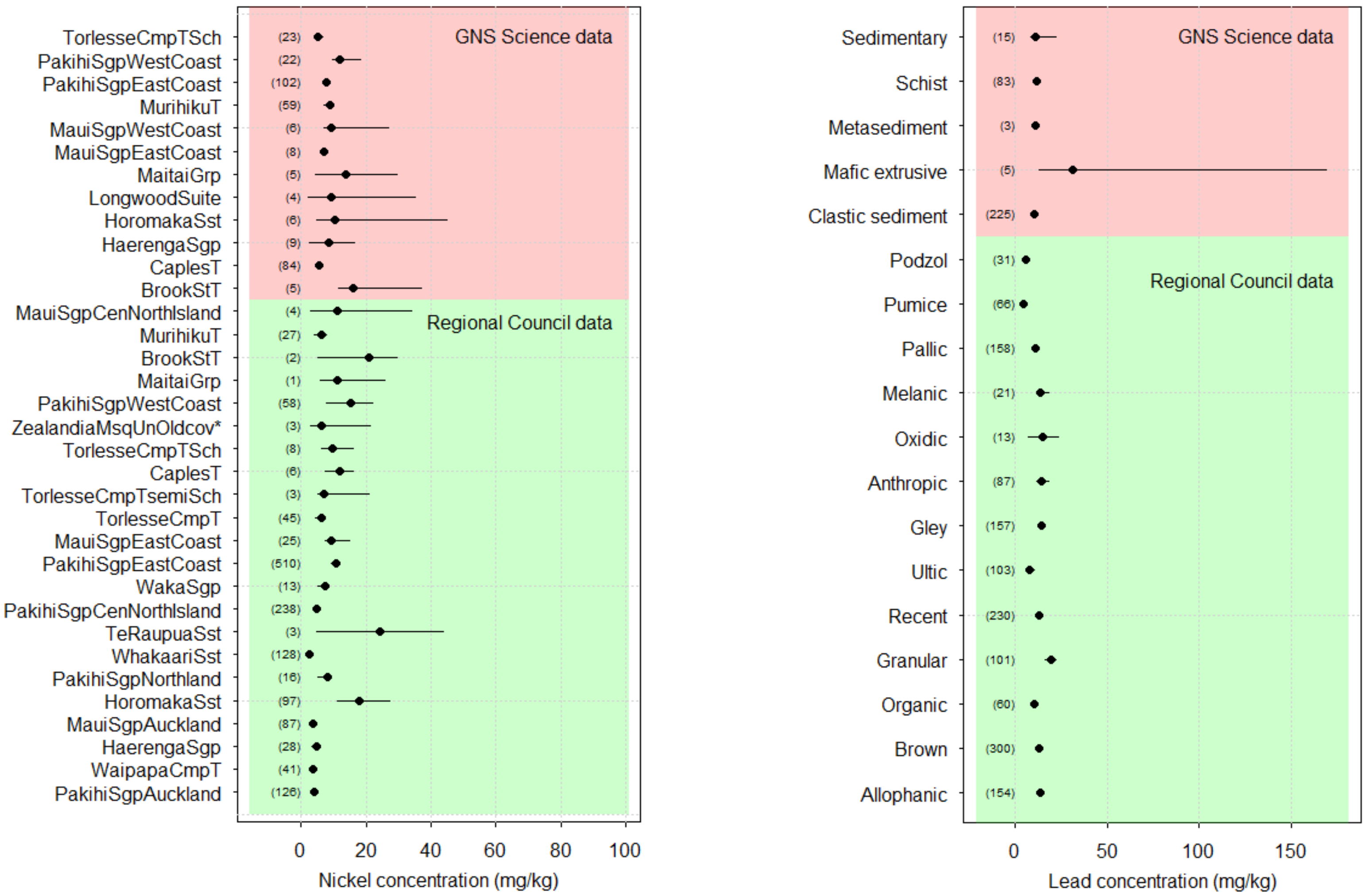

Figure 30 Comparison of predicted background concentrations of nickel (Ni) and lead (Pb) (estimated median concentration (black circle) and estimated 95% confidence interval (black line)) using the favoured 95th quantile models for each of the regional council data (Chemical5 for Ni and Soil Order for Pb) and the Southland-Otago (GNS Science) data (Chemical5 for Ni and Rock Class for Pb). The number of samples in each of the levels for the given explanatory variable is shown at the left of the plot area.

