# RECOMMENDATIONS AND GUIDELINES FOR ACQUIRING AND ANALYZING SOIL GEOCHEMICAL DATA TO ESTIMATE THE RANGE OF BACKGROUND CONCENTRATIONS FOR RISK ASSESSMENTS

*A.N. Rencz, R.G. Garrett, E.C. Grunsky, I.M. Kettles, R.A. Klassen, and Rick J. McNeil,*
*Natural Resources Canada - Geological Survey of Canada, Ottawa, ON*

## PUTTING THE RECOMMENDATIONS IN PERSPECTIVE

The development of robust soil quality guidelines require that methods for collection and analysis of soil geochemical data be consistent and based on sound science. The methodologies described support risk assessment and include a procedure for estimating geochemical background. One aim while developing the recommendations for methodologies, was to promote the use of standardized sampling and analytical protocols for soils. The use of common protocols facilitates the comparison of soil and other data generated by different groups at different times.

The standard set of procedures proposed here serve as a starting point for data comparisons but other methodologies should be considered to provide optimal characterization of the sites being assessed. For example, if information on human exposure to metals or bioaccessibility is required, the use of additional methods of sampling and analyzing soils and other media should be considered. These include the sampling of additional soil horizons or depth intervals (e.g. A-horizon or 0-30 cm interval ) and the use of other techniques for sample preparation and analysis such as the water leach, physiologically-based extraction techniques (PBETs), and element speciation techniques (Garrett et al, 2009a; Dodd, 2011 (this release); and Parsons, 2011 (this release)).

There are four sections in the Recommendations document covering the spectrum from sample collection through data analysis. The first section has the recommendations for *Site Selection, Field Sampling, and the Use of Thematic Map Layers for Data Plotting*. The second covers *Analytical Techniques* and provides information on how the different methods of sample preparation and analysis cause variation in the resultant data. The third outlines an approach and provides a procedure or *Estimating Geochemical Background.* The fourth section has information on how to access existing sets of published geochemical data for Canada. At the end of this document, the *Appendices* provide additional information that pertains to the procedure for estimating geochemical background.

# 1. SITE SELECTION, FIELD SAMPLING AND THE USE OF THEMATIC MAP LAYERS FOR DATA PLOTTING

For the development and implementation of a sampling design, several aspects of the geochemical landscape are considered. Two aspects are the patterns of change in soil composition with increasing depth (vertical variation), and the patterns of variability over an expanse of geographic area (horizontal variation). There are also differences in chemical composition between the different types of sample media (e.g. soil vs. vegetation) and between different phases of the same media (e.g. leaves vs. bark in trees), a few of which are noted here.

## 1.1 Vertical Spatial Variation

a) Within the uppermost metre of the soil there are variations in the textural, mineralogical, moisture, and organic composition that are expressed as soil horizons (Fig. 1). In most places, the horizons present in the soil profile can be distinguished visually. The chemical makeup of the materials in each horizon reflects the composition of the soil parent materials and the physical, chemical, and biological processes that have affected them over time. Each horizon formed as a result of a characteristic set of soil-forming processes. From one horizon to the next, there is variation in the amounts of sequestering mineral substances such as Fe- and Mn-sesquioxides, clay minerals and organic matter. The patterns are mostly predictable as are the effects of the variation on chemical composition. Consequently, based on

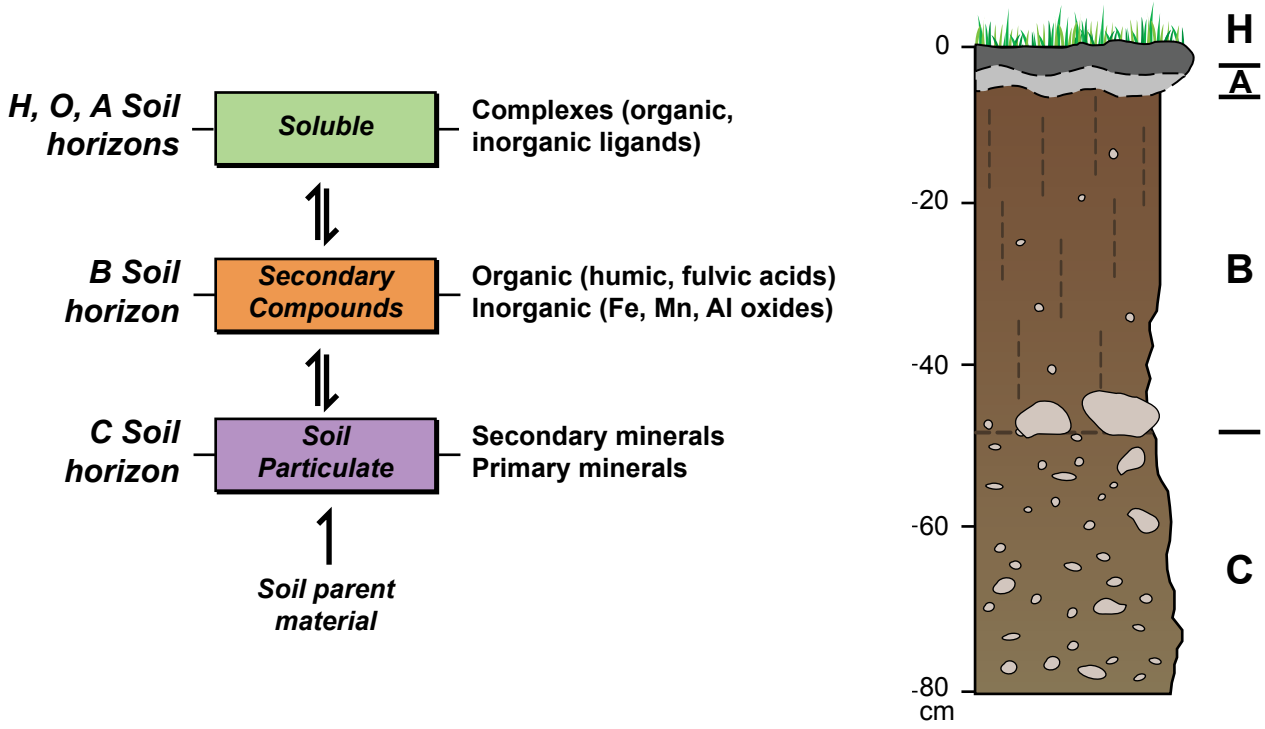



***Figure 1.*** *Diagram shows an idealized soil profile (left side) and descriptions of soil materials and mineral phases.*

the assumption that similar soil-forming processes have affected similar horizons, it is easier to explain the causes of geochemical variation in soils if samples are collected from similar pedologic horizons. The collection of samples from intervals of fixed depth introduces more variation into the resultant data because the patterns of horizon development and the thicknesses of the individual soil horizons vary significantly from site to site and region to region.

To establish geochemical background and make reliable comparisons of data from different soil surveys, sampling by soil horizon is preferable to sampling by depth interval (McNeil (2011 (this release)). Procedures for collecting horizon-based soil samples are documented in Friske et al. (2011) and information on changes in soil composition with depth is presented in Grunsky and McNeil (2011 (this release)).

b) At a bare minimum, the collection of two types of samples is recommended for risk assessment purposes. The first is from the C horizon, the data from which are used to establish geochemical background. The C horizon is considered to be the layer least affected by soil forming processes. The second sample type is the 0-5 cm depth interval and is referred to as the Public Health (PH) layer. Data from this sample type concerning bioaccessibility provide information on the risks due to exposure to the soils. Although the PH is a depth interval for sampling as opposed to a pedologic horizon, as recommended above, it is the sampling interval currently used by Health Canada. It was chosen because it is the soil layer that living organisms are in direct physical contact.

If there is to be additional sampling, it is recommended that it follows the principle that horizon-based is preferable to depth interval-based sampling. For example, the United States Geological Survey, as part of the North American Soil Geochemical Landscapes Project, elected to collect and analyze samples from the A horizon (excluding the leached materials from the Ae subhorizon) (Woodruff, 2010). The resulting data provide information on the biologically active zone that is more easily compared at local and regional scales. Also, samples from the upper B horizon were collected in many areas of Canada for mineral exploration purposes.

### 1.2. Horizontal spatial variation

a) A spatially random sample design should be used to ensure statistically defensible estimates of the background range of element concentrations in soils (Garrett, 1983). A stratified random design is preferable because it ensures that all possible sites have an equal opportunity of being sampled. Details on the design used for the North American Soil Geochemical Landscapes Project are provided in McNeil (2011 (this release)), Friske et al. (2010) and Garrett and Kettles (2010).

Across Canada soils have developed on different types of surficial materials (till, glaciofluvial sands and gravels, glacial lake sediments, glaciomarine sediments, etc.), that were, in turn, derived from diverse

types of bedrock. More information on the geological controls on the geochemical composition of soils is provided in Klassen (2011 (this release)). When collecting samples to assess geochemical background, it is prudent to assess how representative the targeted site is with respect to overall variability in the area of interest. If the immediate study area is contaminated, a nearby area in the terrane that is pedologically and geologically similar to the contaminated study area should be sampled in order to estimate the most likely properties of background in the contaminated area.

b) If the objective is to detect contamination arising from natural sources or as the consequence of human activities, prior knowledge of the expected size of the contaminated area is required. This information may be garnered from published geological and geochemical maps, field observations, and historical documents. Samples should be collected on a grid pattern (Garrett and Grunsky, 2011 (this release)). The area of the individual cells of the grid should be equal to or smaller, preferably by half, than the expected area of the contamination target. Knowledge of the dispersal process from the source of contamination is helpful when setting the orientation of the sampling design. If the 'targets' are 'elliptical' rather than circular the grid should be rectangular rather than square.

c) An adequate number of samples are needed (Garrett and Grunsky, 2011 (this release); Garrett, 2010). For calculation of geochemical background using univariate statistical methods, 30 samples is the fewest number of samples, with 60 being better. However, after 120 samples, it is likely that not much precision is being added to the estimation. When multivariate statistical methods are used, there should be at least 8 or 9 times as many samples as the number of variables. For example, if there are 10 variables there needs to be at least 80 or 90 samples, with a couple of hundred being better.

### 1.3 Geochemical variation between different sample media

Numerous media (e.g. soils, rocks, water, surficial sediments, plants) have been used to characterize the geochemical nature of the environment. For terrestrial environments soil are a useful sampling media for risk assessment. They are a direct source of elements for biotic systems and they are relatively easy to sample.

a) An ecosystem is composed of numerous biotic and abiotic components and there is considerable variation in chemical composition between these "media". For example there may be up to a magnitude of difference in the concentration of elements in two types of tissue collected from the same biological species (e.g. Ni in the roots versus the leaves of sugar maples as described in Ford et al. (1988)). To avoid this variation select only one type of media and sample the same types of materials from the selected media type.

b) Biotic and abiotic components are linked in an ecosystem and elements flow from one compartment to the next. Researchers have used biological samples to infer element concentrations in soil. However, there is not necessarily a strong correlation between the two (Ford et al., 1988). Consequently, caution should be exercised in making such comparisons.

c) Soils are recommended as the preferred media for sampling to establish geochemical background in the terrestrial environment. Unlike vegetation or living organisms, soil chemical characteristics are not significantly affected by seasonal variations. Samples collected at lower depths, notably the C horizon, are less affected by temporal changes including contamination. Soils are present in most places and readily accessible. In addition, there is a large body of knowledge related to the pedological processes that characterize the different soil horizons and their effects on geochemical composition.

### 1.4 Use of Thematic Map Layers for Data Plotting

In a GIS environment, geochemical data may be contoured or plotted as proportional symbols on thematic map layers (e.g. bedrock geology, surficial geology, soil groups, drainage basin, and land classification). When the distribution patterns of soil data are examined on themed map layers it provides a visual context for the interpretation of the geochemical variability. Thematic layers also provide a framework for data integration and the generating data subsets for statistical analysis. A useful regional-scale map base available for North America is one of eco-classifications, presented in Kettles (2011 (this release)). Further information of systems of eco-classification is available in the State of the Environment Reporting by the CANSIS group at Agriculture and Agri-Food Canada (Ecological Stratification Working Group. 1996).

