DATABASE MANAGEMENT AND MAP PRODUCTION

T. Tarvainen[1], S. Reeder[2], and S. Albanese[3]

[1]Geological Survey of Finland, Espoo, Finland
[2]British Geological Survey, Keyworth, Nottingham, UK
[3]University of Napoli Federico II, Naples, Italy

Introduction

The FOREGS Geochemical Baseline Mapping Programme was initiated in 1998 to provide high quality environmental geochemical baseline data in Europe. Geological surveys and related institutions from 26 countries have taken part in the mapping programme. During the five working years, several European-wide data archives have been established. The FOREGS databases and material archives comprise: archived sample materials (topsoil, subsoil, floodplain and stream sediments, and humus) stored at the Geological Survey of the Slovak Republic; field observation sheets; work maps; Microsoft Access databases for field observations; analytical data files; databases of combined field and analytical data; GIS layers; work maps and tables; collections of field photographs; and a digital photo archive. The Geological Survey of Finland (GTK) has been responsible for database management and map production for the FOREGS group. Data handling and map production methods have been selected in the database management group workshops. Data verification and validation has been carried out by each participating country.

Database Structure

MS Access field databases, consisting of tables and entry forms for the sampled media, were completed by each country and merged at the Geological Survey of Finland (GTK). Input rules and constraints within the database were used to guarantee the consistency of the field data. The data entry forms were based on the field observation sheets presented in the FOREGS field manual (Salminen *et al.* 1998). Basic information collected for all sample sites included: sample identifiers (using agreed FOREGS sampling code); date of collection; name of sampler; sampling site location, including geographical region, map sheet, sampling coordinates (easting and northing and latitude and longitude) and altitude; number of subsites; a site description, including, landscape/topography, land use (agriculture, pasture, forest, wetland, *etc*), bedrock lithology and type of overburden); and photograph identifiers. In addition, more specific information was collected for each sample media. For soils, for example, the additional observations included: soil type; ploughing depth; subsoil horizon; sampling intervals; clast abundances; texture; humidity; organic content estimates; and gamma-radiation measurements.

Analytical data from nine European geochemical laboratories (see chapter on chemical analysis) were combined with the field data at GTK. The merged dataset was reviewed by the participating countries. ANOVA interpretation of the sampling and analytical data was used to select the suitability of elements and analytical methods for map production (see chapter on Quality Assurance). GIS data layers were then created and preliminary dot maps, basic tables and distribution graphs prepared. After review of the preliminary maps, all the field and analytical data were once again merged at the GTK, all mismatches of sample identifiers were reported to the laboratories or participating countries, and the reviewed, combined dataset was used to produce final maps and summary tables.

Selection of the Most Appropriate Analytical Data

In some cases, data for selected elements were available from more than one technique for the same sample media. For example, both XRF and total acid digest ICP-AES data were available for

stream and floodplain sediments and top and subsoils; similarly, both ICP-AES and ICP-MS data were available for water samples. Decisions on which technique to use were based on the sensitivity (limits of detection) of the analytical techniques, and the quality of the analytical data based on the ANOVA tests (see chapter Sample Preparation and Analysis and Tarvainen *et al.* 2004). To achieve consistency, the analytical data used for map production were normally selected exclusively from only one of the available

techniques. For XRF analyses of stream and floodplain, however, some analytical data exceeded the calibration range of the selected technique. In such cases, these values were replaced by ICP MS measurements. For example, the lead (Pb) concentration reported by XRFS for Greek stream sediment sample N26E14S3 was >1000 mg kg^{-1}. Lead was also analysed by ICP MS, giving a value of 1484 mg kg^{-1}, and this result was used in map production for this sample.

Procedure for Handling Data Below the Analytical Limit of Detection

For most analytical methods, the participating laboratories applied a fixed detection limit throughout the duration of the project, for example, the detection limit for tellurium (Te) determined by ICP-MS was 0.2 mg kg^{-1} in soil samples. In such cases, all values reported as less than the limit of detection were converted to half of the detection limit (0.1 mg kg^{-1} in the case of Te) before map production. On the distribution maps, values lower than the detection limit are shown with the smallest symbol size and belong to the lowest concentration range in the colour scale. These values are shown as DL/2 in the cumulative distribution function curve.

For water samples, detection limits have varied throughout the project. In part this has arisen because of changes (mainly improvements) to the analytical techniques with time, and in part it is

because of the need to carry out dilutions on some samples to get major ions within calibration range. As a result, values below the general detection limit have sometimes been reported. This variation is usually not shown in European-wide maps, but the values below the general detection limit can be seen as variation within the lowest dot size class in the cumulative distribution function curve.