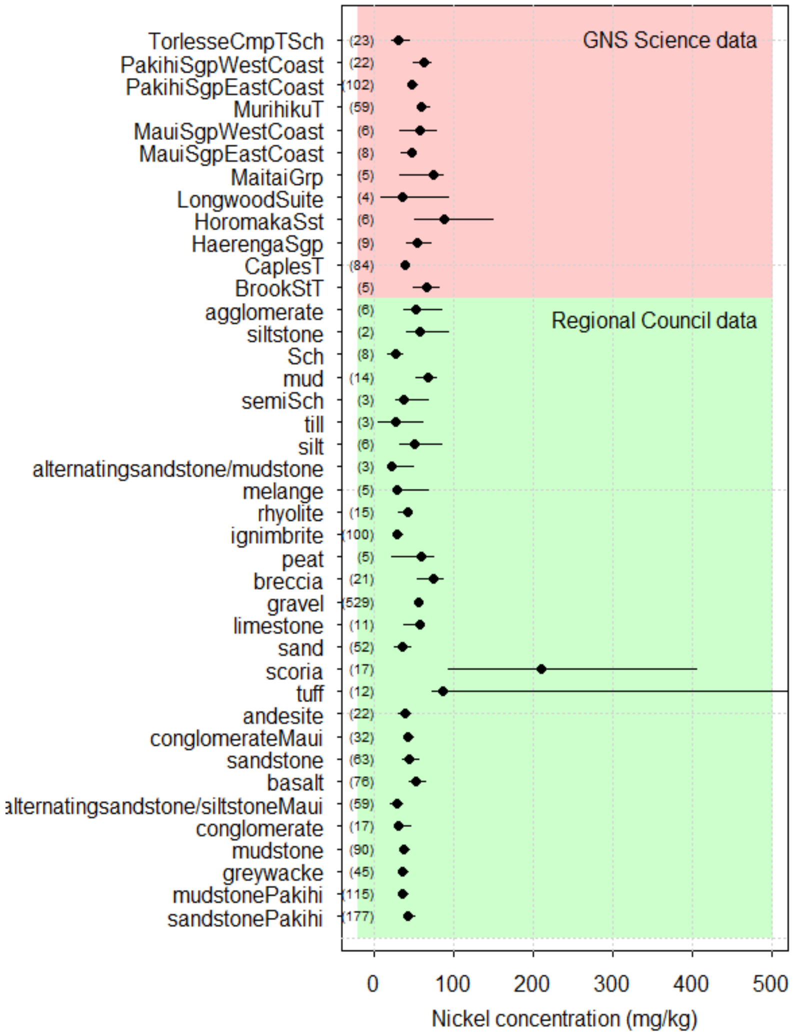

Figure 31 Comparison of predicted background concentrations of zinc (Zn) (estimated median concentration (black circle) and estimated 95% confidence interval (black line)) using the favoured models for each of the regional council data (Chemical4) and the Southland-Otago (GNS Science) data (Chemical5). The number of samples in each of the levels for the given explanatory variable is shown at the left of the plot area.