For studies undertaken at detailed scales or to address specific concerns, comparisons of soil geochemical data based on other spatial criteria may be advantageous. Some examples are soil type, surficial materials unit, underlying bedrock unit, and catchment areas for hydrological drainage basins. There is information on how to acquire existing data sets in Kettles et al. (2011 (this release).

## 2. ANALYTICAL TECHNIQUES

### 2.1 Sample preparation for chemical analyses

Sample preparation and chemical analyses are discussed in McNeil and Garrett (2011 (this release)). Existing practices used for soil sample preparation are documented in Girard et al. (2004) and Sheldrick (1984).

a) For practical purposes, the < 2 mm fraction is recommended as a standard for geochemical analysis. This fraction has traditionally been used by agronomists and the practice has continued in the field of

environmental research and risk assessment. Hence, there is a large body of existing data based on analysis of this fraction. However, there are limitations associated with its use, some of which are considered in Klassen (2011 (this release)).

There is considerable variation in chemical composition between textural classes, For many trace and minor elements, concentrations increase as particle size deceases. The geochemical substrates that have the capacity to concentrate trace elements have large surface areas, high cation exchange capacities, and high surface charges (Horowitz, 1991). The most common of such substrates are hydrous manganese oxides, hydrous iron oxides, organic matter, and clay minerals, all of which tend to be concentrated in the finer size fractions. Data from analysis of the silt-sized and finer fractions (<0.063 mm) may provide more information on the mineral phases and residence sites of elements in soils and the patterns of regional variation for many trace elements (Shilts, 1975, 1984).

b) Ball or ring mill pulverizing should not be considered for risk assessments as the process may bias the final results. Such milling is used only for total analysis (using 4-acid digestions or x-ray or nuclear methods), but not for preparing samples to be treated with aqua regia and its variants, the water leach or other partial extractions. Although sample materials are dried and screened prior to chemical analysis, if not touched by pulverizing, the mineral fragments resistant to physical and chemical weathering are not disaggregated and hence, remain closer to their natural state.

c) Splits from all samples should be archived and stored. This provides materials to be used for cross-checking purposes and also for further analyses as new issues arise and/or new analytical methodologies are developed.

**2.2. Different types of chemical analysis**

a) There are many techniques for chemical analyses, each having their advantages. To support consistency, a commonly used aqua regia digestion is recommended. A study undertaken at the Geological Survey of Canada (Garrett et al., 2009b) has shown that there are only small differences between various aqua regia-like digestions involving various combinations of HCl, $HNO_3$ and water for most trace metals commonly subject to environmental reviews. For this reason it is recommended that the widely used and accepted USEPA 3050B aqua regia variant be used, i.e. a 4:1 HCL – $HNO_3$ digestion. This is a relatively strong leach. When compared to results obtained using a total method (INAA), it provides similar total element recovery for some elements (e.g. Cu, Pb and Zn) but poorer recovery for others such as Cr and V. Selected results from the aqua regia study are shown in McNeil and Garrett 2011 (this release).

b) In addition to the aqua regia analyses a method for estimating the amount of loosely held ‘bioaccessible’ amounts of the total element concentration should be considered. There are a number of techniques for

assessing bioavailability. The water leach method is one method that it is relatively easy to use (Garrett et al., 2009b).

c) Speciation is relevant for estimating soil toxicity. For example, there is considerable difference between the toxicity of $As^{+3}$ compared to $As^{+5}$. Some information on speciation is provided in Parsons (2011 (this release)), but additional information and more research are required before methodologies can be recommended. These types of analysis are very costly at present and, hence, any requests for them to be undertaken need to be soundly justified.

**2.3. Quality Assurance/Quality control (QA/QC)**

The quality of the geochemical survey and the resulting chemical data can be evaluated by the insertion, analysis and monitoring of QA/QC samples. Procedures for this undertaking were developed as part of the National Geochemical Reconnaissance Programme at Geological Survey of Canada (Friske and Hornbrook, 1991) and further information is provided in McNeil (2011 (this release)) and McNeil and Garrett (2011 (this release)).

a) Collect a field duplicate some appropriate distance from a 'regular' sample site. The distance will be a small proportion of the distance between 'regular' sample sites and reflect the probable uncertainty in reoccupying the site at a later date, usually some distance between 5-10 metres.

b) It is necessary to include analytical duplicates, i.e. a split of one of the field samples in order to determine the analytical precision of the data. When field duplicates have been collected, it is advantageous to generate the analytical duplicate data from a split of one of each field duplicate pair.

c) It is essential to include an aliquot of a control reference material (CRM) in each batch of analyses, to ensure that there is no 'analytical drift' during the course of a project, and, if appropriate, between projects. For large projects more that one CRM should be used.

Some internationally certified CRMs are available where the analyses have been undertaken with an aqua regia-related digestion. To the greatest possible extent the CRM(s) should be derived from similar types of surficial materials to those of the collected samples. Where CRM matrices are unrelated to that of the sampled materials, instrumental-mineral effects relevant to assessing either analytical accuracy or 'drift' may not be recognized.

d) A frequency of insertion of QA/QC materials at 1 in 20 is a good compromise between adequate QA/QC and minimizing overhead costs. Thus each group of 20 field samples should include three QA/QC samples- field duplicate, analytical sample and a CRM.

e) To ensure that the QA/QC samples are not treated with additional care by laboratory staff, they should be inconspicuously labelled and placed randomly among the routine survey samples so their locations are not obvious.

f) A variety of graphical methods are available to monitor and evaluate data quality. Control charts can be plotted for CRMs, using date of analysis or batch number for plotting. Where an established CRM is used tolerance bounds can be established, and the laboratory can be alerted if samples fall 'out of tolerance'. Field and analytical duplicates can be plotted on Thompson-Howarth diagrams (Thompson and Howarth, 1978). If more than the expected number of duplicate analyses fall 'out of tolerance', follow-up can be initiated with the laboratory or a field investigation can be undertaken.

g) Statistical summaries and analyses may be undertaken. Precision, quantified by the percentage relative standard deviation (RSD) or coefficient of variation (CV) can be estimated from the repeat analyses of the CRM(s) and analytical duplicate pairs. Analysis of Variance (ANOVA) can be used to determine if the field or analytical variability are sufficiently small relative to variability across the study area to be able to discuss spatial distribution with confidence. Where the analytical duplicates have been split from the field duplicates, ANOVAs can be undertaken where study area, local field and analytical variability are evaluated as a whole.

h) It is the responsibility of the investigating scientist to ensure, independently form a service analytical laboratory, the integrity of the geochemical survey data. This is achieved by generating and analyzing the QA/QC data.

# 3. ESTIMATION OF GEOCHEMICAL BACKGROUND

The discussion of geochemical background that follows is provided in Garrett and Grunsky (2011 (this release)). The concept of geochemical background was developed in the 1940s and 50s. It was introduced to differentiate between normal element concentrations and chemical anomalies that might be indicative of ore mineral occurrences. Background was defined by Hawkes and Webb (1962) as: "The normal abundance of an element in barren earth material". They also concluded that "It is more realistic to view background as a range rather than an absolute value". Today geochemical surveys have equal applicability to environmental issues and concerns.

Geochemically there is no immediate difference between an anomaly arising from a natural process, e.g., the formation of a mineral deposit, or as a result of contamination of the natural environment by an anthropogenic process. Both types of processes impose an 'overprint' on the natural regional geochemical

background. Although some would argue that natural background no longer exists, anthropogenic processes have distributed contaminants world-wide, a look at regional geochemical maps demonstrate that natural processes still dominate the distribution of trace-elements on regional and continental scales (Reimann et al., 2009, 2010).

Two terms related to geochemical background are commonly used: natural background and ambient background. Ambient background is the sum of natural background levels and any anthropogenic additions. Background levels usually differ between sample media, thus levels in air, water, lake, stream/river and marine sediments, and soils will be different. However, all will be related to their source materials, i.e. rocks, and these will be modified by the physical, chemical and transport processes characteristic of the sample media. Difference in rock chemistry (geochemistry) for a single element may vary by orders of magnitude (Garrett, 2005).

### 3.1 Univariate methods

A wide variety of procedures are available for estimating the range of geochemical background variation, see Reimann et al. (2005) and Reimann and Garrett (2005), and none give similar estimates. On the understanding that geochemical background is a range and not a single value that represents "the normal abundance of an element in barren earth material" (Hawkes and Webb, 1962), where 'barren' indicates devoid of the influence of ore deposits or site specific anthropogenic contamination, a simple approach is to use the percentiles of a set of background geochemical data representing the study area.

To demonstrate the steps needed to estimate geochemical background we have used a set of as yet unpublished geochemical data for soil samples collected from selected horizons at sites in the Maritime Provinces of Canada (Friske et al., 2011). These data are for the <2 mm fraction of samples collected between the ground surface and a depth of 5 cm. The samples were analyzed using ICP-MS after digestion with an aqua regia variant (USA-EPA 3050B). The format of the data set is shown in Figure 2.

#### a) Software to construct maps and graphs

The maps and graphs that follow were plotted, except where indicated, using the 'rgr' library (Garrett and Chen, 2007) and functions in R. R is an Open Source software environment for statistical computing and graphics. It runs on a wide variety of UNIX, Windows and MacOS platforms. It is managed by the Comprehensive R Archive Network (CRAN) and R and associated packages can be downloaded from a number of CRAN mirror sites linked to http://www.r-project.org/. Should it be decided to use R and 'rgr', it is recommended that before downloading any software you set up a working directory on the data drive of your computer (e.g. D:\\R\WD) to store the files produced using 'rgr'.

One mirror site is at the University of Toronto, http://probability.ca/cran/. Within the site there versions of R and also 'rgr'. The 'rgr' package is available to be downloaded from the 'Contributed Packages' section.

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ID | Lat | Lng | Prov | EcoP | Ecoprovince | EcoR | Ecoregion | Soil Type_SLC | SubRx Type (Wheeler' |
| 2 | NB071001 | 45.9173 | -66.7055 | NB | 7.2 | Northumberland Lowlands | 122 | Maritime Lowlands | Podzolic Gray Luvisolic | Nonmarine Sedimentary |
| 3 | NB071002 | 45.7888 | -66.5366 | NB | 7.2 | Northumberland Lowlands | 122 | Maritime Lowlands | Gleysolic | Nonmarine Sedimentary |
| 4 | NB071003 | 45.7734 | -66.1785 | NB | 7.2 | Northumberland Lowlands | 122 | Maritime Lowlands | Podzolic Gray Luvisolic | |
| 5 | NB071004 | 46.1842 | -67.0483 | NB | 7.3 | Fundy Uplands | 121 | Southern New Brunswick Uplands | Ferro-Humic Podzolic | Sedimentary Rocks Un |

| | K | L | M | N | O | P | Q | R | S | T | U | V |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | As_PH | Pb_PH | Corg_PH | As_Ahor | Pb_Ahor | Corg_Ahor | As_Bhor | Pb_Bhor | Corg_B-hor | As_Chor | Pb_Chor | Corg_Chor |
| 2 | 0.7 | 6.67 | 43.1 | | | | 10.9 | 12.51 | 3.55 | 12.7 | 14.17 | 0.73 |
| 3 | 4 | 22.48 | 27.09 | 4 | 17.8 | 6.28 | 3.4 | 6.88 | 0.89 | 5.9 | 8.91 | 0.11 |
| 4 | 2.9 | 10.79 | 3.64 | | | | 6.1 | 10.01 | 2.47 | 8.6 | 8.9 | 0.37 |
| 5 | 2.1 | 31.2 | 12.31 | 1.1 | 13.3 | 7.74 | 3.4 | 8.64 | 1.64 | 4.3 | 7.49 | 0.45 |

***Figure 2.*** *Example from dataset (Friske et al., in preparation) used for demonstration purposes to estimate geochemical background. Data were obtained from soils collected as part of the North American Soil Geochemical Landscape Project. The abbreviations represent the following: ID – Sample number; Lat – Latitude; Lng – Longitude; Prov – Province of Canada; EcoP – Ecoprovince of Canada (numerically coded), based on Ecological Stratification Working Group (1996); EcoR – Ecoregion of Canada (numerically coded); Soil Type_SLC – soil unit based on the Soil Landscapes of Canada (1996); SubRx Type – underlying bedrock type after Wheeler et al. (1997); As – arsenic; Pb – lead; C org - % organic carbon; PH – public health interval (0 – 5 cm); Ahor – A horizon of soil; Bhor – B horizon of soil; Chor – C horizon of soil.*

It is downloaded as a zipped file and should be left in that format on your computer. At the same time it is necessary to download and save one other package - akima - from the 'Available Packages' section. It is a dependency needed to run 'rgr'. Use of the 'rgr' package requires that R is up and running on the computer on which the 'rgr' package is to be installed. Figure 3 shows the first window that appears once R is started. The package 'rgr' is brought into R by first clicking on 'Packages' on the drop down menu, as shown in Figure 2, and then clicking on the last option 'Install packages(s) from local zip files'. The two dependencies, akima and MASS, need to be installed using the same procedure.