All cation concentrations reported as less than the limit of detection were converted to half of the actual reported detection limit. However, detection limits for anions varied more during the project. Thus all anion concentrations reported as less than any limit of detection were converted to half of the most common detection limit before map production.

Basic Tables

The basic tables were calculated on the basis of reviewed combined data sets. The tables show the distribution of elements in all six sampling media as count, minimum, median, arithmetic mean, standard deviation, 90th percentile and maximum values. In some cases, the minimum and median

values are lower than the analytical detection limit. The reported data set includes all the extreme values, which can be higher by a factor oh ten than most values in the data set. The median value better reflects typical concentrations than the arithmetic mean.

Map Production

Representation of data distribution is based on Albers Equal-Area Conic projection with an applied spheroid/ellipsoid of WGS 84 (considered as the most important ellipsoid for international use). The central meridian is 20$°$ and the reference latitude is 0$°$. The first standard parallel is 45$°$ and the second standard parallel is 55$°$. The false Easting was fixed at 5,000,000 m, and the false Northing is 0. Thus, a box with

upper left corner coordinates (2,400,000 m, 7,500,000 m) and lower right corner coordinates (5,800,000 m, 3,400,000 m) covers almost the entire mapping area. The Canary Islands are presented as an inlet in the main frame of the map. The distributions of elements are presented with a combination of dot and colour surface maps (Tarvainen *et al.* 2003).

Geochemical maps should show the data with as little distortion as possible, with a minimum of computational artefacts. Dot maps reveal the actual sampling density, and any error in coordinates can be easily observed. The dot maps were produced using the ArcView GIS®, software, using an Alkemia Circmap type dot size function (Gustavsson *et al*. 1997) to classify data. In most cases, 10% of the lowest values are presented with the smallest symbol size and 2% of the highest concentrations with the largest symbol (16 mm, grey colour). Between these fixed percentiles, the rest of the distribution is divided into 14 symbol size categories using a logarithmic scale (Figure 1). If more than 10% of the samples are under the analytical detection limit, all the samples with concentration lower than the detection limit are shown with the smallest symbol size. The combination of almost continuous symbol size and logarithmic dot size function reveals anomaly patterns at both high and medium concentration levels.
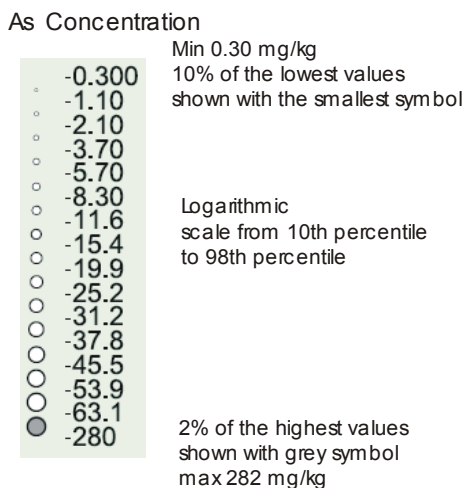
As Concentration

Min 0.30 mg/kg
-0.300   10% of the lowest values
-1.10   shown with the smallest symbol
-2.10
-3.70
-5.70
-8.30   Logarithmic
-11.6   scale from 10th percentile
-15.4   to 98th percentile
-19.9
-25.2
-31.2
-37.8
-45.5
-53.9
-63.1   2% of the highest values
-280   shown with grey symbol
          max 282 mg/kg

Figure 1. Dot size scale for topsoil arsenic concentrations. The minimum of the data set is 0.30 mg kg^{-1} and the maximum is 280 mg kg^{-1}. 10% of the lowest As concentrations are shown by the smallest symbol size and 2% of the highest concentrations by the big grey symbol. Between these fixed percentiles, the scale is logarithmic.

The data were interpolated to generate a regular grid with a 6 km x 6 km output cell size, using the Alkemia Smooth interpolation method (Gustavsson *et al*. 1997). For each cell, values were calculated using a moving weighted median in a circular window with a fixed radius of 150 km. A 10-grade colour scale was selected to present the distribution. The colour scale was based on the following percentiles: 5, 15, 25, 35, 50, 65, 75, 85 and 95 (Figure 2A). If more than

(A) Arsenic

Percentiles   Concentration

5	1.10	Very low
15	1.90	} Low background
25	2.80	
35	4.30	Below background
50	7.03	} Background
65	9.60	
75	12.5	Above background
85	17.7	High
95	33.1	Very high
		Highly anomalous