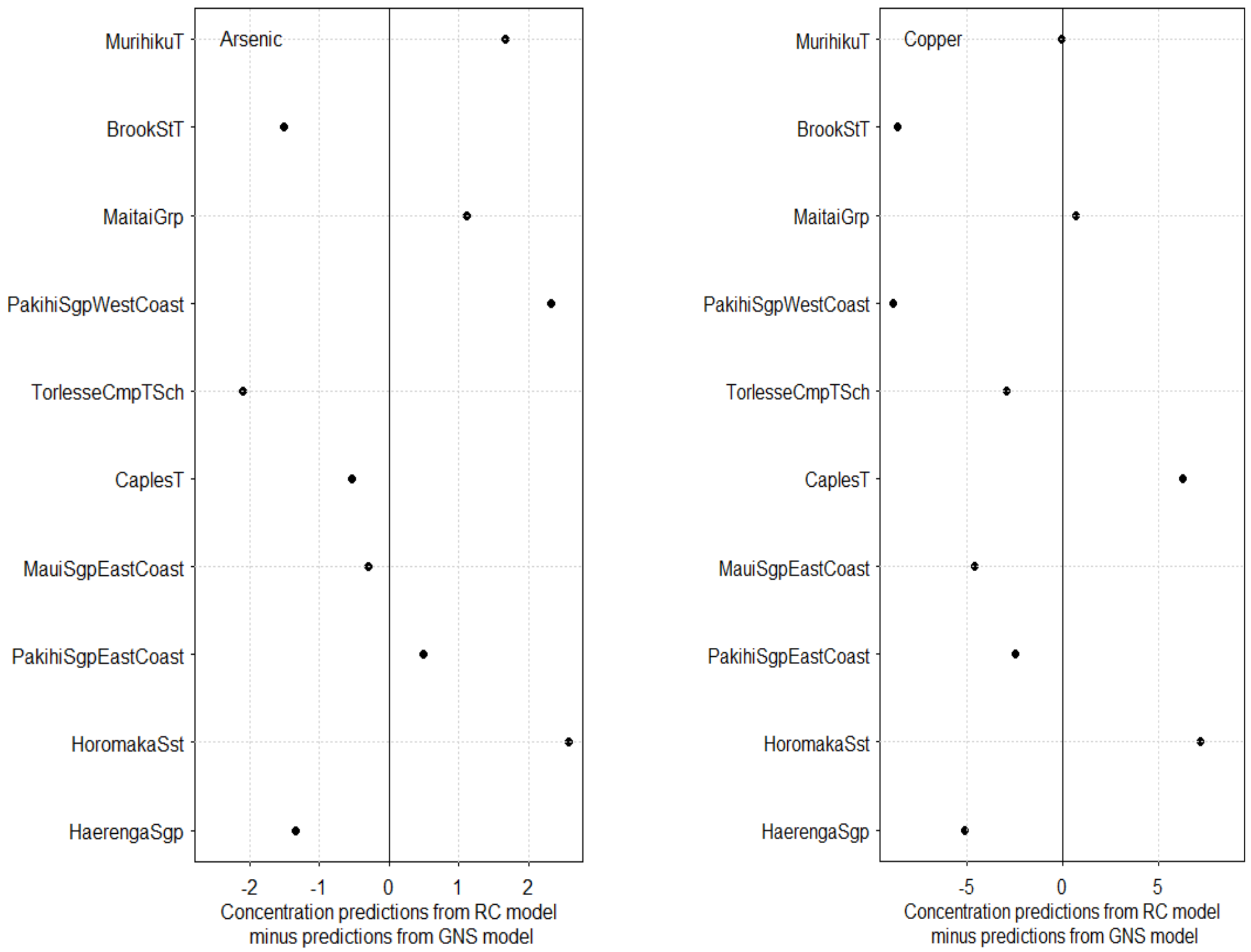

Figure 32 Direct comparison of predicted median concentrations of arsenic and copper based on the models built from the regional council data (RC model) and the models built from the Southland-Otago dataset for Chemical5 parameters.

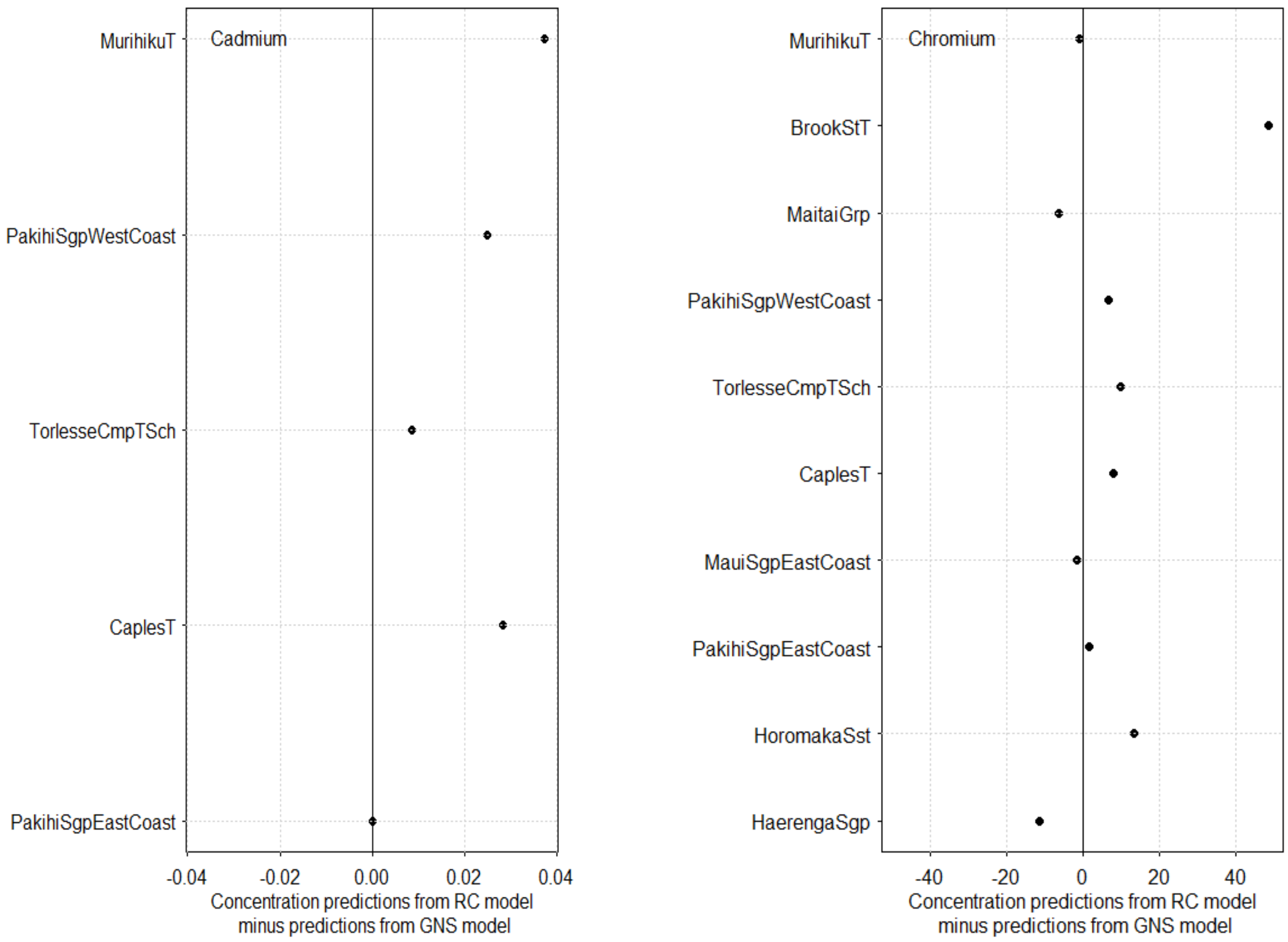

Figure 33 Direct comparison of predicted median concentrations of cadmium and chromium based on the models built from the regional council data (RC model) and the models built from the Southland-Otago dataset for the Chemical5 parameters.