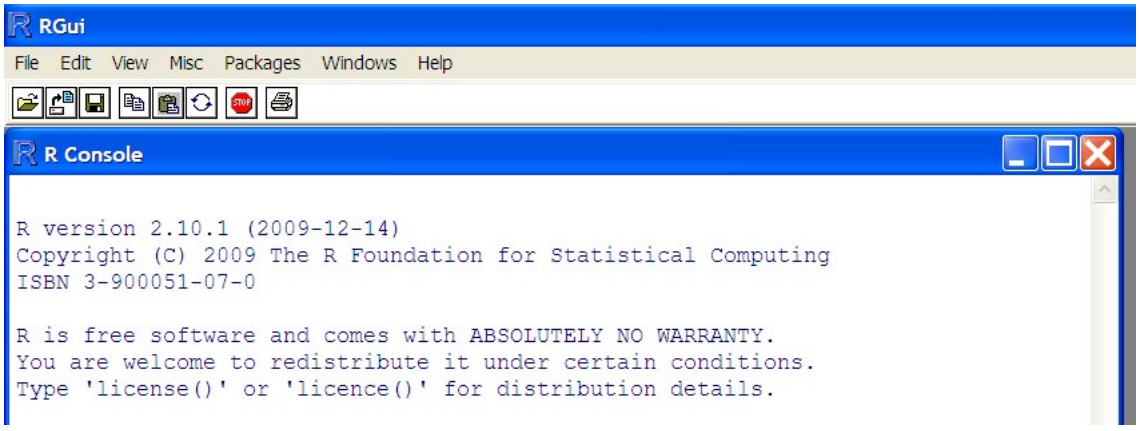



***Figure 3.*** *First window that opens in the R software programme.*

The next step involves installing a 'rgr.first' function, see Appendix 02-01. Open 'rgr.first' in the software program Notepad and copy the contents of the file. Return to the R workspace and type: fix(rgr.first). When the editor screen comes up, paste the copied text into the editor. If needed, edit the first line of

the copied text so that it exactly the same as the first line in the editor, and then save the editor file. To execute the function type rgr.first() after the > prompt. Please, note that from this point on in any new work session, 'rgr' is initiated by typing rgr.first() after the > prompt.

Appendix 02-02 is a workspace generated using R and saved. It shows the installations described above and also the addition of the Maritime geochemical data used in this example. The session in R is ended by typing: q(). Please note that it is important to save the workspace if new data (an object) or any functions have been added during the session so that they can be used again in future. If you have not added any new data (i.e. created a new object) or functions, click the 'no' button.

There were over 100 functions written at GSC to support exploration and applied geochemical survey and research activities. An overview of the 'rgr' functions from Open File 5583 (Garrett and Chen, 2007) is presented in Appendix 02-03. The 'rgr' functions described here fall dominantly under the title of univariate exploratory data analysis tools. Within the R workspace, it is possible to see a list of the functions by typing: help(rgr). Go to the help file to see what can be done in 'rgr'. The 'rgr' package also includes the same test data as that used in Reimann et al. (2008) on applied environmental statistics.

**b) Steps for estimating geochemical background**

In the section that follows, three steps for estimating background are outlined in detail. The first step is undertaken to determine if there is only one population in the data, and if there are multiple background populations to generate data subsets for each. The second step is carried out to identify and remove outliers from the data set or subsets. There are subjective aspects within these first two steps that require the analyst to make observations and choices related to the data set. Some tools are provided to aid this decision-making. The third step, the actual calculation of the background range is straight forward. However, there are several ways to make such calculations.

*Step 1: Ensuring only one population in the geochemical data set*
It is necessary to determine if the data are drawn from a single background population. This is accomplished by inspecting the data, element by element, on maps and generating graphical plots of the data and summary statistical tables as described below. Where there are multiple populations, the data set will need to be divided up according to population and background ranges estimated for each.

Inspecting Data on Maps
The geochemical data for the area of interest should be viewed element by element as maps. If location data for sample sites are available, three 'eda' mapping functions in 'rgr' are available to display simple spatial plots for data inspections. It should be noted, however, that this function does not replace the use

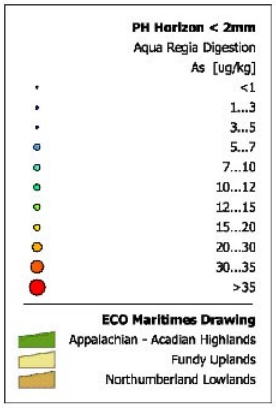



***Figure 4***. *Arsenic in the public health interval (0-5 cm) of soils, based on analysis of the <2 mm fraction using an aqua regia variant (USA-EPA 3050B) digestion and the ICP-MS technique.*

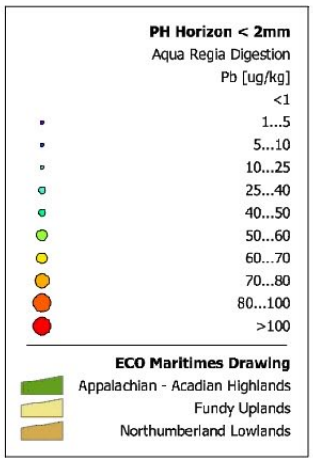
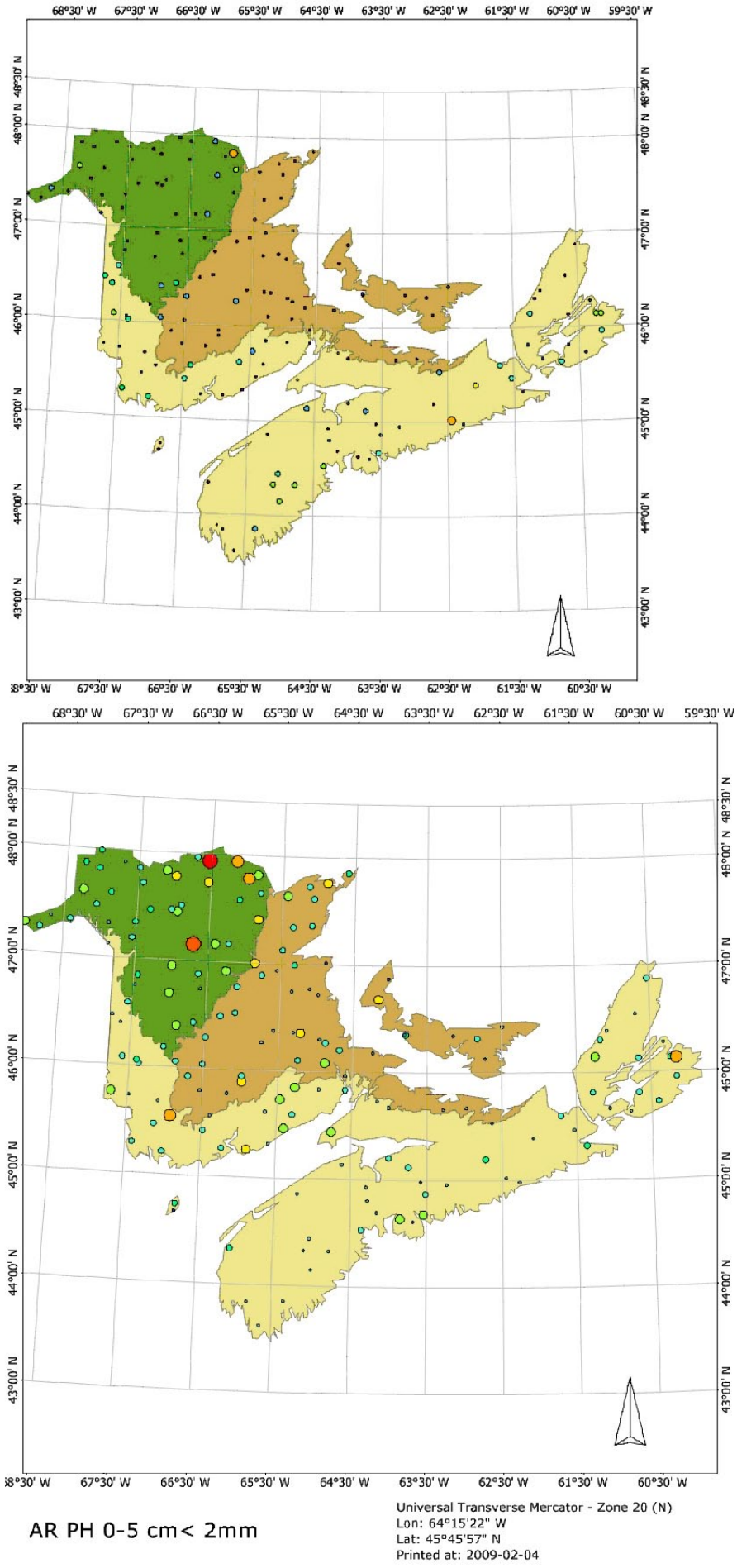
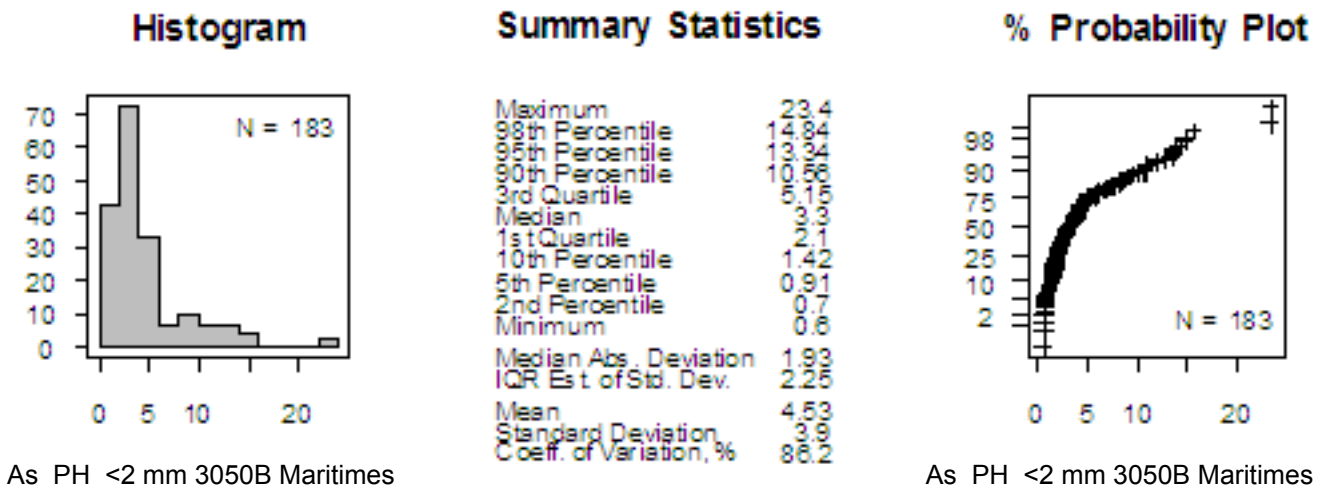