(B) Iodine

Percentiles   Concentration

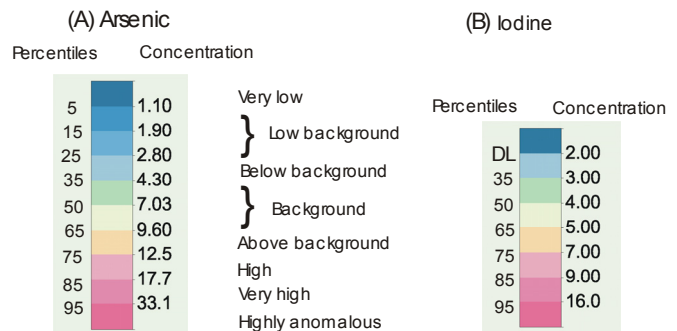| DL | 2.00 |
35	3.00
50	4.00
65	5.00
75	7.00
85	9.00
95	16.0

Figure 2. (A) Colour scale for topsoil arsenic concentrations. The 10-grade colour scale is based on the following percentiles: 5, 15, 25, 35, 50, 65, 75, 85 and 95. The concentration levels are referred to in the text using the term on the right side of the colour scale. For example, all values between 35th and 65th percentile (dark and pale green) are called 'background'.

(B). Colour scale for topsoil iodine concentrations. More then 25% of the observations were lower than the analytical detection limit of 2 mg kg^{-1}. Two blue shades are dropped from the scale and all values below the detection limit (DL) are shown with the dark blue shade. From the 35th percentile to 95th percentile, the scale is similar to the full scale in Fig. 2A.
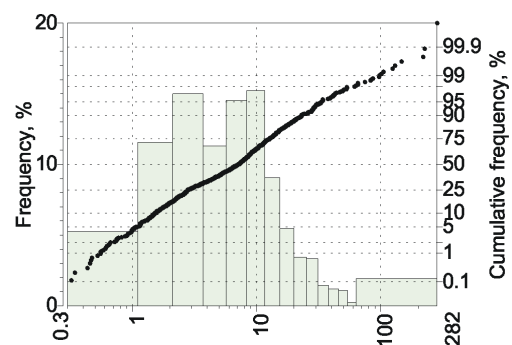
Figure 3. The histogram and cumulative distribution function curve for arsenic in topsoils. The histogram is based on the dot size scale shown in the figure under the histogram. For example, samples with concentrations in the range 2.10 to 3.70 mg kg^{-1} are shown with the third dot size on the map and about 15% of the samples belong to this concentration class. The concentration scale (x-axis) is logarithmic.

5% of the data were lower than the detection limit, one or more of the lower colour classes were not applied (Figure 2B).

The data distribution is shown as a combination of histogram and the cumulative distribution function curve on the upper left corner of each map. The histogram is based on the dot map classification scale (Figure 3). The cumulative distribution is presented with black dots in the graph.

In addition to the distribution maps published in the atlas, several work maps with different scales or combination of elements (*e.g.* factor scores maps, ratio maps topsoil/subsoil) or combination of field data and analytical results were produced for interpretation purposes.

Photograph Archive

Two photographs were taken from all sampling sites: one depicting a general overview of the sampling area, and one a close-up of the sample location. These photographs were scanned and stored in an Internet-based photo-archive. Two copies of each photograph have been stored: a 72 dots per inch (DPI) version for on-line screen viewing, and a 300 DPI version for printing.

References

Gustavsson, N., Lampio, E. & Tarvainen, T., 1997. Visusalization of geochemical data on maps at the Geological Survey of Finland. Journal of Geochemical Exploration, 59, 197-207.

Salminen, R., Tarvainen, T., Demetriades, A., Duris, M., Fordyce, F. M., Gregorauskiene, V., Kahelin, H., Kivisilla, J., Klaver, G., Klein, H., Larson, J. O., Lis, J., Locutura, J., Marsina, K., Mjartanova, H., Mouvet, C., O'Connor, P., Odor, L., Ottonello, G., Paukola, T., Plant, J. A., Reimann, C., Schermann, O., Siewers, U., Steenfelt, A., Van der Sluys, J., Vivo, B. de; & Williams, L., 1998. FOREGS geochemical mapping field manual. Geologian tutkimuskeskus. Opas 47. 36 p. + 1 app.

Tarvainen, T, Ahlsved, C, Reeder, S, Lima, A, Zuccolini, M, Bel Ian, A, Locutura, J & Lampio E., 2003. Overview of database management procedures used for the FOREGS geochemistry mapping project. In: 6th International Symposium on Environmental Geochemistry, Edinburgh 7-11 September 2003. Book of Abstracts. p. 211.

Tarvainen, T., Ahlsved, C., Albanese, S., Reeder, S., Salminen, R., Sandström, H., Savolainen, H. & Siewers, U., 2004. FOREGS Geochemical Mapping Programme: Assessment of analytical and data quality. 32nd International Geological Congress, Italia 2004. Abstracts Part 2. Workshops DWO16. p. 1530.