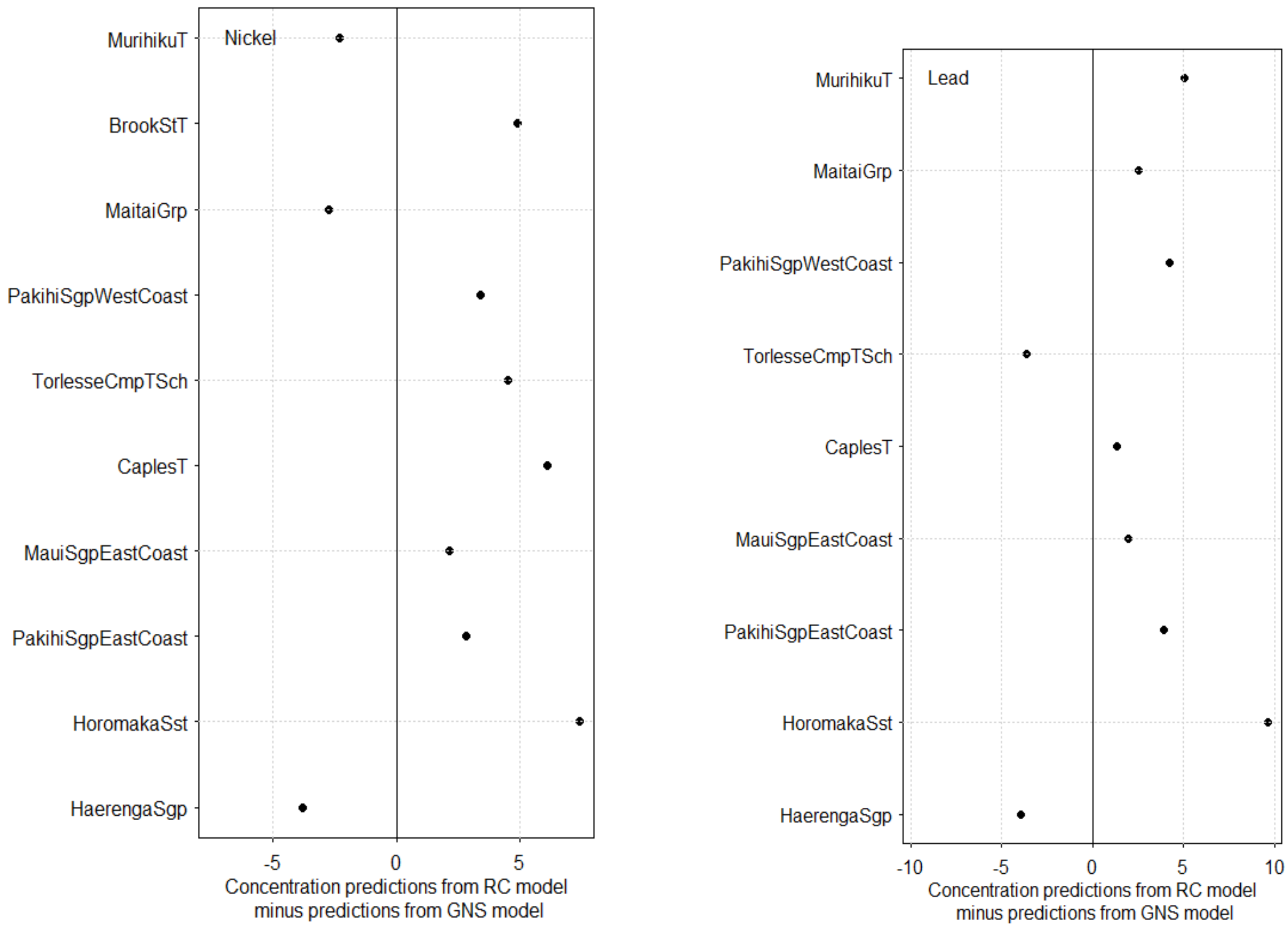

Figure 34 Direct comparison of predicted median concentrations of nickel and lead based on the models built from the regional council data (RC model) and the models built from the Southland-Otago dataset for the Chemical5 parameters.