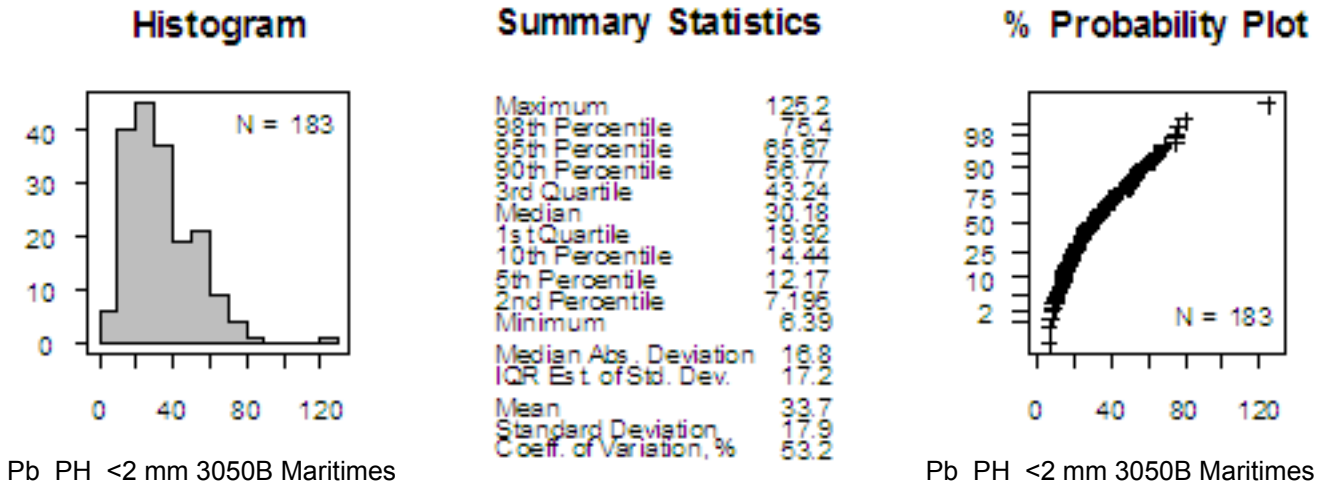

***Figure 5***. *Pb in the 'public health' interval (0-5 cm) of soils, based on analysis of the <2 mm fraction using an aqua regia variant (USA-EPA 3050B) digestion and the ICP-MS technique.*






As PH <2 mm 3050B Maritimes

As PH <2 mm 3050B Maritimes




Pb PH <2 mm 3050B Maritimes

Pb PH <2 mm 3050B Maritimes

***Figure 6***. *Summary statistics, histograms, and cumulative probability plots for As and Pb in the <2 mm fraction of the 'public health' interval (0-5 cm) of soils from the Maritime provinces. Samples were analyzed using an aqua regia variant (USA-EPA 3050B) and the ICP-MS technique.*

of GIS software packages. Maps, based on geochemical data, plotted using proportional or colour-coded symbols are shown below were taken from Garrett and Grunsky (2011, this release) and were plotted using Arcview (Figures 4 and 5). The maps show the distribution of As and Pb in the 'public health' interval (0-5 cm) of the Maritime soil samples. In some cases, locally coherent patterns unrelated to a contamination source (cf. a mineral occurrence/deposit) may be clearly visible on the maps. When this is the case, the data should be split into two or more groups and the cause of the spatial differences determined.

Generating summary statistical tables

It is useful to generate summary statistical tables to help estimate the range of background variation. The statistical tables include listings of the minimum, maximum, mean, median, 1st quartile, and 3rd quartiles

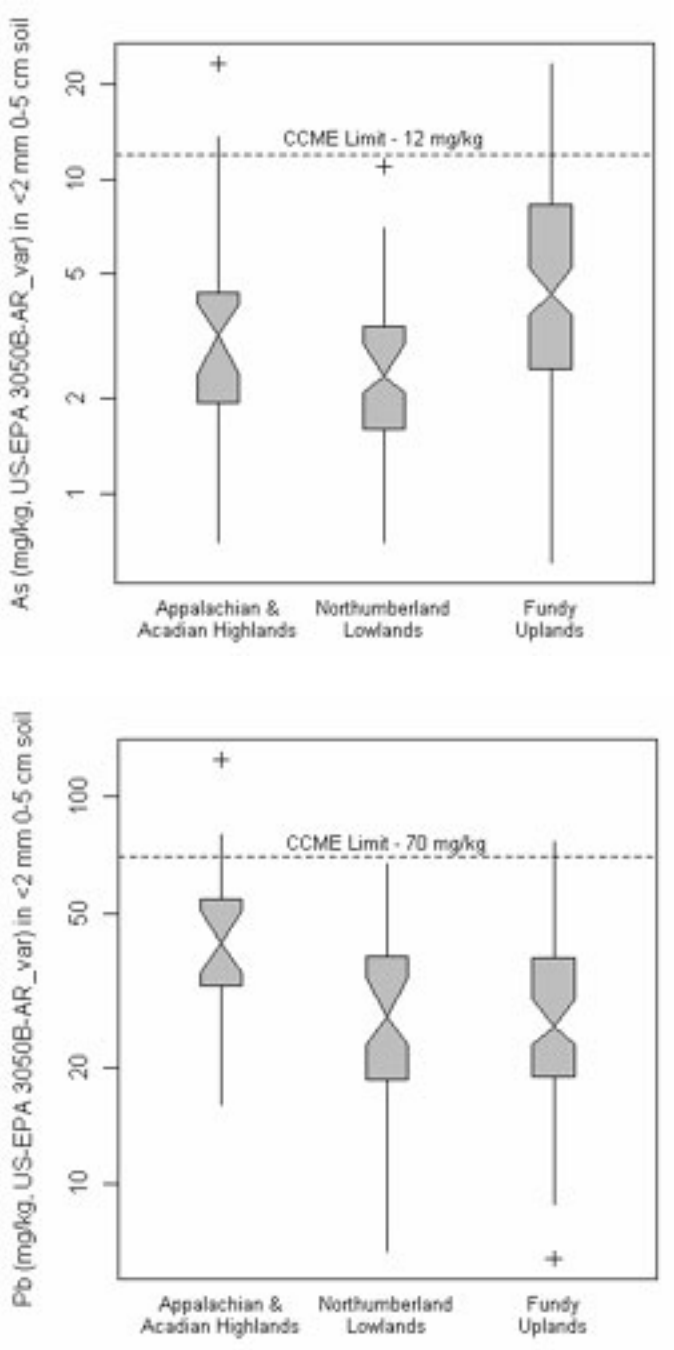



***Figure 7***. *Tukey box plots showing the distribution of As and Pb in the 'public health' interval (0-5 cm) of Maritime soils by ecoprovince.*

*Also shown are the soil quality guideline values for As and Pb set by the CCME (Canadian Council of Ministers of the Environment, 2001).*

and the following measures of dispersion – standard deviation, median absolute deviation (MAD), and the coefficient of variation (CV). Shown in Figure 5 are such tables for As and Pb for samples from the PH interval. These were generated using the 'inset' function in 'rgr'. The output appears in a second window entitled [R2 Graphic: Device 2 (ACTIVE)]. The R scripts with 'rgr' workspace to generate the plots are shown in Appendix 02-04

Inspecting data using statistical graphics displays

Data should also be viewed using statistical graphics displays (histograms, box plots, density plots, and Q-Q plots). Histograms, for example, may indicate the presence of more than one population in the data. There are functions in 'rgr' to accomplish these tasks (see list at the end of Appendix 02-04). Histograms

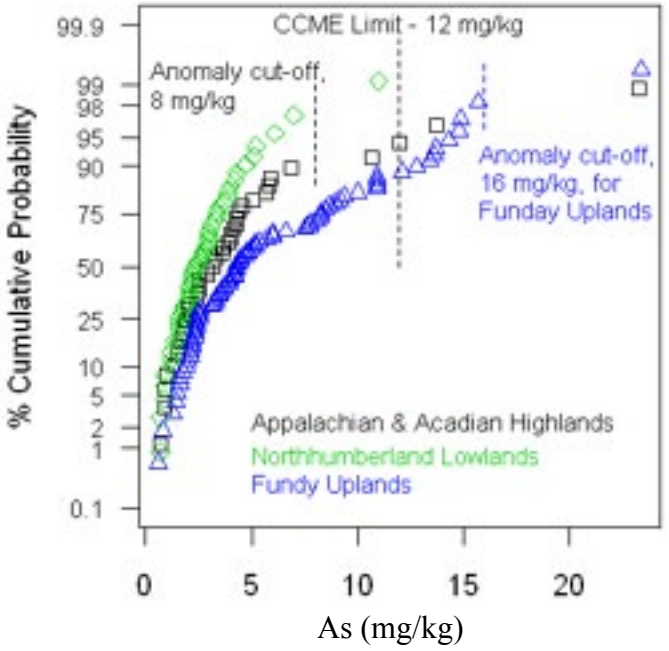



***Figure 8.*** *Q-Q plot of As in samples collected from the 'public health' interval (0-5 cm) of soils in 3 ecoprovinces of the Maritime provinces. Also shown are the cut-off point for data to be used to estimate geochemical background and the upper limit from the CCME guidelines (Canadian Council of Ministers of the Environment, 2001).*

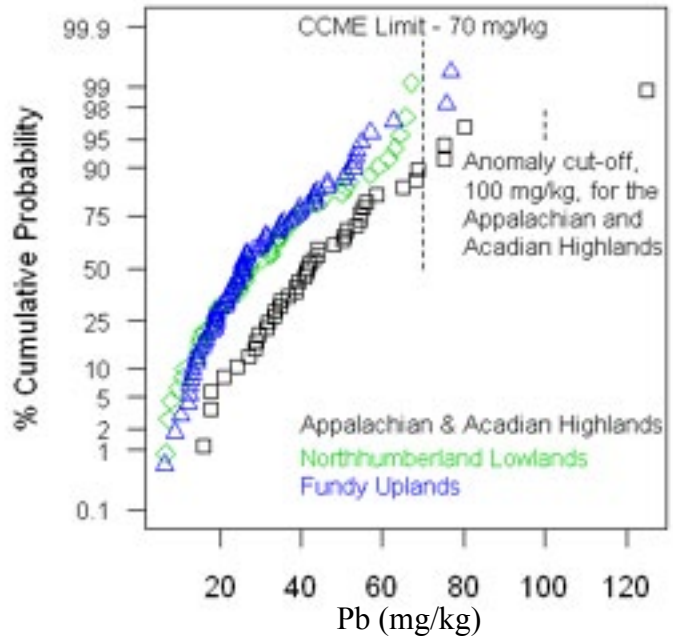



***Figure 9.*** *Q-Q plot of Pb in samples collected from the 'public health' interval (0-5 cm) of soils in 3 ecoprovinces of the Maritime porvinces. Also shown are the cut-off point for data to be used to estimate geochemical background and the upper limit from the CCME guidelines (Canadian Council of Ministers of the Environment, 2001).*

**Table 1.** Calculated estimates of geochemical background As and Pb for the 'public health' interval (0-5 cm) of soil samples collected in the Maritime provinces.