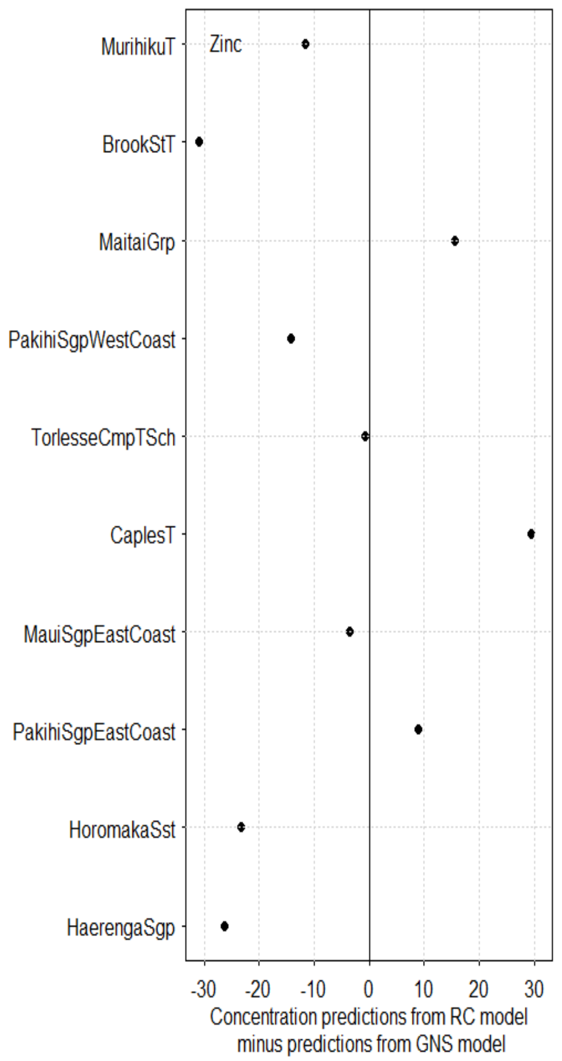

Figure 35 Direct comparison of predicted median concentrations of zinc based on the models built from the regional council data (RC model) and the models built from the Southland-Otago dataset for the Chemical5 parameters.

Appendix 3 – Minerals Exploration Data

The location and concentration of trace elements in the Minerals exploration dataset are shown in the following figures.

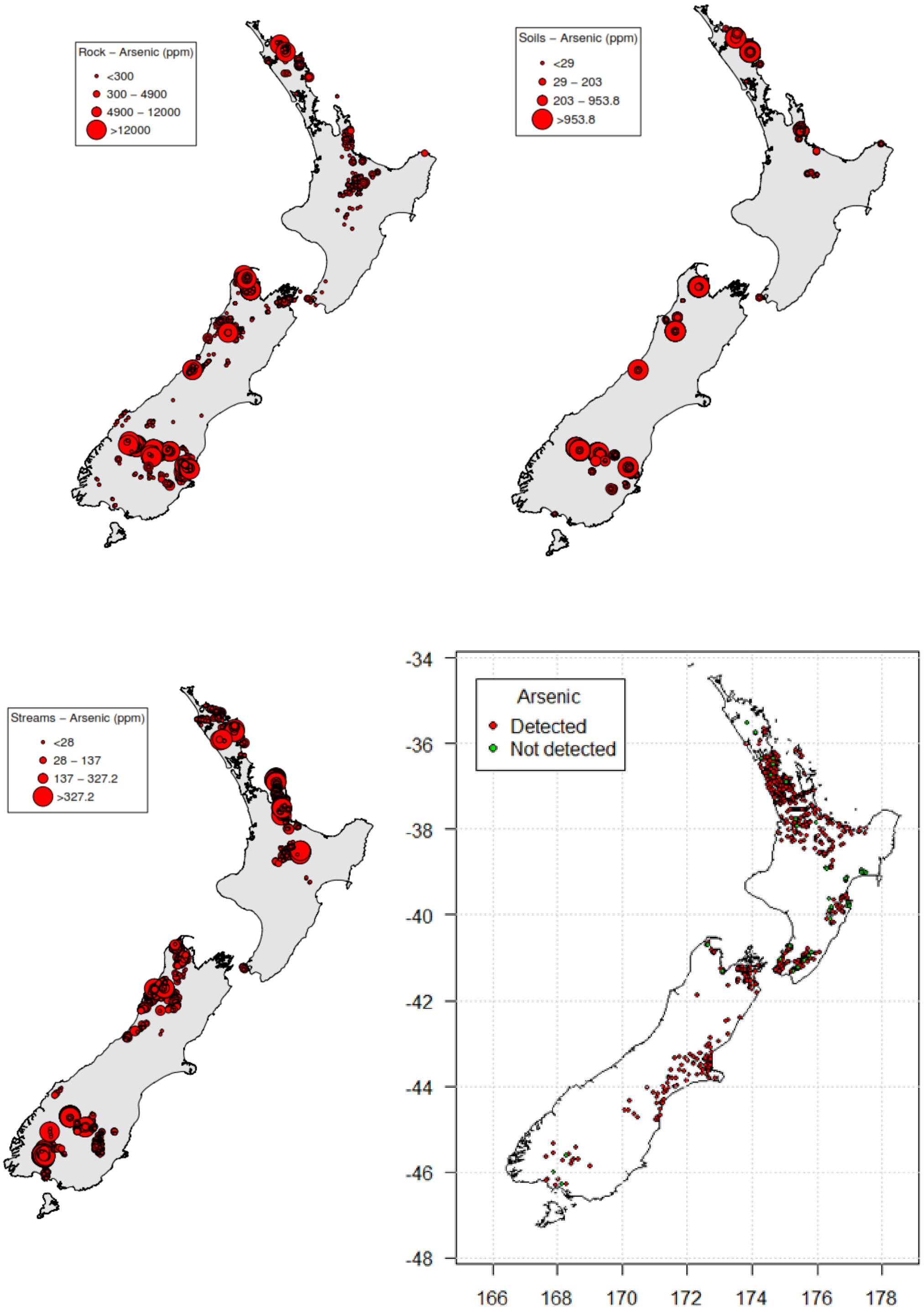

Figure 36 Arsenic in selected rock, soil and sediment samples captured in the Crown Minerals Database for minerals exploration, and location of sampling points within the regional council dataset (in box). Data cut-off points are the 75th, 95th and 99th percentile of the data.

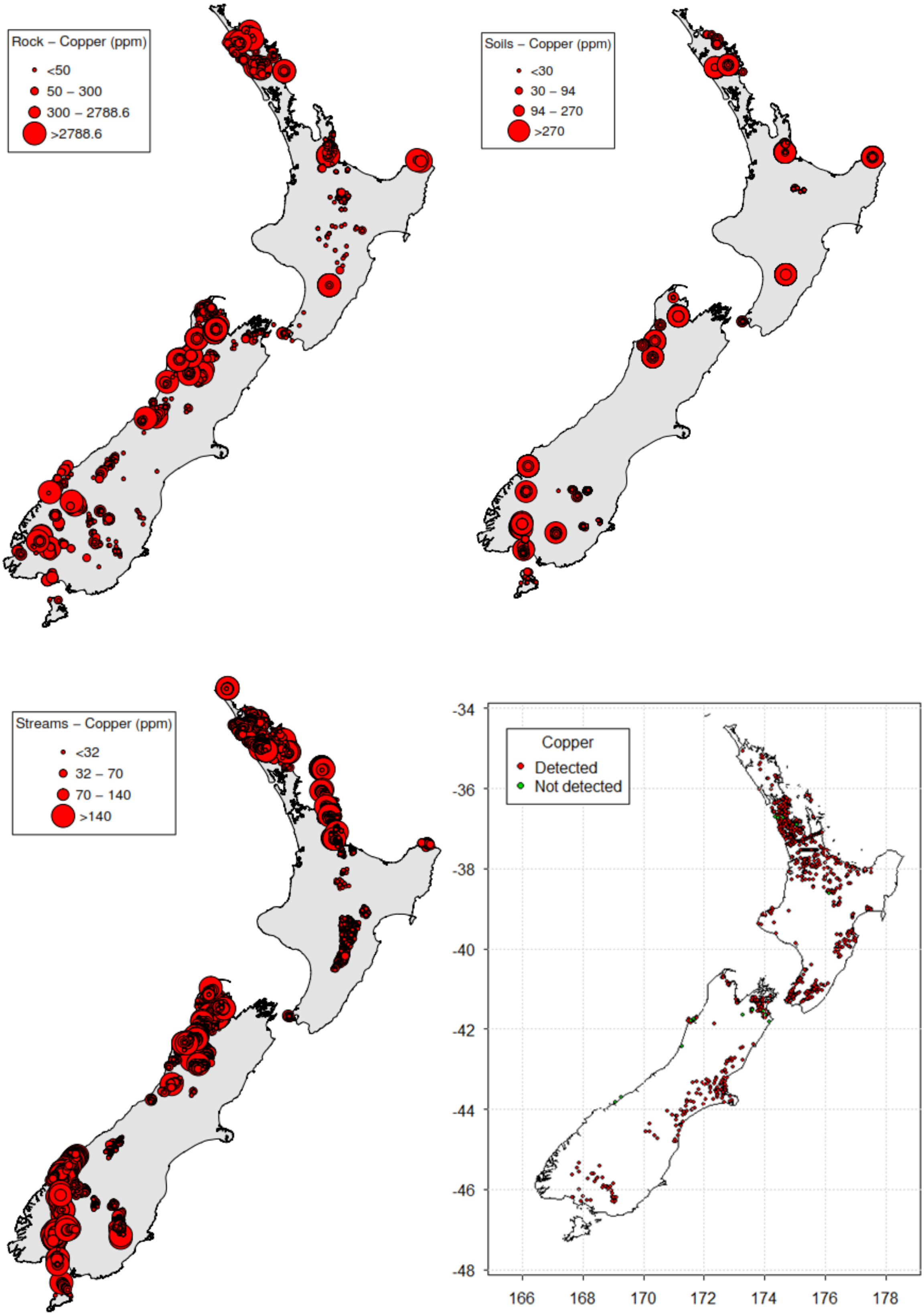

Figure 37 Copper in selected rock, soil and sediment samples captured in the Crown Minerals Database for minerals exploration, and location of sampling points within the regional council dataset. Data cut-off points are the 75th, 95th and 99th percentile of the data.