*As (mg/kg) Appalachian and Acadian Highlands*

| As_PH | N = 43 | | | 2%ile = 0.868 | | 98%ile = 15.2 | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Median | MAD | | Mean±2SD | Med±2MAD | Tukey Fences (actual) |
| | 4.16 | 4.08 | 3.2 | 1.78 | + | 12.3 | 6.76 | 7.95 ( 6.9 ) |
| | | | | | - | -4.01 | -0.358 | -1.65 ( 0.7 ) |
| Log10 | 0.492 | 0.321 | 0.505 | 0.3 | + | 13.6 | 12.7 | 14.5 ( 13.7 ) |
| | | | | | - | 0.708 | 0.803 | 0.585 ( 0.7 ) |

*As (mg/kg) Northumberland Lowlands*

| As_PH | N = 56 | | | 2%ile = 0.72 | | 98%ile = 6.91 | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Median | MAD | | Mean±2SD | Med±2MAD | Tukey Fences (actual) |
| | 2.79 | 1.74 | 2.35 | 1.11 | + | 6.27 | 4.57 | 6.1 ( 5.2 ) |
| | | | | | - | -0.696 | 0.126 | -1.1 ( 0.7 ) |
| Log10 | 0.377 | 0.245 | 0.371 | 0.243 | + | 7.38 | 7.18 | 10.5 ( 7 ) |
| | | | | | - | 0.77 | 0.769 | 0.517 ( 0.7 ) |

*As (mg/kg) Fundy Uplands*

| As_PH | N = 84 | | | 2%ile = 1.13 | | 98%ile = 15.2 | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Median | MAD | | Mean±2SD | Med±2MAD | Tukey Fences (actual) |
| | 5.87 | 4.36 | 4.3 | 3.11 | + | 14.6 | 10.5 | 17 ( 15.7 ) |
| | | | | | - | -2.85 | -1.93 | -6.26 ( 0.6 ) |
| Log10 | 0.656 | 0.323 | 0.633 | 0.393 | + | 20.1 | 26.3 | 51 ( 23.4 ) |
| | | | | | - | 1.02 | 0.703 | 0.403 ( 0.6 ) |

*Pb (mg/kg) Appalachian and Acadian Highlands*

| Pb_PH | N = 43 | | | 2%ile = 17.5 | | 98%ile = 87.4 | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Median | MAD | | Mean±2SD | Med±2MAD | Tukey Fences (actual) |
| | 45.4 | 20.3 | 41.7 | 17.7 | + | 85.9 | 77.2 | 87.2 ( 80.2 ) |
| | | | | | - | 4.88 | 6.26 | -0.152 ( 15.9 ) |
| Log10 | 1.62 | 0.186 | 1.62 | 0.172 | + | 97.8 | 92.1 | 118 ( 80.2 ) |
| | | | | | - | 17.6 | 18.9 | 15.1 ( 15.9 ) |

*Pb (mg/kg) Northumberland Lowlands*

| Pb_PH | N = 56 | | | 2%ile = 9.82 | | 98%ile = 67.2 | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Median | MAD | | Mean±2SD | Med±2MAD | Tukey Fences (actual) |
| | 30.9 | 16.4 | 27.1 | 15.2 | + | 63.7 | 57.5 | 69.2 ( 67 ) |
| | | | | | - | -1.9 | -3.22 | -11.8 ( 6.67 ) |
| Log10 | 1.42 | 0.252 | 1.43 | 0.242 | + | 84.8 | 82.6 | 117 ( 67 ) |
| | | | | | - | 8.31 | 8.92 | 6.15 ( 6.67 ) |

*Pb (mg/kg) Fundy Uplands*

| Pb_PH | N = 84 | | | 2%ile = 9.82 | | 98%ile = 67.2 | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Median | MAD | | Mean±2SD | Med±2MAD | Tukey Fences (actual) |
| | 29.5 | 14.9 | 25.5 | 13.2 | + | 59.3 | 51.9 | 67.8 ( 62.9 ) |
| | | | | | - | -0.331 | -0.9 | -10.5 ( 6.39 ) |
| Log10 | 1.42 | 0.219 | 1.41 | 0.212 | + | 71.4 | 67.7 | 112 ( 76.9 ) |
| | | | | | - | 9.55 | 9.59 | 6.47 ( 8.81 ) |

Note: The column with the + and - indicate that on that row the values are for Median+2MAD, Mean+2SD, etc. and the - row is for Median-2MAD, etc.

and cumulative probability plots are shown in Figure 6. In the case of the Maritime sampling project, the dataset was divided, based on the distribution of samples in the three ecoprovinces in the Maritimes - Appalachian and Acadian Highlands, Northumberland Lowlands, and Fundy Uplands (see Kettles, 2011 (this release)). The data subsets are composed of 43, 56 and 84 samples, respectively. Tukey box plots (Tukey, 1977) and larger-scale cumulative probability plots for the Maritime data are shown in Figures 7, 8, and 9. These were generated using the 'tbplot' and 'gx.cnpplts' functions from the 'rgr' package , discussed above. The scripts used are presented in Appendix 02-05. There are clear differences between the statistical distributions for data from the three ecoprovinces. Note that cumulative probability plots, based on the whole data set, as shown on Figure 6 with the statistical tables, do not clearly indicate the presence of multiple populations. This is not uncommon for mixtures of data with similar means.

*Step 2 - Identification and removal of outliers in the data set*

The data set or subsets should be inspected for outliers. Outliers are individual data points that do not appear to "belong" to the data population. There is a need for visual inspection via probability (Q-Q) plots. Outliers may also be identified through calculation.. Methods used for calculation are as follows: (1) mean ± 2 standard deviations (SD); (2) median ± 2 median absolute deviations (MAD); and (3) Tukey Boxplot 'normal' range (Tukey, 1977). If a calculation is required, Reimann et al. (2005) recommend procedure (2), the use of medians and MADs. These provide upper, and lower, limits of the expected range of the data based on 'normal law'. Once the cut-off values are set, outliers, if present, should be removed from the data set or subset.

**Table 2** Estimates of background based on statistical analyses of remaining geochemical data obtained using a hybrid procedure.

| As - mg/kg | N | Min | Max | Cut-off | N | Min | 98th %ile | Max |
|---|---|---|---|---|---|---|---|---|
| Appalachian and Acadian Highlands | 43 | 0.7 | 23.3 | 8 | 39 | 0.7 | 6.1 | 6.9 |
| Northumberland Lowlands | 56 | 0.7 | 11 | 8 | 55 | 0.7 | 6 | 7 |
| Acadian Uplands | 84 | 0.6 | 23.4 | 16 | 83 | 0.6 | 14.8 | 15.7 |
| **Pb - mg/kg** | | | | | | | | |
| Appalachian and Acadian Highlands | 43 | 15.9 | 125.2 | 100 | 42 | 15.9 | 76.2 | 80.2 |
| Northumberland Lowlands | 56 | 6.7 | 67 | 70 | 56 | 6.7 | 66 | 67 |
| Acadian Uplands | 84 | 6.4 | 76.9 | 70 | 82 | 6.4 | 55.7 | 62.8 |

Estimates of background range for As in the Acadian Uplands are 0.6 to 15 mg/kg, and for Pb in the Appalachian and Acadian Highlands are 15 to 76 mg/kg

Figures 8 and 9 show visually, based on Q-Q plots, the identification of outliers and selection of cut-off values for As and Pb concentrations in the 0-5 cm interval of the Maritime soil samples. The R and 'rgr' workspace used to generate these plots is shown in Appendix 02-06. Some calculated estimates for As and Pb based on background ranges for the Maritime dataset are shown in Table 1. These estimates were obtained using the 'fences' function from the 'rgr' package and, as noted previously, the different procedures yield different estimates for the range of background. The workspace used is presented in Appendix 02-07. Background range estimates may also be made using a hybrid procedure involving graphical tools to identify and eliminate outliers, see Figures 8 and 9, and then estimating the percentiles of the remaining background data.

***Step 3 - Estimating background***

Once it is established that the data represent one population and that there are no outliers, an estimate of geochemical background can be made. Using the remaining data in the set or subset, calculate the percentiles for individual elements. Results for As and Pb using the remaining data following outlier elimination via Figures 8 and 9 are shown in Table 2. Before undertaking these calculations, however, ensure that there are a sufficient number of samples to make a valid statistical comparison. As a rule of thumb, data subsets should have a minimum size of 30, but further information is available from Reimann et al., (2005) and Reimann and Garrett (2005). Select from the percentiles, or even the minimum and maximum (if appropriate), a probable background range of concentrations for individual elements. For example, the 2nd and 98th percentiles are Ontario's $OTR_{98}$ estimates, and, for mineral exploration purposes, the background range is sometimes based on the 5th and 95th percentiles. The choice of percentile range is set depending on the degree of caution required by the risk assessment or remediation activity. It is also possible to estimate a range of background values based on different themes or criteria, for example a specific lithologic group, surficial materials unit, or ecosystem classification unit.

## 3.2 Multivariate Methods

Other types of analyses may be employed for value-added interpretation (Filzmoser et al., 2005; Reimann et al., 2008; Garrett and Grunsky, 2011, this release). An extensive discussion of these methods is beyond the scope of this article so they are only mentioned here briefly. For example, by analyzing and interpreting the occurrence of groups of elements in soil samples, it may be possible to identify and discriminate between natural and anthropogenic sources.

Where background ranges for several elements are to be determined the univariate procedure can lead to situations where, for different elements, different samples and their data are removed. This is not ideal.

A multivariate equivalent of the probability plot exists, the chi-square plot, where all elements of interest are investigated simultaneously This procedure requires special transformations of the data as geochemical data are a 'closed number system', they sum to a constant, 100%, 1,000,000 mg/kg, etc. Using this approach, a background data set can be prepared and range estimates made for the various elements of interest determined univariately as appropriate. Principal Component Analysis, with a similar caveat re data transformation, may be used to identify outliers and gain understanding of inter-element relationships (Reimann et al., 2008; Grunsky, 2010).

### 4. AVAILABILITY OF GEOCHEMICAL DATA

A catalogue was developed and contains searchable metadata and raw data for approximately 700 geochemical surveys carried out by the GSC and provincial geological agencies since the 1950s (Spirito et al., 2007). It can be accessed, along with many other types of data, in the Geochemical Data Repository (GDR) at Natural Resources Canada (http://gdr.nrcan.gc.ca/index_e.php [accessed March 11, 2010]). Information on the geochemical data catalogue in the GDR is available in Kettles et al. (2011 (this release)).

Digital and user friendly databases, catalogues, and free downloads are available on-line from federal and provincial library facilities (e.g. GSC Bookstore; on-line at http://gsc.nrcan.gc.ca/bookstore/index_e.php [accessed March 11, 2010]).

## REFERENCES

Canadian Council of Ministers of the Environment, 2001. Canadian soil quality guidelines for the protection of environmental and human health: Arsenic (inorganic) (1997). Updated In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Dodd, M., 2011. Metal bioaccessibility in Tri-National Survey soil samples; *in* Proceedings and recommendations from the Workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

Ecological Stratification Working Group. 1996. A National Ecological Framework for Canada. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research and Environment Canada, State of Environment Directorate, Ottawa/Hull. 125 p. Map at scale 1:7.5 million. http://www.ec.gc.ca/soer-ree/English/Framework/framework.cfm. Pdf copy available from /cansis/publications/ecostrat/intro.html [accessed April 2010]

Filzmoser, P., Garrett, R.G. and Reimann, C., 2005. Multivariate outlier detection in exploration geochemistry. Computers & Geosciences, v. 31, p. 579-587.

Ford, K.L., DiLabio, R.N.W., Rencz, A.N. (1988). Geological, geophysical and geochemical studies around the Allan Lake carbonatite, Algonquin Park, Ontario. Journal of Geochemical Exploration, v. 30, p. 99-121.

Friske, P.W.B. and Hornbrook, E.H.W., 1991. Canada's National Geochemical Reconnaissance programme; Source Institution of Mining and Metallurgy, Transactions, Section B: Applied Earth Sciences v. 100, p. B47-B56.

Friske, P W B; Ford, K L; Kettles, I M; McCurdy, M W; McNeil, R J; Harvey, B A., 2010. North American Soil Geochemical Landscapes Project: Canadian field protocols for collecting mineral soils and measuring soil gas radon and natural radioactivity; Geological Survey of Canada, Open File 6282; 177 p.

Friske, P.W.B., Ford, K.L., and McNeil, R.J., 2011. Geochemical and mineralogical soil data from the 2007 North American Soil Geochemical Landscapes Project in New Brunswick, Nova Scotia and Prince Edward Island. Geological Survey of Canada, Open File 6433.

Garrett, R.G., 1983. Sampling methodology. Chapter 4 of Handbook of Exploration Geochemistry, Vol. 2, Statistics and Data Analysis in Geochemical Prospecting, (ed.) R.J. Howarth, Elsevier, p. 83-110.

Garrett, R.G., 2005. Natural distribution and abundance of elements. Chapter 7 of The Essentials of Medical Geology (Ed. in Chief O. Selinus). Elsevier Academic Press, Amsterdam, The Netherlands, p. 17-41.

Garrett, R.G., 2010. How many samples are enough. Association of Applied Geochemists, Newsletter, Explore, v. 149, p. 9-14.

Garrett, R.G. and Chen, Y., 2007. rgr: The GSC (Geological Survey of Canada) Applied Geochemistry EDA Package - R tools for determining background ranges and thresholds; Geological Survey of Canada, Open File 5583, 2007; 1 CD-ROM

Garrett, R.G. and Grunsky, E.C., 2011. Geochemical background - What it is and how it varies; *in* Proceedings and recommendations from the Workshop on the role of geochemical data in environmental

and human health risk assessment, Halifax, 2010, (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

Garrett, R.G. and Kettles, I.M., 2009. North American Soil Geochemical Landscapes Project (NASGLP): Database of GRTS sample sites with notes on the sampling design and site selection procedure. Geological Survey of Canada, Open File 6300. CD-ROM..