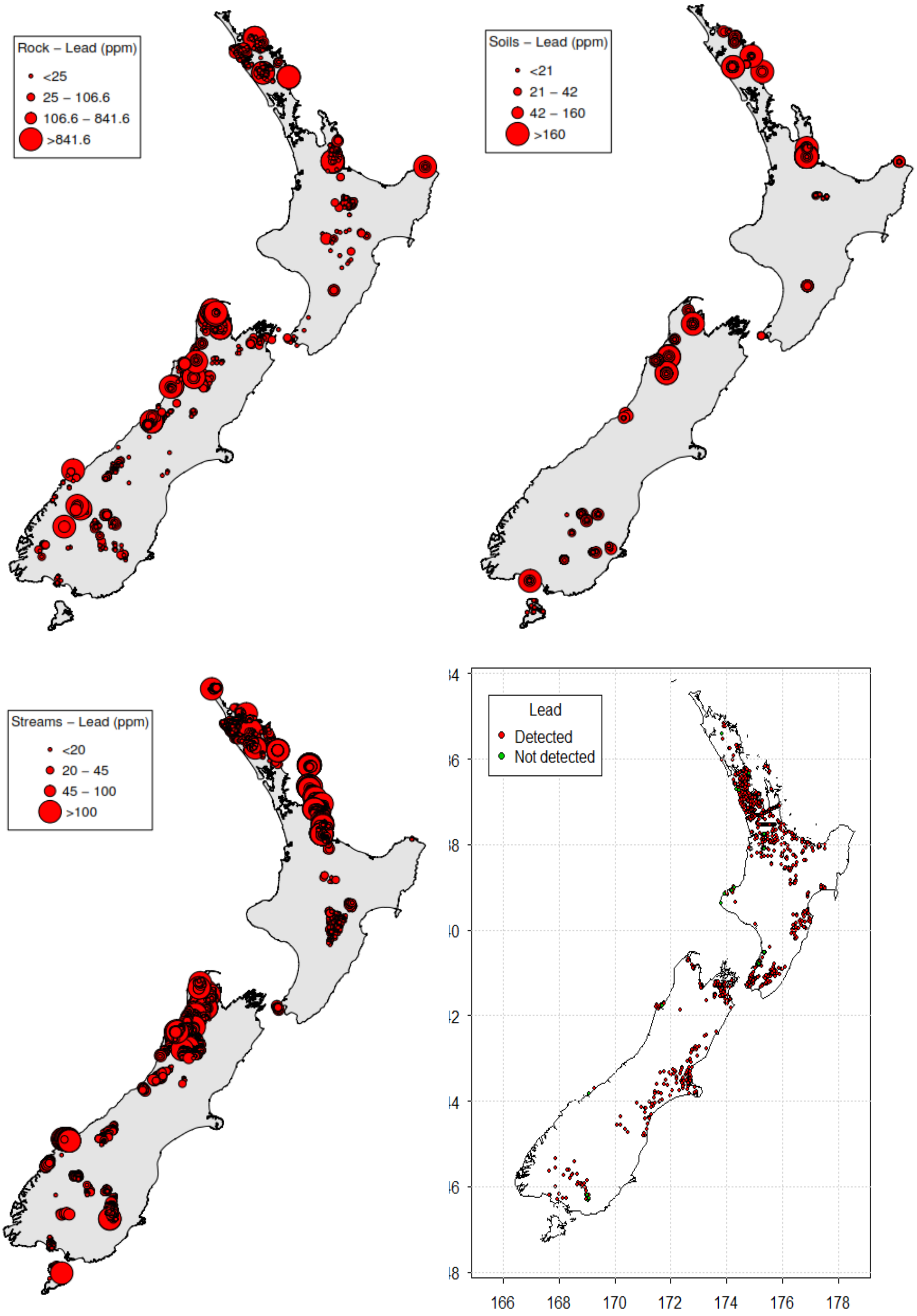

Figure 38 Lead in selected rock, soil and sediment samples captured in the Crown Minerals Database for minerals exploration, and location of sampling points within the regional council dataset. Data cut-off points are the 75th, 95th and 99th percentile of the data.

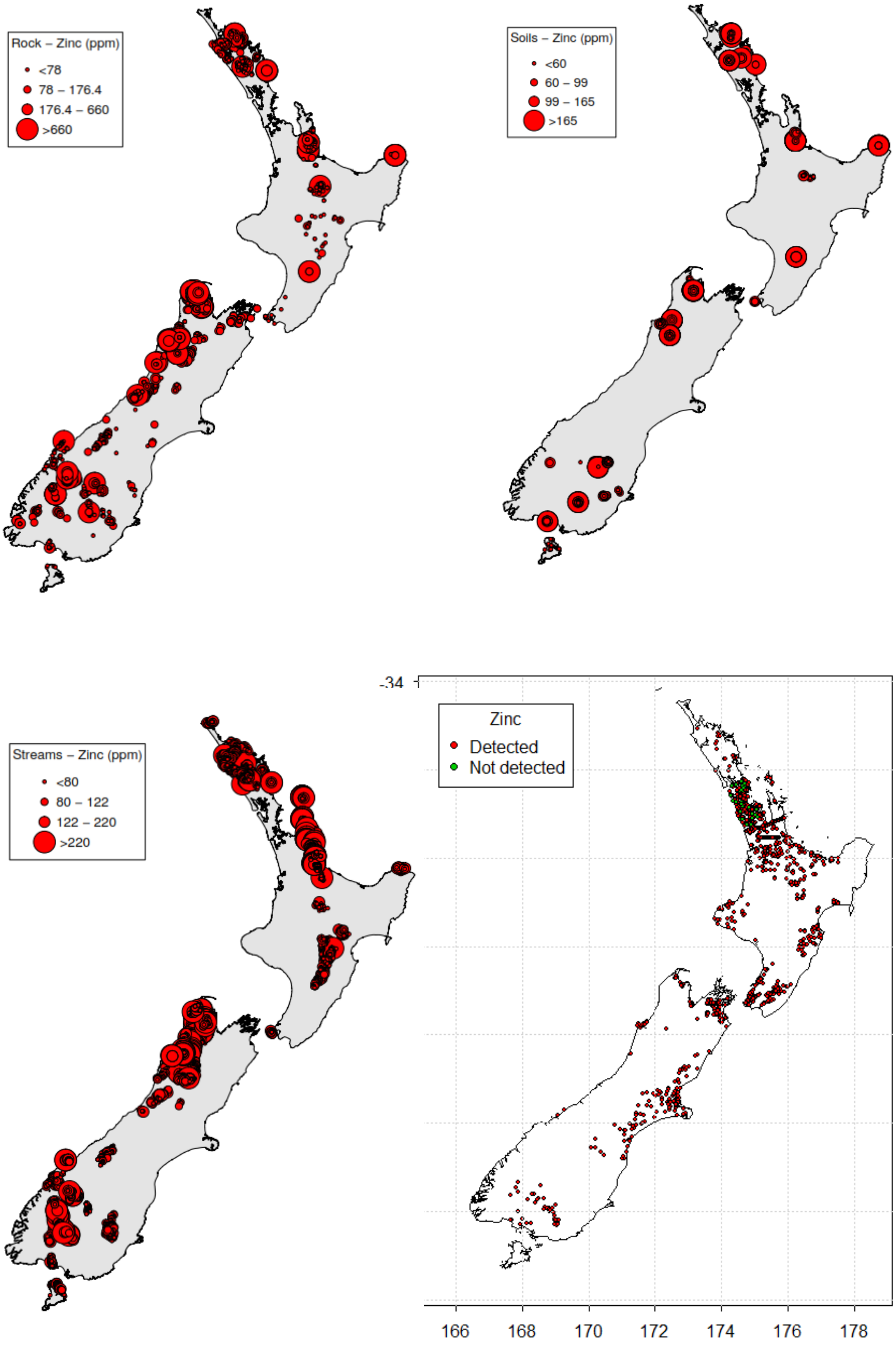

Figure 39 Zinc in selected rock, soil and sediment samples captured in the Crown Minerals Database for minerals exploration, and location of sampling points within the regional council dataset. Data cut-off points are the 75th, 95th and 99th percentile of the data.

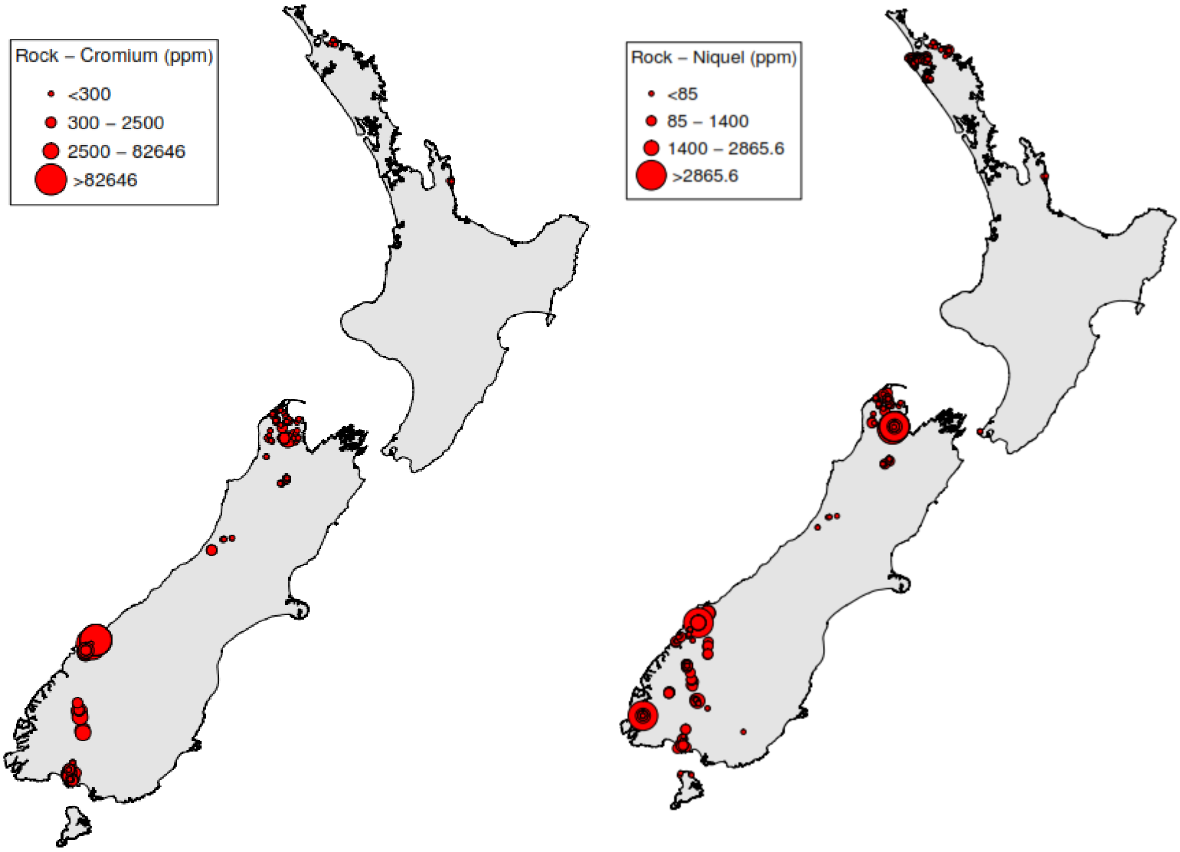

Figure 40 Chromium and Nickel in selected rock samples in the Crown Minerals Database for minerals exploration. Data cut-off points are the 75th, 95th and 99th percentile of the data.