Garrett, R.G., Hall, G.E.M., Vaive, J.E., and Pelchat, P., 2009a. A water-leach procedure for estimating bioaccessibility of elements in soils from transects across the United States and Canada: Applied Geochemistry, v. 24, p. 1438-1453.

Garrett, R.G., Grunsky, E.C., Friske, P.W.B., and McCurdy, M.W., 2009b. To Aqua Regia or not… that is the question; *in* (ed.) Rencz, A.N. and Kettles, I.M., 2009; North American Soil Geochemical Landscapes Project (NASGLP): Proceedings of Workshop II, Ottawa, 2008; Geological Survey of Canada, Open File 6209, CD-ROM

Girard, I., Klassen, R.A. and Laframboise, R.R., 2004. Sedimentology laboratory manual, Terrain Sciences Division; Geological Survey of Canada, Open File 4823, 134 p.

Grunsky, E.C., 2010. The interpretation of geochemical survey data. Geochemistry, Exploration, Environment and Analysis Geochemistry, v. 10, p. 27-74.

Grunsky, E.C. and McNeil, R.J., 2011. Spatial variation - Horizontal and vertical; *in* Proceedings and recommendations from the workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

Hawkes, H.E., and Webb, J.S., 1962. Geochemistry in Mineral Exploration. Harper and Rowe, New York, 415 p.

Horowitz, A.J., 1991. A primer on sediment-trace element chemistry (second edition), Lewis Publishers, 136 p.

Kettles, I.M., 2011. Note on eco-classification systems for Canada; *in* Proceedings and recommendations from the Workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, in (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

Kettles, I.M., Spirito, W.A. and Rencz, A.N., 2011. Availability of geochemical data via the Geoscience Data Repository; *in* Proceedings and recommendations from the workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

Klassen, R.A., 2011. Geoscience in ecological and human health risk assessment; *in* Proceedings and recommendations from the Workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

McNeil, R.J., 2011. Field sampling protocols; *in* Proceedings and recommendations from the Workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

McNeil, R.J. and Garrett, R.G., 2011. Soil sample preparation, analytical techniques and QA/QC; *in* Proceedings and recommendations from the Workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

Parsons, M.B., 2011. Analytical methods used to characterize the solid phase speciation of metal(loid)s. *in* Proceedings and recommendations from the Workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, (ed.) A.N. Rencz and I.M. Kettles; Geological Survey of Canada, Open File 6645, CD-ROM.

Reimann, C. and Garrett, R.G., 2005. Geochemical background - Concept and reality. Science of the Total Environment, v. 350, p. 12-27.

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

Reimann, C., Filzmoser, P., Garrett, R. and Dutter, R., 2008. Statistical data analysis explained: Applied Environmental statistics; R. John Wiley and Sons, New York.

Reimann, C., Matschullat, J., Birlee, M. and Salminen, R., 2009. Arsenic distribution in the environment: the effects of scale. Applied Geochemistry, v. 24, p. 1147-1167.

Reimann, C., Matschullat, J., Birlee, M. and Salminen, R., 2010. Antimony in the environment: Lessons from geochemical mapping. Applied Geochemistry, v. 25, 175-198.

Soil Landscapes of Canada Working Group, 1996. Soil Landscapes of Canada v. 2.2; Agriculture and Agri-Food Canada. (digital map and database at 1:1 million scale).

Sheldrick, B.H., 1984. Chapter 3 Physical analysis, 3.1 Particle size distribution; in Analytical Methods Manual 1984, (ed.) B.H. Sheldrick; Land Resource Research Institute, Research Branch, Agriculture Canada, LRRI Contribution no 84-3-, Ottawa, Ontario; p. 26/1-27/4.

Shilts, W.W., 1975. Principles of geochemical exploration for sulphide deposits using shallow samples of glacial drift; Canadian Institute of Mining and Metallurgy, Bulletin 68, p. 73-80.

Shilts, W.W., 1984. Till geochemistry in Finland and Canada; Journal of Geochemical Exploration, v. 21, p. 95-117.

Spirito, W. A., Rencz, A.N, Kettles, I.M., Adcock, S.W., and Stacey, AP., 2007. Compilation of soil and till geochemical metadata for Canada - an update; Geological Survey of Canada 5085.

Thompson, M. and Howarth, R.J. 1978: A new approach to the estimate of analytical precision; Journal of Geochemical Exploration, v. 9, p. 23-30.

Tukey, J., 1977. Exploratory Data Analysis. Addison Wegley, Reading, MA.

Wheeler, J.O., Hoffman, P.F., Card, K.D., Davidson, A., Sanford, B.V., Okulitch, A.V., and Roest, W.R. (comp.), 1997: Geological Map of Canada, Geological Survey of Canada, Map D1860A.

Woodruff, L.G., 2011. North American Soil Geochemical Landscapes - 2010 USGS Perspective; *in* Presentations and recommendations from the workshop on the role of geochemical data in environmental and human health risk assessment, Halifax, 2010, A.N. Rencz and I.M. Kettles (ed.); Geological Survey of Canada, Open File 6645, CD-ROM.

## Appendix 02-01. File ‘rgr.first’ needed to run ‘rgr’package

```
“rgr.first” <-
        function (wd = NULL)
        {
        if(is.null(wd)) wd = “D:\\R\\WD”
        setwd(wd)
        library(rgr)
        par(pty = “s”, pch = 3)
        cat(“Default plot shape set to ‘square’ and plot symbol to ‘plus’\n”)
        cat(“Working directory set to:”, wd, “\n”)
         options(warn = -1)
        cat(“R options set to warn = -1 to suppress unwanted graphics related messages from\n
                rgr multi-panel functions and functions calling eqscplot in Library MASS\n”)
        }
```

## Appendix 02-02. Example of "getting started" workspace installing R and 'rgr' and attaching geochemical data set

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> ls()
character(0)
> fix(rgr.first)
> ls()
[1] "rgr.first"
> ls()
[1] "gx.cnpplts" "rgr.first"
> utils:::menuInstallLocal()
package 'akima' successfully unpacked and MD5 sums checked
> utils:::menuInstallLocal()
package 'MASS' successfully unpacked and MD5 sums checked
> library(rgr)
Loading required package: akima
Loading required package: MASS
Warning messages:
1: package 'rgr' was built under R version 2.8.0 and help will not work correctly
Please re-install it
2: package 'akima' was built under R version 2.11.0
3: package 'MASS' was built under R version 2.11.1
> utils:::menuInstallLocal()
Warning: package 'rgr' is in use and will not be installed
> help(rgr)
starting httpd help server ... done
```

```
> ?bwplot
> pb.as.2mm.3050B.halifax<-read.table("D:\\R\\As&Pb4Halifax.4R",header=T,sep="","")
> attach(pb.as.2mm.3050B.halifax)
> names(pb.as.2mm.3050B.halifax)
 [1] "ID"                "Lat"               "Lng"
 [4] "Prov"              "EcoP"              "Ecoprovince"
 [7] "EcoR"              "Ecoregion"         "Soil.Type_SLC"
[10] "SubRx.Type..Wheeler." "As_PH"             "Pb_PH"
[13] "Corg_PH"           "As_Ahor"           "Pb_Ahor"
[16] "Corg_Ahor"         "As_Bhor"           "Pb_Bhor"
[19] "Corg_B.hor"        "As_Chor"           "Pb_Chor"
[22] "Corg_Chor"
> shape(As_PH)
  1 row(s) with missing value(s), NA(s), removed from vector
Warning messages:
1: In par(oldpar) : graphical parameter "cin" cannot be set
2: In par(oldpar) : graphical parameter "cra" cannot be set
3: In par(oldpar) : graphical parameter "csi" cannot be set
4: In par(oldpar) : graphical parameter "cxy" cannot be set
5: In par(oldpar) : graphical parameter "din" cannot be set
>
```

Now we are set up properly:

```
Working directory set to: D:\R\WD
R options set to warn = -1 to suppress unwanted graphics related messages from

        rgr multi-panel functions and functions calling eqscplot in Library MASS
Warning messages:
1: package 'rgr' was built under R version 2.8.0 and help will not work correctly
Please re-install it
2: package 'akima' was built under R version 2.11.0
3: package 'MASS' was built under R version 2.11.1
> ls()
[1] "as.pb.2mm.3050B.halifax" "gx.cnpplts"
[3] "rgr.first"
> attach(as.pb.2mm.3050B.halifax)
> names(as.pb.2mm.3050B.halifax)
 [1] "ID"                "Lat"               "Lng"
 [4] "Prov"              "EcoP"              "Ecoprovince"
 [7] "EcoR"              "Ecoregion"         "Soil.Type_SLC"
[10] "SubRx.Type..Wheeler." "As_PH"             "Pb_PH"
[13] "Corg_PH"           "As_Ahor"           "Pb_Ahor"
[16] "Corg_Ahor"         "As_Bhor"           "Pb_Bhor"
[19] "Corg_B.hor"        "As_Chor"           "Pb_Chor"
[22] "Corg_Chor"
> quit()
```

# Appendix 02-03. Overview of the 'rgr' package and functions

Over 100 'rg' functions have been written in the last sixteen years at the Geological Survey of Canada (GSC) for the S-Plus proprietary statistical software to support exploration and applied geochemical survey and research activities. Most of these function scripts have been written from 'scratch', however, others are based on scripts shared within the S user community on S-News. The 'rgr' functions are a subset of the 'rg' functions that run under the R system. R is the open source version of the S language, it is extensively used in academia and other institutions, see http://www.r-project.org/ . R may be downloaded from any of the CRAN sites listed there. Version 'rgr_1.0.5'comprises 67 of the 'rg' functions; most of these are exploratory data analysis (EDA) and data inspection tools, many of which rely on graphics as a way to communicate the nature of the data sets under investigation/interpretation. All future development is being carried the R environment. While the functions in 'rgr_1.0.5' are essentially univariate, with exception of the computation of log-ratio transformations and the calculation of correlation matrices and weighted sums, future releases will contain multivariate tools for the graphical display of data.

The 'rgr' functions have been prepared so that they may be used by GSC staff, and be distributed externally to other government departments (OGDs), agencies and individuals wishing to use the Applied Geochemistry Section's graphical and other procedures. In some instances the GSC has undertaken specific geochemical compilations for OGDs of National Geochemical Reconnaiss-ance and other geochemical survey data sets held by the GSC. Many of these are publically available as GSC Open Files. The 'rgr' functions may be used to process applied geochemical data to generate summary statistics, both numeric and graphical, in support of the estimation of the boundaries of ambient and natural geochemical background variations. Implicitly this includes the setting of threshold or action levels that may trigger further field activities to determine if the causes of outliers, i.e. observations with above threshold or action levels, are due to natural or anthropogenic causes. The 'rgr' functions described in this document fall, dominantly, under the title of univariate exploratory data analysis tools. If observation site coordinates are available, four functions are available to display simple spatial plots, 'maps', for data inspection (note: these do not replace the use of a Graphical Information System, GIS, for spatial data display and analysis). A further spatial function displays a concentration-area plot that assists in determining if multi-fractal patterns are present in the data that can be used to identify boundaries between data populations related to different spatial - fractal - processes; e.g., background and anomalous.

The following notes are for those as yet unfamiliar with the use of R.

Although the help files use TRUE and FALSE in the examples, etc., the capital letters T and F, respectively, may be used when running the functions..

Quotes, " ", are used to enclose character strings that will be printed or displayed. To obtain a Greek µ, use \265, thus to display (µg/kg) use "... (\265g/kg) ...", if a tab is required use \t, and \n forces a new line. The available codes for special characters such as the Greek µ may be displayed with function display.ascii.o.

Where justification of text is an option, adj = 0 results in left justification, adj =1 in right justification, and adj = 0.5 in centring. Defaults are always provided.

NA is an explicit way that the S language, and R, has of conveying the fact that there is no information. A blank numeric field, i.e. ‘, ,’(note the space between the commas) in a table entered into R by the ‘read. table’ command is converted to a NA, for an actual value of zero a zero has to be explicitly entered. In some geochemical data files blanks have been converted to zeros by other software packages, in others they are set to a coded value, e.g. -9999. Tools are available in functions ‘ltdl.fix.df’ and ‘ltdl.fix’ to set these zeros or coded values to NAs if that is appropriate.

Many computational tasks cannot accept NAs, therefore software, function ‘remove.na’, is used internally within the ‘rgr’ functions to remove NAs from data when required.

It is common practice to set geochemical results less than the detection or quantification limit (<dl) to the negative value of the detection limit. Tools, functions ‘ltdl.fix.df’ and ‘ltdl.fix’ for data frames and vectors, respectively, are provided to convert these negative values to half the positive value of the detection limit. This is essential if logarithmic scaling is to be used in plots, or calculations are to be undertaken in logarithms.

If logarithmic scaling is required for plots or computations ensure that the parameter log = T and, if required, logx = T. In some functions where two variables are plotted against each other, log = “x”, log = “y” and log = “xy”, control scaling of the axes.

A common construct for storing data in S and R is the dataframe, this includes not only the data but also the variable names (columns) and the observation identifiers (rows). The latter commonly known as the sample numbers or IDs. In this there is a difference between natural scientists and statisticians, to a natural scientist a sample is an individual ‘something’ that is collected, described and measured, whereas to a statistician the sample is the whole collection of individual ‘somethings’ and has some size N. The data for any one variable or measurement is a column vector. The data in a dataframe may be made easily accessible by attaching the dataframe with ‘attach(dfname)’. The attached dataframe may be removed by ‘detach(dfname)’. The function ‘df.test(dfname)’ may be run to see if a dataframe is attached or present in the R work space and identify the names of the variables it contains.

The construct ‘deparse(substitute(x))’ is used to generate a default variable name label, the column variable name, if no more informative text is provided by the user, e.g., Cu (a column variable name) rather than a user defined text string like “Cu (mg/kg) in surface soil”. The contents of the variable name label are variously defined as xlab, xname, or name, or a similar parameter name for a y or z variable, where required.
A construct useful in the execution time selection of a subset of the values for a variable is conditioning. While Cu leads to the processing of the entire vector of data for the column variable Cu, Cu[Cu<200] would result in the processing of only those data where the Cu value was <200. Similarly, Cu[Cu>10 & Cu<200] would result in the processing of only those data with values between, exclusively, 10 and 200. The condition may be based on the values of any variable available in the data frame, thus Cu[Zn>200] would result in only those Cu values where the Zn value exceeded 200 being processed. Similarly for a ‘factor’ (text string variable), Cu[PM ==”Till”] (note the double = signs) results in only those Cu values where PM (the soil parent material) was recorded as Till being processed. For generating more permanent subsets from a dataframe or matrix the R function ‘subset’ or the rgr package function ‘gx. subset’ may be used.

In some cases it is required to split the data for a variable, a column vector, into subsets based on the value of some classificatory variable, factor, which appears as a column variable. For example, split(Cu, GSG) would split the data for the variable Cu into subsets on the basis of the values of GSG (Great Soil Group). The values of the criterion may be either character strings or integer numbers, there will be as many subsets as there are unique values of the criterion (factor). This technique may be used with functions 'bwplot' and 'tbplot'. The value of the criterion also may be computed, for example, Distance%/%10 generates a truncated (integer) value of the Distance from a fixed point divided by 10. Thus all Distances between 0 and 9.99... have a value of 0, those from 10 to 19.99... have a value of 1, and so on. Thus if Distance is the distance from a point source of contaminants, e.g., a smelter stack, a Tukey boxplot display can be generated where the individual plots graphically summarize the data in 10 km units from the source.

Where options exist for the colour infill of polygons the default is grey, 'colr = 8'. The following are the available default R colours: 1 = black; 2 = red; 3 = green; 4 = dark blue; 5= light blue; 6 = purple; 7 = yellow; and 8 = grey. Setting colr = 0 results in no infill. To display the actual colours use function 'display.lty()' that also displays line styles.

Some users may find it convenient to use a 'first' function. Such a function can be used to load the rgr package for use along with another required R Library, akima, set certain defaults to the user's preferences, and set the R Working Directory to one appropriate for the data under investigation. In earlier versions of R the MASS library had also to be explicitly made available, currently it comes bundled with the standard R installation. Both MASS and akima are 'lazy loaded' by loading 'rgr', they just have to be in the library folder in the user's PC. The following is an example:

```
"first" <-
function (wd = NULL)
{
if(is.null(wd)) wd = "D:\\R\\Project 3\\WD"
setwd(wd)
library(rgr)
par(pty = "s", pch = 3)
cat("Default plot shape set to 'square' and plot symbol to 'plus'\n")
cat("Working directory set to:", wd, "\n")
options(warn =  1)
cat("R options set to warn =  1 to suppress unwanted graphics related messages from\n  rgr multi
panel functions and functions calling eqscplot in Library MASS\n")
}
```

The above function assumes that a folder D:\R\Project 3\WD has been set up outside the Program Files for R where data and output files are to be stored, where Project 3, or some other appropriate name, is a subdirectory where data and files are to be kept. In general it is good practice to store the data and project specific files in a different place than the software. If a 'data' drive, e.g., D:\, is not available a subdirectory in My Documents can be used. The above 'first' function reflects the preference of the author for 'square' xy-plots and the use of a '+' sign as the plotting symbol. Another plotting symbol can be chosen from the available plotting marks, see 'display.marks', and if the user consistently wants

rectangular plots that make maximum use of the display space, change pty = “s” to pty = “m”.

To complete the set-up for a project a R icon should be placed on the desk top and edited so that R uses the defined Working Directory for storing the R files: .Rdata and .Rhistory. To place an extra R icon on the desktop for a particular project, go to the C:\Program Files\R\bin subdirectory and make a shortcut to R.exe and then drag it to the desktop, where it can be renamed appropriately, e.g., Project 3. In this manner different R sessions for different projects or data investigations may be set up, with the result that only relevant files are accessible and the R workspace is less cluttered. This is done by right clicking on the R icon placed on the desktop and selecting Properties and editing the Start in field to D:\R\Project3\WD and clicking on Apply.

If the user has set up a folder D:\R\Project3\WD for data and files for Project 3, then clicking on the ‘Project 3' R icon will start the session in the correct WD subdirectory. If a different Working Directory is required entering ‘first(“D:\\R\\Project 4\\WD") at the > in the R session will result in the rgr package being made available, the Working Directory being set to D:\R\Project 4\WD, and any other defaults the user wishes to set being implemented. Note two things: 1) the subdirectory D:\R\Project 4\WD must have been created, and 2) the use of \\ to cause the correct backslash for the file name in the execution of the ‘first’ function.

The last instruction in the ‘first’ function is ‘options(warn = -1)’. The use of multi-panel displays in some rgr functions and the use of function ‘eqscplot’ from the MASS Library in others causes warning messages to be displayed concerning certain graphics parameters. These are not relevant to the user and the ‘options(warn = -1)’ statement leads to their suppression.

**In practice it is important to appropriately handle any -ve values due to the presence of <dl data and any zeros or coded values indicating missing data prior to undertaking any plotting or computations. The easiest way to achieve this is by use of the ‘ltdl.fix.df’ function. Thus, if the data are in a dataframe ‘dfname' the command ‘dfname.fixed <- ltdl.fix.df(dfname)’ is executed. The newly created object ‘dfname.fixed’ is attached, ‘attach(dfname.fixed)’ so that its column vectors are directly accessible. Any resulting NAs are handled appropriately in each ‘rgr’ function using the ‘remove.na’ function.**

The following list describes the functions available to the user, functions that are only called internally are marked with an asterisk (*), together with the test data sets used in the examples. Full details are available through the on-line help files in the rgr Library.

| | |
|---|---|
| alr | Undertake an Arithmetic Log-Ratio transformation |
| anova1 | Duplicate Sample Analysis of Variance (ANOVA) |
| anova2 | Duplicate Sample Analysis of Variance (ANOVA), alternate input |
| bwplot | Plot Vertical Box-and-Whisker Plots |
| bwplot.by.var | Plot Vertical Box-and-Whisker Plots for Variables |
| bxplot | Plot a Horizontal Boxplot or Box-and-Whisker Plot |
| caplot | Prepare a Concentration-Area (C-A) Plot |
| cat2list * | Divides Data into Subsets by Factor |
| clr | Undertake a Centred Log-Ratio Transformation |

| | |
|---|---|
| cnpplt | Displays a Cumulative Normal Percentage Probability (CPP) Plot |
| cutter * | Function to Identify into which Interval a Value Falls |
| df.test | Check for the Existence of a Dataframe |
| display.ascii.o | Display the Windows Latin 1 Font Octal Table |
| display.lty | Display Available Line Styles and Colour Codes |
| display.marks | Display the Available Plotting Symbols |
| display.rainbow | Display the Colours of the Rainbow(36) Pallette |
| fences | Generate and Display Fence Values |
| fences.summary | Generate and Save Fence Values for Data Subsets |
| fix.test | Test Data for Function ltdl.fix.df |
| framework.stats * | Compile Framework/Subset Summary Statistics |
| framework.summary | Generate and Save Framework/Subset Summary Statistics |
| gx.cnpplts | Plot up to nine CPP in a single Display |
| gx.cnpplts.setup | Define Symbology and Colours for use in gx.cnpplts |
| gx.ecdf | Plot an Emprical Cumulative Distribution Function (ECDF) |
| gx.fractile | Estimate the Fractile for a specified Quantile of a Distribution |
| gx.hist | Plot a Histogram |
| gx.hypergeom | Estimate the Probability of Anomaly Location is Informative |
| gx.ks.test | Plot ECDFs of two Distributions with a Kolmogorov-Smirnov Test |
| gx.pearson | Compute Pearson Correlation Coefficients and estimate Probabilities |
| gx.quantile | Estimate the Quantile for a specified Fractile of a Distribution |
| gx.rma | Estimate the Reduced Major Axis Coefficients and Test for (0,1) |
| gx.runs | Carry out a Wald-Wolfowitz, Runs, Test |
| gx.sort | Sort a Dataframe or Matrix on a single variable and Display |
| gx.sort.df | Sort a Dataframe on multiple variables and Display |
| gx.spearman | Compute Spearman Correlation Coefficients and estimate Probabilities |
| gx.stats | Compute Summary Statistics and optionally Display |
| gx.subset | Extract a Subset of Rows from a Data Frame |
| gx.summary * | Compile Summary Statistics for other displays |
| gx.summary1 | Display a Concise Single Line Summary Statistics Report |
| gx.summary2 | Display a ten-line Summary Statistics Report |
| gx.summary.mat | Display gx.summary1 style Reports for Selected Columns of a Dataframe |
| gx.summary.groups | Display gx.summary1 style Reports for Factors in a Dataframe |
| gx.triples.aov | Undertake a Staggered 3-Level Design ANOVA for Sampling and Analytical Variability and Estimate Variance Components |
| gx.triple.fgx | Undertake ANOVAs to estimate the Regional Representivity of Triples |
| ilr | Undertake an Isometric Log-Ratio Transformation |
| inset | An EDA Graphical and Statistical Summary |
| inset.exporter | Saves an EDA Graphical and Statistical Summary |
| kola.c | Kola Project C-horizon Soil Data |
| kola.o | Kola Project O-horizon Soil Data |
| ltdl.fix | Replace Negative Values Representing Less Than Detects for a Vector |
| ltdl.fix.df | Replace Negative Values Representing Less Than Detects for a Dataframe |
| map.eda7 | Display a Symbol Map of Numeric Data Based on the Tukey Boxplot |
| map.eda8 | Display a Symbol Map of Numeric Data Based on their Percentiles |
| map.tags | Display a Map of Posted Values |

| | |
|---|---|
| map.z | Display a Map of Numeric Data using Proportional Symbols |
| ms.data1 | Measurement Variability Test Data |
| ms.data2 | Measurement Variability Test Data |
| ms.data3 | Measurement Variability Test Data |
| remove.na | Remove and Count NAs |
| shape | An EDA Graphical Summary |
| sind | Howarth and Sinding-Larsen Test Data for Weighted Sums |
| syms * | Function to Compute the Diameters of Proportional Symbols |
| syms.pfunc | Function to Demonstrate the Effect of Different Values of p |
| tbplot | Plot Vertical Tukey Boxplots |
| tbplot.by.var | Plot Vertical Tukey Boxplots for Variables |
| thplot1 | Display a Thompson-Howarth Plot of Duplicate Measurements |
| thplot2 | Display a Thompson-Howarth Plot of Duplicate Measurements, alternate input |
| triples.test1 | Test Data for Function gx.triples.aov |
| triples.test2 | Test Data for Function gx.triples.fgx |
| var2fact * | Rearranges Data for Variables as Factors |
| which.na | Function to Identify Row Numbers of a Vector with NAs |
| wtd.sums | Computation of Weighted Sums |
| xyplot.eda7 | Display a 'XY' Plot of Numeric Data Based on the Tukey Boxplot |
| xyplot.eda8 | Display a 'XY' plot of Numeric Data Based on their Percentiles |
| xyplot.tags | Display a 'XY' Plot of Posted Values |
| xyplot.z | Display a 'XY' Plot of Numeric Data using Proportional Symbols |

The 'rgr' Library was first built in 2007 at the Geological Survey of Canada, Ottawa, by Yiwen Chen working in collaboration with the author. Development has continued since then in R.

Robert G. Garrett December 7, 2010
Emeritus Scientist
Geological Survey of Canada
Earth Sciences Sector, Natural Resources Canada
Ottawa, Ontario K1A 0E8

E_mail: garrett@NRCan.gc.ca
Web_site: http://cgc.nrcan.gc.ca/dir/index_e.php?id=4961

# Appendix 02-04. Workspace Using the 'gx.stats'and 'inset' Functions to Generate Summary Statistics with Associated Graphs

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> rgr.first()
Loading required package: akima
Loading required package: MASS
Default plot shape set to 'square' and plot symbol to 'plus'
Working directory set to: D:\R\WD
R options set to warn = -1 to suppress unwanted graphics related messages from

            rgr multi-panel functions and functions calling eqscplot in Library MASS
Warning messages:
1: package 'rgr' was built under R version 2.8.0 and help will not work correctly
Please re-install it
2: package 'akima' was built under R version 2.11.0
3: package 'MASS' was built under R version 2.11.1
> ls()
[1] "as.pb.2mm.3050B.halifax" "gx.cnpplts"            "rgr.first"
> names(as.pb.2mm.3050B.halifax)
 [1] "ID"                "Lat"               "Lng"               "Prov"
 [5] "EcoP"              "Ecoprovince"       "EcoR"              "Ecoregion"
 [9] "Soil.Type_SLC"     "SubRx.Type..Wheeler." "As_PH"          "Pb_PH"
[13] "Corg_PH"           "As_Ahor"           "Pb_Ahor"           "Corg_Ahor"
[17] "As_Bhor"           "Pb_Bhor"           "Corg_B.hor"        "As_Chor"
[21] "Pb_Chor"           "Corg_Chor"
```

> attach (as.pb.2mm.3050B.halifax)
> > gx.stats(As_PH,xlab = “As in <2mm of PH layer with 3050B”, display = TRUE)
1 row(s) with missing value(s), NA(s), removed from vector

Summary Statistics Display for: As in <2mm of PH layer with 3050B

Data Set N = 183
Minimum = 0.6 Maximum = 23.4
Median = 3.3 MAD Est = 1.927
IQR Est = 2.246
Mean = 4.528 S.D. = 3.903
Variance = 15.23 C.V. % = 86.2

Table of Percentiles

| | |
|---|---|
| Maximum Value | 23.4 |
| 99th Percentile | 17.07 |
| 98th Percentile | 14.84 |
| 95th Percentile | 13.34 |
| 90th Percentile | 10.56 |
| 80th Percentile | 6 |
| 3rd Quartile (75th) | 5.15 |
| 70th Percentile | 4.5 |
| 60th Percentile | 3.92 |
| Median (50th) | 3.3 |
| 40th Percentile | 2.58 |
| 30th Percentile | 2.26 |
| 1st Quartile (25th) | 2.1 |
| 20th Percentile | 1.9 |
| 10th Percentile | 1.42 |
| 5th Percentile | 0.91 |
| 2nd Percentile | 0.764 |
| 1st Percentile | 0.7 |
| Minimum Value | 0.6 |

> gx.stats(As_PH,xlab = “Pb in <2mm of PH layer with 3050B”, display = TRUE)
1 row(s) with missing value(s), NA(s), removed from vector

Summary Statistics Display for: Pb in <2mm of PH layer with 3050B

Data Set N = 183
Minimum = 0.6 Maximum = 23.4
Median = 3.3 MAD Est = 1.927
IQR Est = 2.246
Mean = 4.528 S.D. = 3.903
Variance = 15.23 C.V. % = 86.2

Table of Percentiles

| | |
|---|---|
| Maximum Value | 23.4 |
| 99th Percentile | 17.07 |
| 98th Percentile | 14.84 |
| 95th Percentile | 13.34 |
| 90th Percentile | 10.56 |
| 80th Percentile | 6 |
| 3rd Quartile (75th) | 5.15 |
| 70th Percentile | 4.5 |
| 60th Percentile | 3.92 |
| Median (50th) | 3.3 |
| 40th Percentile | 2.58 |
| 30th Percentile | 2.26 |
| 1st Quartile (25th) | 2.1 |
| 20th Percentile | 1.9 |
| 10th Percentile | 1.42 |
| 5th Percentile | 0.91 |
| 2nd Percentile | 0.764 |
| 1st Percentile | 0.7 |
| Minimum Value | 0.6 |

> inset(As_PH,xlab= “As PH <2mm 3050B Maritimes”, log = FALSE, xlim = NULL, nclass = NULL, ifnright = TRUE)
1 row(s) with missing value(s), NA(s), removed from vector
> inset(Pb_PH,xlab= “Pb PH <2mm 3050B Maritimes”, log = FALSE, xlim = NULL, nclass = NULL, ifnright = TRUE)
1 row(s) with missing value(s), NA(s), removed from vector
> quit()

# Appendix 02-05  Scripts for Generating Tukey Box Plots

```
> tbplot(split(As_PH,EcoP),log=T,logx=T,ylab="As (mg/kg, US-EPA 3050B-AR_var) in <2 mm
0-5 cm soil",label=c("Appalachian &\nAcadian Highlands","Northumberland\nLowlands","Fundy\
nUplands"),cex=0.8)
> abline(h=12,lty=2)
> text(2,13.5,"CCME Limit - 12 mg/kg",adj=0.5,cex=0.8)

> tbplot(split(Pb_PH,EcoP),log=T,logx=T,ylab="Pb (mg/kg, US-EPA 3050B-AR_var) in <2 mm
0-5 cm soil",label=c("Appalachian &\nAcadian Highlands","Northumberland\nLowlands","Fundy\
nUplands"),cex=0.8)
> abline(h=70,lty=2)
> text(2,76,"CCME Limit - 70 mg/kg",adj=0.5,cex=0.8)
```

# Appendix 02-06. Workspace Using the 'gx.cnpplts' Function to Generate Q-Q Plots

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> rgr.first()
Loading required package: akima
Loading required package: MASS
Default plot shape set to 'square' and plot symbol to 'plus'
Working directory set to: D:\R\WD
R options set to warn = -1 to suppress unwanted graphics related messages from
  rgr multi-panel functions and functions calling eqscplot in Library MASS
> attach(as.pb.2mm.3050B.halifax)

> gx.cnpplts(xlab="As (mg/kg, US-EPA 3050B-AR_var) in <2 mm 0-5 cm soil",log=F,main="
",iflgnd=F,As_PH[EcoP==7.1],As_PH[EcoP==7.2],As_PH[EcoP==7.3])
  Data set 1 ; As_PH[EcoP == 7.1] ; range = 0.7 - 23.3 ; nx = 43 ; pch = 0 ; col = 1
  Data set 2 ; As_PH[EcoP == 7.2] ; range = 0.7 - 11 ; nx = 56 ; pch = 5 ; col = 3
  1 row(s) with missing value(s), NA(s), removed from vector
  Data set 3 ; As_PH[EcoP == 7.3] ; range = 0.6 - 23.4 ; nx = 84 ; pch = 2 ; col = 4
> lines(c(8,8),c(1.036,2.326),lty=2)
> text(locator(1),"Anomaly cut-off,\n8 mg/kg",cex=0.8,adj=0)
> lines(c(16,16),c(1.751,2.576),lty=2,col=4)
> text(locator(1),"Anomaly cut-off,\n16 mg/kg, for\nFunday Uplands",cex=0.8,adj=0,col=4)
> lines(c(12,12),c(0,2.9),lty=2)
```

```
> text(12,3.1,”CCME Limit - 12 mg/kg”,adj=0.5,cex=0.8)
> text(5,-2.0,”Appalachian & Acadian Highlands”,col=1,cex=0.8,adj=0)
> text(5,-2.4,”Northhumberland Lowlands”,col=3,cex=0.8,adj=0)
> text(5,-2.8,”Fundy Uplands”,col=4,cex=0.8,adj=0)

> gx.cnpplts(xlab=”Pb (mg/kg, US-EPA 3050B-AR_var) in <2 mm 0-5 cm soil”,log=F,main=”
“,iflgnd=F,Pb_PH[EcoP==7.1],Pb_PH[EcoP==7.2],Pb_PH[EcoP==7.3])
  Data set 1 ; Pb_PH[EcoP == 7.1] ; range = 15.94 - 125.17 ; nx = 43 ; pch = 0 ; col = 1
  Data set 2 ; Pb_PH[EcoP == 7.2] ; range = 6.67 - 67.03 ; nx = 56 ; pch = 5 ; col = 3
  1 row(s) with missing value(s), NA(s), removed from vector
  Data set 3 ; Pb_PH[EcoP == 7.3] ; range = 6.39 - 76.91 ; nx = 84 ; pch = 2 ; col = 4
> lines(c(70,70),c(0,2.9),lty=2)
> text(70,3.1,”CCME Limit - 70 mg/kg”,adj=0.5,cex=0.8)
> text(40,-2.0,”Appalachian & Acadian Highlands”,col=1,cex=0.8,adj=0)
> text(40,-2.4,”Northhumberland Lowlands”,col=3,cex=0.8,adj=0)
> text(40,-2.8,”Fundy Uplands”,col=4,cex=0.8,adj=0)
> lines(c(100,100),c(1.645,2.054),lty=2)
> text(80,0.842,”Anomaly cut-off,\n100 mg/kg, for the\nAppalachian and\nAcadian
Highlands”,cex=0.8,adj=0)
>quit()
```

# Appendix 02-07. Workspace Using the 'fences' Function to Make Calculated Estimates of Background

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> rgr.first()
Loading required package: akima
Loading required package: MASS
Default plot shape set to 'square' and plot symbol to 'plus'
Working directory set to: D:\R\WD
R options set to warn = -1 to suppress unwanted graphics related messages from

rgr multi-panel functions and functions calling eqscplot in Library MASS
Warning messages:
1: package 'rgr' was built under R version 2.8.0 and help will not work correctly
Please re-install it
2: package 'akima' was built under R version 2.11.0
3: package 'MASS' was built under R version 2.11.1
> attach(as.pb.2mm.3050B.halifax)
> names(as.pb.2mm.3050B.halifax)
 [1] "ID"  "Lat"  "Lng"
 [4] "Prov"  "EcoP"  "Ecoprovince"
 [7] "EcoR"  "Ecoregion"  "Soil.Type_SLC"
[10] "SubRx.Type..Wheeler." "As_PH"  "Pb_PH"
[13] "Corg_PH"  "As_Ahor"  "Pb_Ahor"
[16] "Corg_Ahor"  "As_Bhor"  "Pb_Bhor"
[19] "Corg_B.hor"  "As_Chor"  "Pb_Chor"

```
[22] “Corg_Chor”
> fences.summary(EcoP,As_PH,file=”Halifax”)
  Variable As_PH subset by EcoP - output will be in D:/R/WD/Halifax_EcoP_As_PH_fences.txt
> fences.summary(EcoP,Pb_PH,file=”Halifax”)
  Variable Pb_PH subset by EcoP - output will be in D:/R/WD/Halifax_EcoP_Pb_PH_fences.txt
> quit()
```

Note: Output from ‘fences’ function was the basis of Table 1: