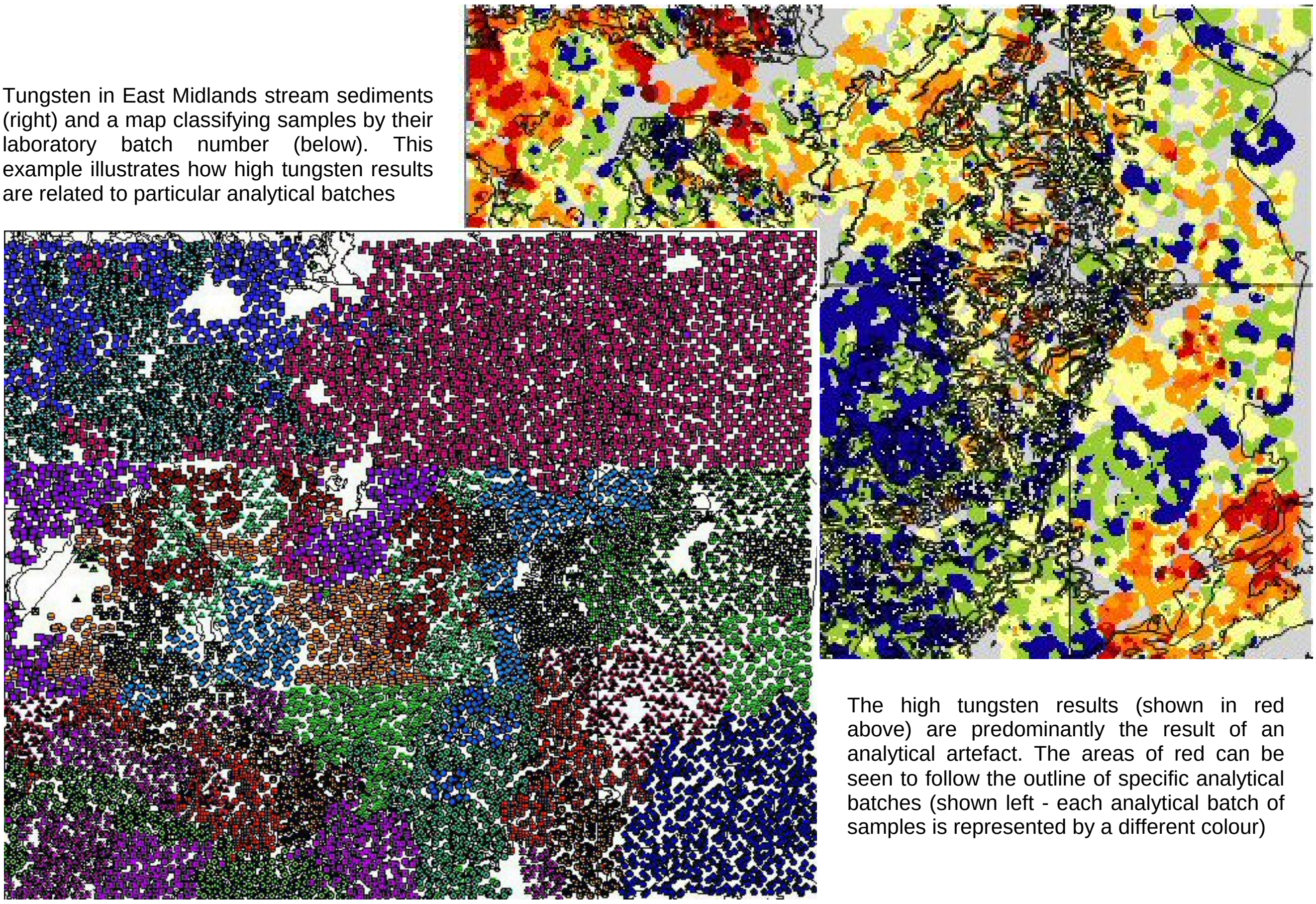
G-BASE data conditioning procedures for stream sediment and soil chemical analyses

Information Management & Chemical and Biological Hazards Programmes

Internal Report IR/05/150



Tungsten in East Midlands stream sediments (right) and a map classifying samples by their laboratory batch number (below). This example illustrates how high tungsten results are related to particular analytical batches

The high tungsten results (shown in red above) are predominantly the result of an analytical artefact. The areas of red can be seen to follow the outline of specific analytical batches (shown left - each analytical batch of samples is represented by a different colour)

BRITISH GEOLOGICAL SURVEY

INTERNAL REPORT IR/05/150

G-BASE data conditioning procedures for stream sediment and soil chemical analyses

T.R.Lister and C.C.Johnson

Editor

D.M.A. Flight

Keyworth, Nottingham British Geological Survey 2005

Keywords

Quality control, quality assurance, levelling, normalisation, G-BASE.

Front cover

Example of geochemical pattern generated by analytical error, Humber-Trent stream sediment raw analytical data with reference to the different laboratory batches submitted

Bibliographical reference

LISTER, T.R. AND JOHNSON, C.C.. 2005. G-BASE data conditioning procedures for stream sediment and soil chemical analyses. *British Geological Survey Internal Report*, IR/05/150. 85pp.

BRITISH GEOLOGICAL SURVEY

The full range of Survey publications is available from the BGS Sales Desks at Nottingham, Edinburgh and London; see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications including maps for consultation.

The Survey publishes an annual catalogue of its maps and other publications; this catalogue is available from any of the BGS Sales Desks.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as its basic research projects. It also undertakes programmes of British technical aid in geology in developing countries as arranged by the Department for International Development and other agencies.

The British Geological Survey is a component body of the Natural Environment Research Council.

British Geological Survey offices

Keyworth, Nottingham NG12 5GG
☎ 0115-936 3241 Fax 0115-936 3488
e-mail: sales@bgs.ac.uk
www.bgs.ac.uk
Shop online at: www.geologyshop.com

Murchison House, West Mains Road, Edinburgh EH9 3LA
☎ 0131-667 1000 Fax 0131-668 2683
e-mail: scotsales@bgs.ac.uk

London Information Office at the Natural History Museum (Earth Galleries), Exhibition Road, South Kensington, London SW7 2DE
☎ 020-7589 4090 Fax 020-7584 8270
☎ 020-7942 5344/45 email: bgslondon@bgs.ac.uk

Forde House, Park Five Business Centre, Harrier Way, Sowton, Exeter, Devon EX2 7HU
☎ 01392-445271 Fax 01392-445371

Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Belfast, BT9 5BF
☎ 028-9038 8462 Fax 028-9038 8461

Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB
☎ 01491-838800 Fax 01491-692345

Sophia House, 28 Cathedral Road, Cardiff, CF11 9LJ
☎ 029–2066 0147 Fax 029–2066 0159

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1EU
☎ 01793-411500 Fax 01793-411501
www.nerc.ac.uk

Contents

FIGURES

TABLES

Summary

Data conditioning is the process of making data fit for the purpose for which it is to be used and forms a significant component of the G-BASE project. This report is part of a series of manuals to record G-BASE project methodology. For data conditioning this has been difficult as applications used for processing data and the way in which data are reported continue to evolve rapidly and sections of this report have had to be continually updated to reflect this fact. However, the principals of data conditioning have changed little since the BGS regional geochemical mapping started in the late 1960s.

The process of data conditioning is based on one or more quality control procedures applied to the geochemical results as received from the laboratories, the degree of conditioning depending on how the data is to be used. The task is based on "blind" control samples being inserted prior to analysis, a system of quality control described in the G-BASE procedures manual. The first of the data conditioning processes is data verification and error checking, essentially assessing whether the laboratory has done what it was asked to do and results are being reported with reasonable accuracy. Shewhart or control charts form an important part of this process.

Once the data has been error checked, verified and accepted from the laboratory, further analysis of the data is carried out. These processes include: a series of x-y plots (of duplicate and replicate samples), more detailed control chart plots, and ANOVA analysis of the duplicate/replicate pairs to allocate variance in the results to sampling, analytical or between site variability. Analysis of both primary and secondary reference material can quantify analytical accuracy and precision. An important part of the data conditioning is the quality assurance and this includes procedures used for dealing with results that have data quality issues and documenting all parts of the data conditioning procedure.

The final part of the data conditioning procedure is necessary in order to use the data in context of other previously analysed data sets. This is the process of normalisation and levelling of the data. In G-BASE this is a very necessary step in order to create seamless geochemical maps and images across campaign boundaries and varying analytical methodologies that have spanned several decades.

1 Introduction

This report details the procedures used by the Geochemical Baseline Survey of the Environment Project (G-BASE) for conditioning analytical results received from laboratories prior to their inclusion in the British Geological Survey (BGS) Geochemistry Database. Here the term conditioning is used in the sense "*to render fit for work or use*". Results are received primarily from the BGS laboratories, although external laboratories have been used, and such results are defined as being the "raw data". These procedures refer to soils and stream sediments collected for a geochemical baseline survey. Stream water procedures are different and are described in another report (Ander, In Prep).

The data initially goes through a series of error checking and verification procedures that relate to data reporting; element ranges; absent, not determined and not detected results; and collation or mis-numbering errors. These procedures are essentially a check that the laboratory has carried out what they were asked to do and the results are reported to an acceptable standard. The quality of the data is then tested by statistical and graphical analysis of the data, element by element, using control samples inserted before submission for analysis. The use of duplicate, replicate and reference samples, which are "blind" to the analyst, is described in the G-BASE field procedures manual (Johnson, 2005) and was introduced in the early days of the G-BASE project (Plant et al., 1975). These samples are part of the G-BASE quality control procedures. The laboratories carry out their own quality control procedures during instrument validation, calibration and recalibration. Quality control (QC) measures are part of a quality assurance (QA) process, the latter requires procedures to deal with problem data and a final decision to accept or reject results. In the last part of the data conditioning process the results are levelled with reference to data that exists within the Geochemistry Database. This is essential to ensure seamless geochemical maps across field campaign boundaries that, as in the case of G-BASE, may have spanned several decades. The normalisation process also allows us to combine data produced by different analytical calibrations and methods, though this cannot be successfully achieved for all elements, particularly those where the majority of results are at or below the lower detection limit.

The error checking and quality control procedures refer to results generated by the X-Ray Fluorescence Spectrometry Section of the BGS laboratory services. The normalisation and levelling procedures cover results generated by different analytical methods, different laboratories and over a period of time.

Data sets used in G-BASE can generally be classified into one of three geographical groupings, namely atlas areas, urban areas or drainage catchments. Atlas areas (see Figure 1) are the standard data set for regional baseline data and have been the main classification for data conditioning in the past. This was because results have been published by atlas area and the combination of three or four years of sampling gave a large number of quality control samples for the quality assurance process. However, it had the disadvantage that it was several years before errors were found and consequently these were less easy to rectify after a long period of time. G-BASE will increasingly use drainage catchment reporting of its regional data as was undertaken for the co-funded Tamar drainage catchment survey (Rawlins et al., 2003). Urban baseline data are usually processed by combining data from a number of urban areas

(Lister, 2002a). Urban data sets are generally smaller than regional data sets and so a specific urban area may only be associated with a limited number of control samples.

Currently, analytical results are received as a laboratory batch consisting nominally of five hundred samples, and identified by a laboratory batch number. The laboratory batch number is the fundamental key in the laboratory information management system and an important parameter in the quality assurance procedure.

The overall flow of data conditioning is summarised in Figure 2 and the various stages in this process are discussed in the subsequent sections of this report.

Data conditioning is a time consuming task but as the geochemical database it creates is central to all the geochemical map production, interpretating and reporting it must be afforded the necessary time. For large geochemical mapping programmes, even if highly accredited laboratories are used for the analysis, it can take several years to bring the data up to an acceptable level of quality to produce final products (Reiman, 2005).

1. Shetland
2. Orkney
3. South Orkney and Caithness
4. Sutherland
5. Hebrides
6. Great Glen
7. East Grampian
8. Argyll
9. Southern Scotland
10. Lake District
11. NE England
12. NW England and N Wales
13. Humber-Trent
14. Wales & W Midlands (soil and sediment)
15. Wales & W Midlands (surface water)
16. East Midlands
17. East Anglia
18. SE England
19. SW England
20. Northern Ireland

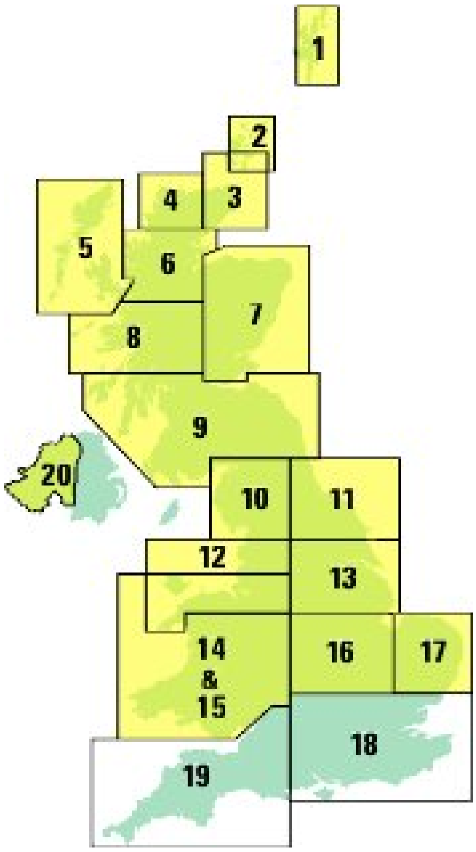


Figure 1: G-BASE atlas areas.

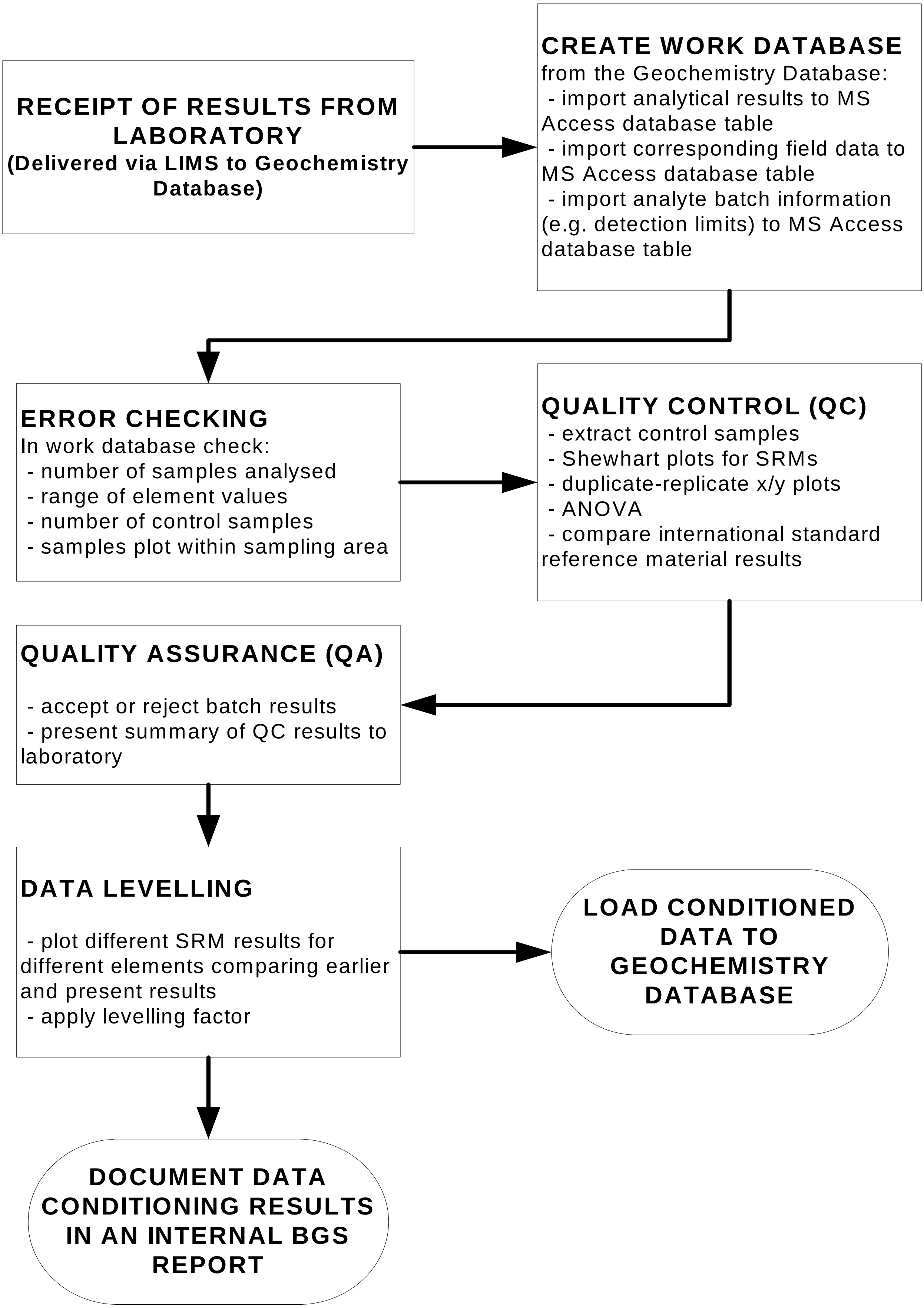


Figure 2: Figure showing flow of data conditioning procedures

2 Receipt of Results

Data received from the BGS laboratories consist of: lists of analytical results listed by site number; a cover note which describes the analytical protocol used and data quality statements (example is given in Appendix 1); and ancillary information required to interpret and use the results such as detection limits and international standard reference material results. These data have been received as both digital and hardcopy format. The actual format of reported data has rapidly evolved in recent years. At the end of 2004 analytical results for soils and stream sediments were transferred by the BGS Laboratory Information Management System (LIMS) direct to raw data tables in the Oracle Geochemistry Database. This involved the introduction of a new set of codes to represent missing or semi-quantitative results (Table 1).

Code	Comment
-94	Insufficient sample (e.g. sample collected but not enough to analyse or sample lost)
-95	Not determined because of high concentration; but exceeds calibration limit
-96	Not determined because of interference; probably of high concentration
-97	Not determined because of interference; probably of low concentration
-98	Not determined because of interference; no estimate
-99	Absent data (e.g. not requested)

Table 1: Codes for missing or semi-quantitative results

The procedures for the receipt of the results up to the point where the conditioning of the data can commence are summarised in Figure 3. A key field in the monitoring of the progress of data analysis and interpretation is the laboratory batch number, a unique reference number issued by the BGS laboratories. When the BGS LIMS transfers raw data to the Geochemistry Database other data tables concerning detection limits and sample registration information for each batch are also automatically populated. The laboratory batch number is therefore present in all these data tables. The Oracle data tables populated by the LIMS are summarised in Table 2. These data tables are described in more detail in Appendix 2. Raw data transferred to the Geochemistry Database contains G-BASE control samples with site numbers that are indistinguishable from normal samples. They can be retrieved as standards from the database by using the SAMP_STD field loaded to the DTA_SEDIMENTS, DTA_WATERS and DTA_OVERBURDENS data tables coded as shown in Table 3.

Results for primary reference material are transferred from the laboratory (currently only for sediments and soils) to an Oracle data table BGS_ DTA_REFERENCE_MATERIAL_DATA and the reference materials used are described in a dictionary table (BGS_DIC_ REFERENCE_MATERIAL).

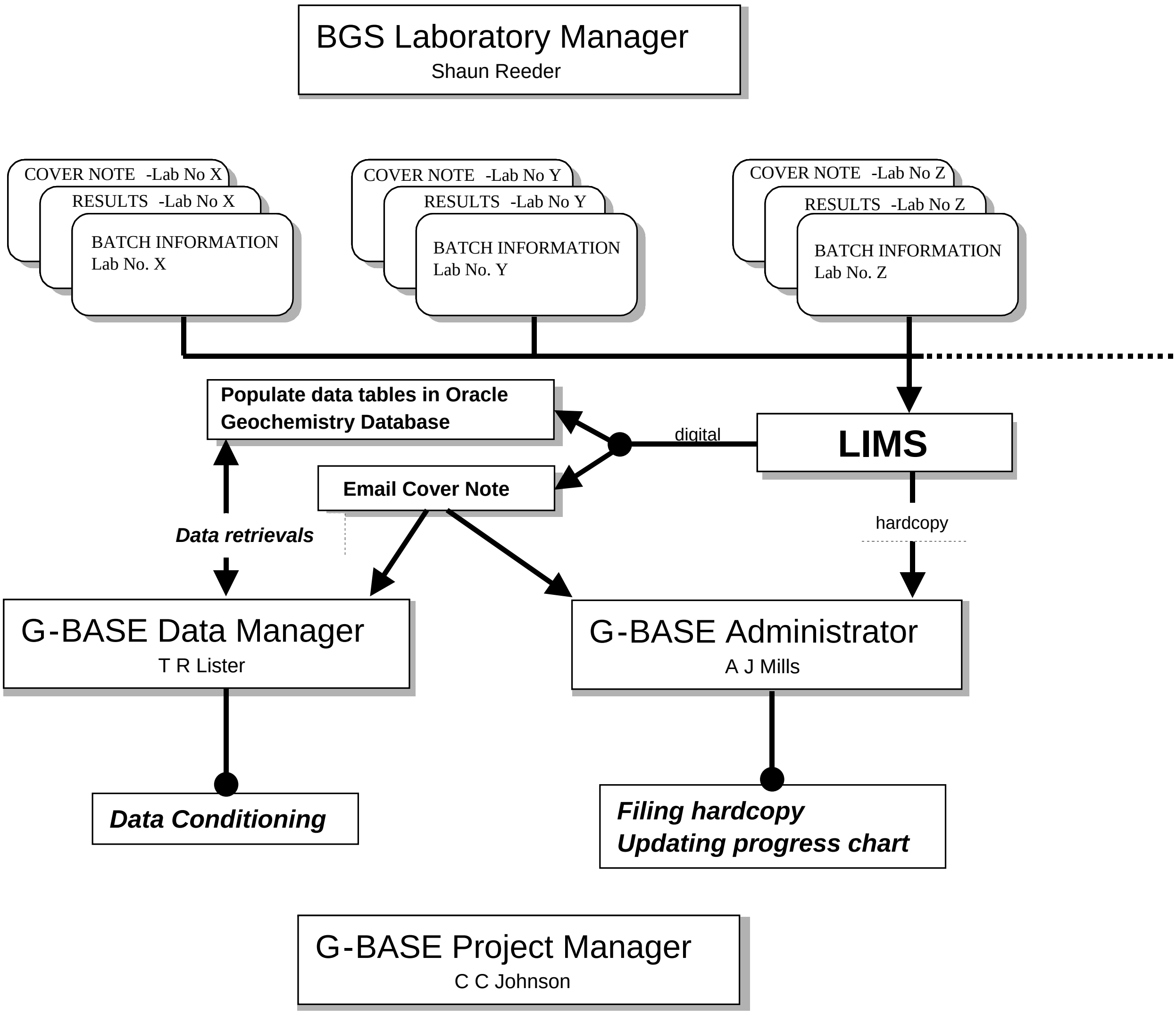


Figure 3: Figure showing the flow of procedures when results are sent from the laboratories

Table Name	Description
Analyte BGS_DTA_RAW_ANALYTES	List of analytes determined, the method used and the limits of detection
Analyte determinations BGS_DTA_RAW_GBASE_ANALYTE_DETAILS	Raw analytical results
Analysis BGS_DTA_ANALYSIS	The analysis carried out by a particular laboratory by a specified method, for the owner of the samples, costed to a particular project
Batch BGS_DTA_BATCHES	Details of the batch of samples including name of area from which samples were collected
Project Batch BGS_DTA_PROJECT_BATCHES	Table listing maximum and minimum site numbers and number of samples in the submitted batch
Primary Reference Materials BGS_ DTA_REFERENCE_MATERIAL_DATA	Table containing primary reference material determined before and after each analytical batch is run

Table 2: Summary of Oracle tables populated during transfer of raw data to the Geochemistry Database

DUPA	Duplicate A (original sample)
DUPB	Duplicate B (collected at same site as Dup A)
DUPC	Duplicate C (original sample)
DUPD	Duplicate D (collected at same site as Dup C)
SSA	Subsample A (laboratory replicate of DUPA)
SSB	Subsample B (laboratory replicate of DUPB)
SSC	Subsample C (laboratory replicate of DUPC)
SSD	Subsample D (laboratory replicate of DUPD)
STD	Secondary ref. material (SRM) for A,S,C and W
BW	Blank water used only for W
MON	Water monitor site sample used only for W

Table 3: List of control sample codes used in the SAMP_STD field of the field database

3 Creating a Work Database

3.1 PREPARING THE DATA

Before commencing the data conditioning exercise all the relevant data needs to be loaded into tables in a MS ACCESS database. It is important that the data used is complete and finalised - it is not efficient to work on partially completed or preliminary data sets. Results are considered to be complete when signed off by the Laboratory Manager (or their representative) and the G-BASE Data Manager must consider how to group data for the process of conditioning. The data required are:

a) The raw analytical results (including control samples)

b) Corresponding field data that has been prepared as described by Lister et al. (2005)

c) Analytical batch information (such as detection limits)

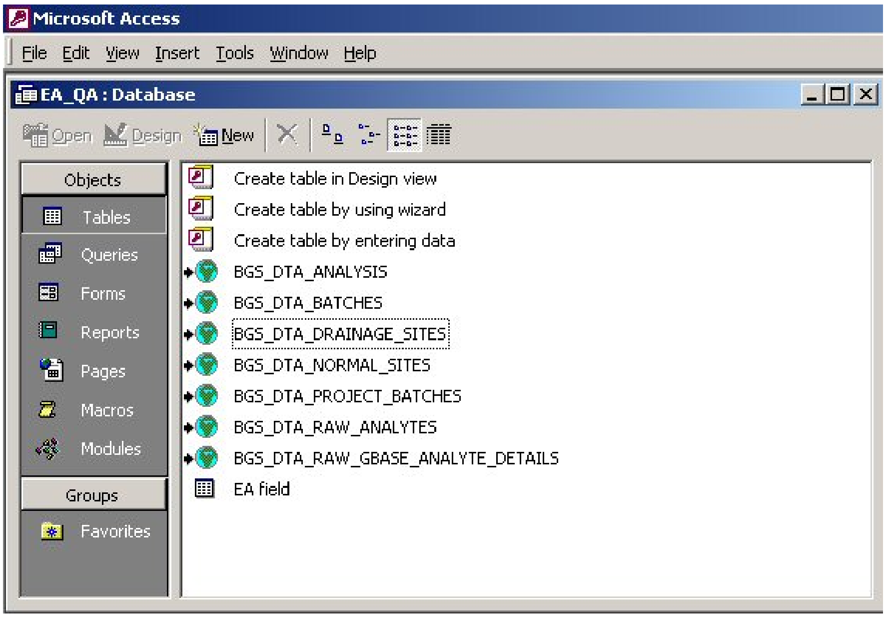
These data should be available in the BGS corporate Oracle Geochemistry Database (see Table 2), and in the numerous site information tables that are created from the field database as documented by Lister et al. (2005). The principal site information tables are BGS_DTA_DRAINAGE_SITES, BGS_DTA_SEDIMENTS and BGS_DTA_WATERS (for drainage samples), and BGS_DTA_NORMAL_SITES and BGS_DTA_OVERBURDENS (for soils). Therefore, the first step in the process is a series of retrieval queries from the Oracle database using MS ACCESS.

Prior to 2005 this data was not routinely input to the Geochemistry Database and so pre-2005 data needs to be gathered from a number of different sources. As all the data is now routinely input to the Geochemistry Database, it is envisaged that in the future a certain amount of the data processing described later in this section could be done automatically as the data is retrieved from the Oracle database.

For the purpose of this report the starting point of data conditioning is a MS ACCESS database populated with the three sets of data listed above. This database is referred to as the "work database".

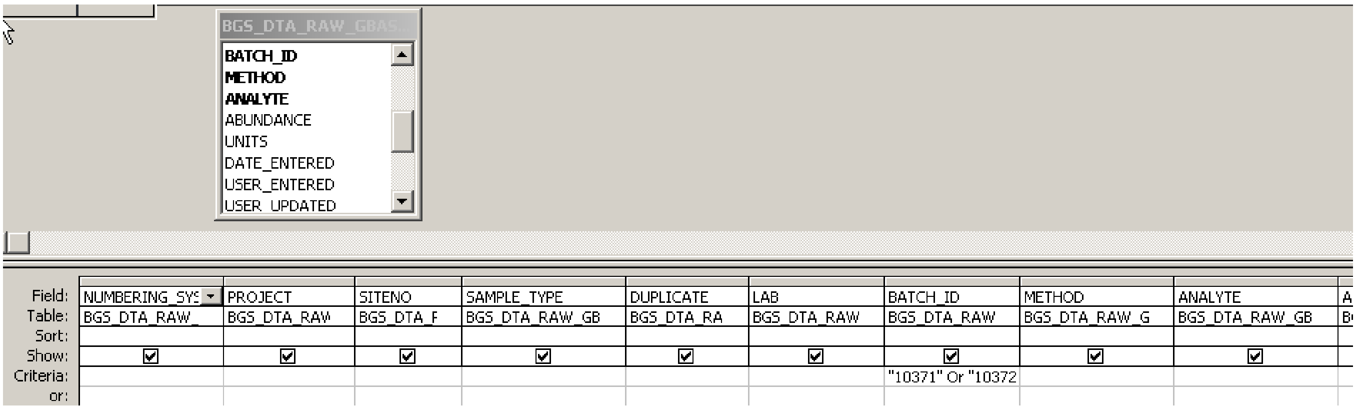
The raw data as received from the laboratories must be preserved in an unaltered state so it is available for future reference. In the BGS Geochemistry Database there are tables for G-BASE of "raw data" (see Table 2) that remain as the original record of the raw data. Once the raw data has been conditioned it is loaded into a different analyte table of the Geochemistry Database (BGS_DTA_ ANALYTE_DETERMINATIONS). Prior to the direct transfer of data from LIMS to the database raw data were received as MS EXCEL files and these are maintained in a data archive by the G-BASE Data Manager

3.2 EXAMPLE WORK DATABASE



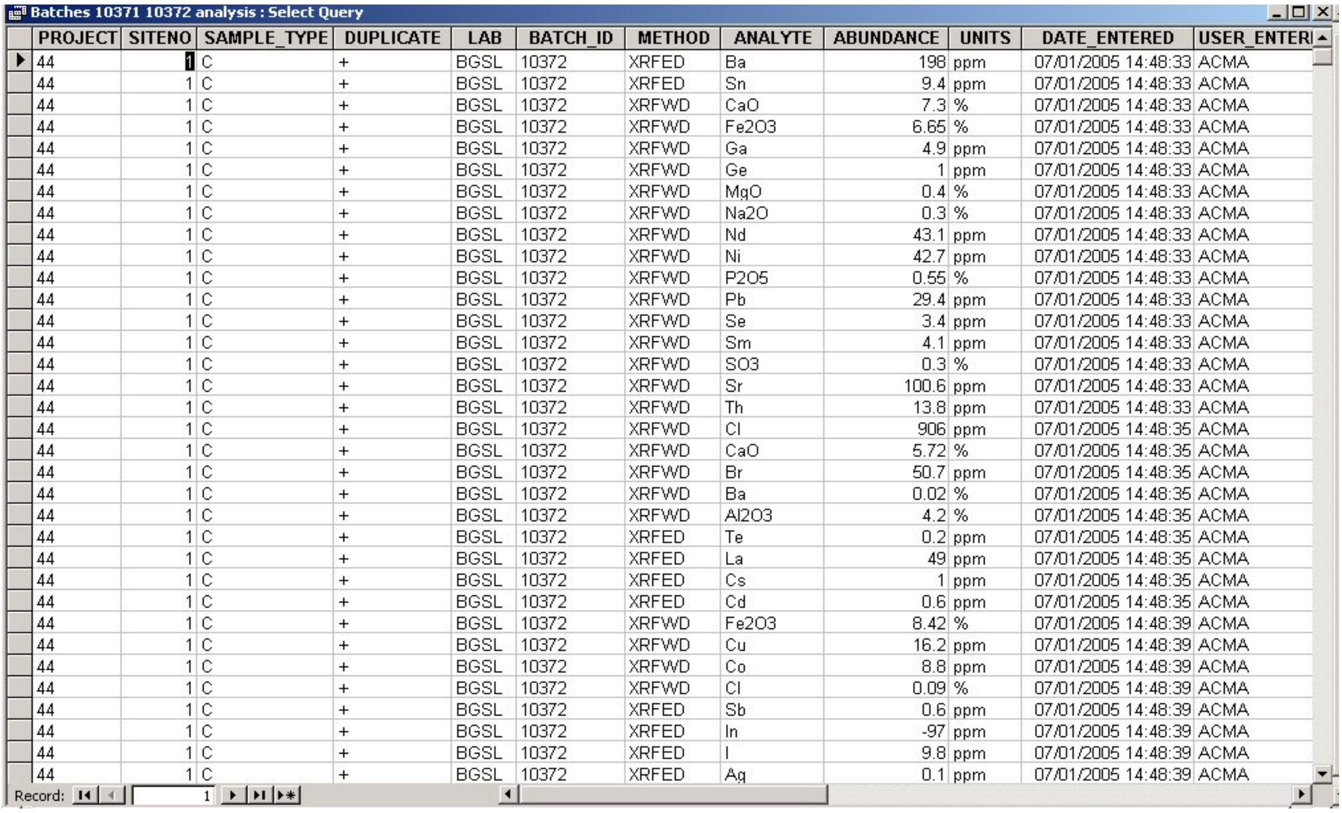
Create table in Design view
Create table by using wizard
Create table by entering data
BGS_DTA_ANALYSIS
BGS_DTA_BATCHES
BGS_DTA_DRAINAGE_SITES
BGS_DTA_NORMAL_SITES
BGS_DTA_PROJECT_BATCHES
BGS_DTA_RAW_ANALYTES
BGS_DTA_RAW_GBASE_ANALYTE_DETAILS
EA field

Figure 4: Example of linked tables in ACCESS database ready for creating retrieval queries



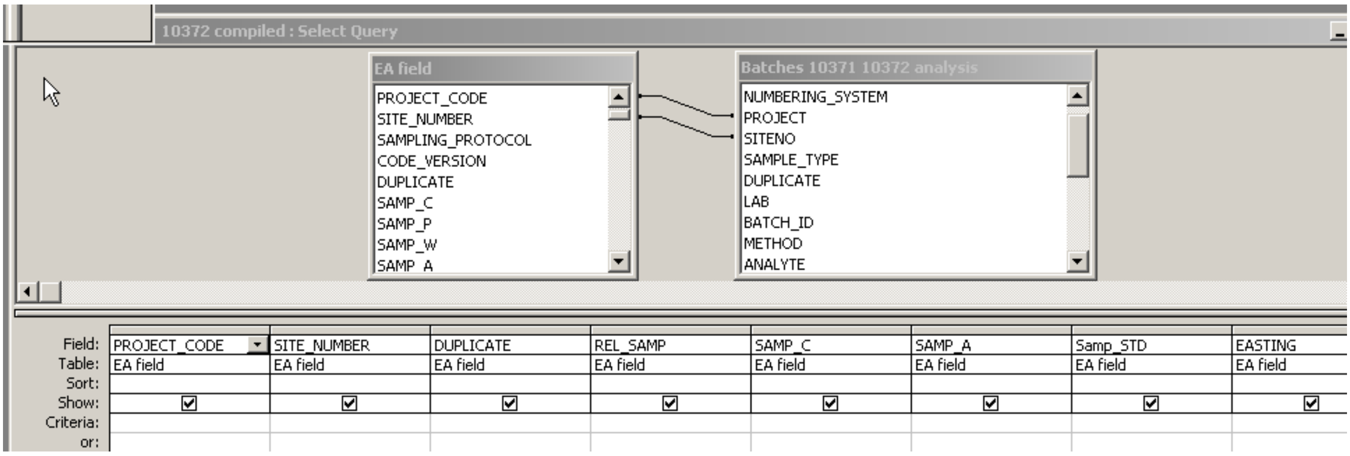
Field:	NUMBERING_SYS	PROJECT	SITENO	SAMPLE_TYPE	DUPLICATE	LAB	BATCH_ID	METHOD	ANALYTE	A
Table:	BGS_DTA_RAW_	BGS_DTA_RAW	BGS_DTA_F	BGS_DTA_RAW_GB	BGS_DTA_RA	BGS_DTA_RAW	BGS_DTA_RAW	BGS_DTA_RAW_G	BGS_DTA_RAW_GB	B
Sort:										
Show:	☑	☑	☑	☑	☑	☑	☑	☑	☑	
Criteria:							"10371" Or "10372			
or:										

Figure 5: Example of the query used initially to retrieve the analytical data for two analytical batches (from data table **BGS_DTA_RAW_GBASE_ANALYTE_DETAILS)**



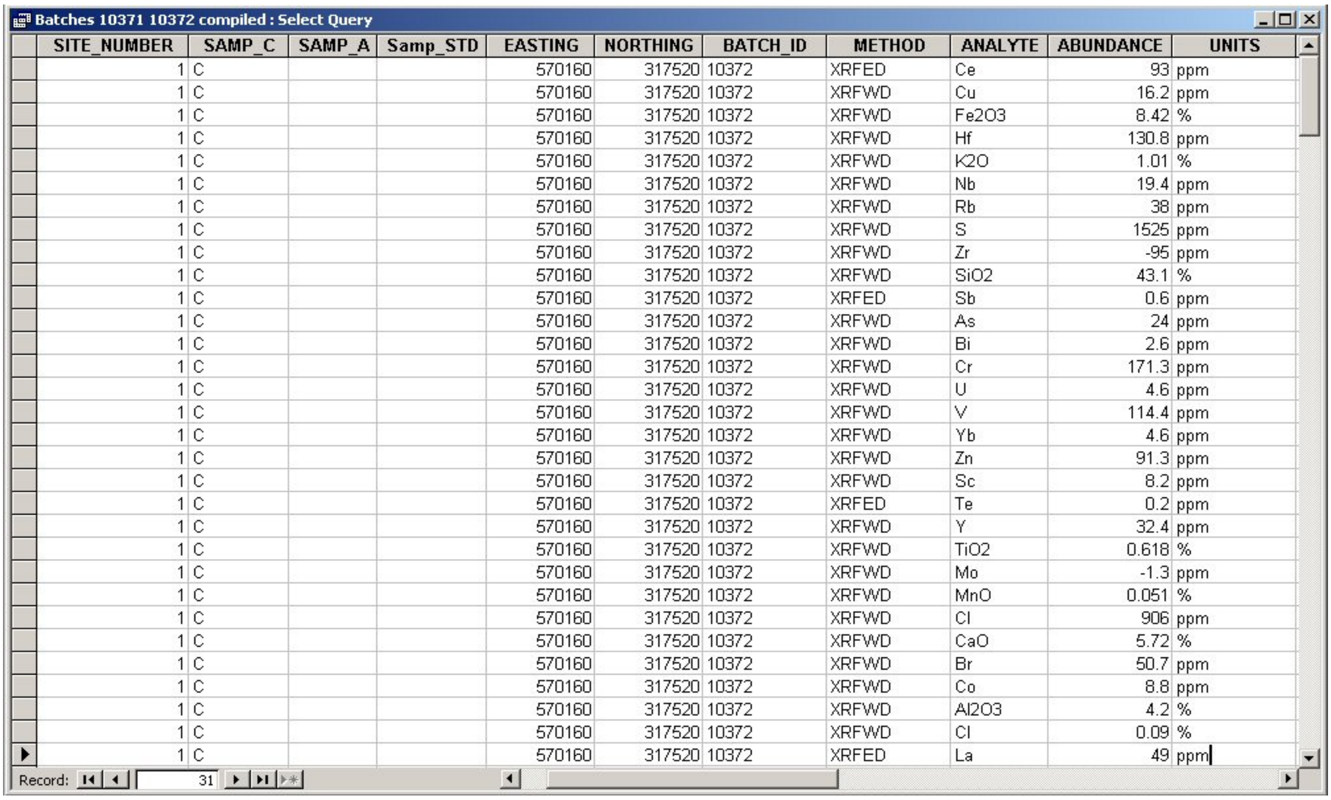
PROJECT	SITENO	SAMPLE_TYPE	DUPLICATE	LAB	BATCH_ID	METHOD	ANALYTE	ABUNDANCE	UNITS	DATE_ENTERED	USER_ENTER
44	1	C	+	BGSL	10372	XRFED	Ba	198	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFED	Sn	9.4	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	CaO	7.3	%	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Fe2O3	6.65	%	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Ga	4.9	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Ge	1	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	MgO	0.4	%	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Na2O	0.3	%	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Nd	43.1	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Ni	42.7	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	P2O5	0.55	%	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Pb	29.4	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Se	3.4	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Sm	4.1	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	SO3	0.3	%	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Sr	100.6	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Th	13.8	ppm	07/01/2005 14:48:33	ACMA
44	1	C	+	BGSL	10372	XRFWD	Cl	906	ppm	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFWD	CaO	5.72	%	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFWD	Br	50.7	ppm	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFWD	Ba	0.02	%	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFWD	Al2O3	4.2	%	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFED	Te	0.2	ppm	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFED	La	49	ppm	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFED	Cs	1	ppm	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFED	Cd	0.6	ppm	07/01/2005 14:48:35	ACMA
44	1	C	+	BGSL	10372	XRFWD	Fe2O3	8.42	%	07/01/2005 14:48:39	ACMA
44	1	C	+	BGSL	10372	XRFWD	Cu	16.2	ppm	07/01/2005 14:48:39	ACMA
44	1	C	+	BGSL	10372	XRFWD	Co	8.8	ppm	07/01/2005 14:48:39	ACMA
44	1	C	+	BGSL	10372	XRFWD	Cl	0.09	%	07/01/2005 14:48:39	ACMA
44	1	C	+	BGSL	10372	XRFED	Sb	0.6	ppm	07/01/2005 14:48:39	ACMA
44	1	C	+	BGSL	10372	XRFED	In	-97	ppm	07/01/2005 14:48:39	ACMA
44	1	C	+	BGSL	10372	XRFED	I	9.8	ppm	07/01/2005 14:48:39	ACMA
44	1	C	+	BGSL	10372	XRFED	Ag	0.1	ppm	07/01/2005 14:48:39	ACMA

Figure 6: Example output from the query shown in Figure 5



Field:	PROJECT_CODE	SITE_NUMBER	DUPLICATE	REL_SAMP	SAMP_C	SAMP_A	Samp_STD	EASTING
Table:	EA field	EA field	EA field	EA field	EA field	EA field	EA field	EA field
Sort:								
Show:	☑	☑	☑	☑	☑	☑	☑	☑
Criteria:								
or:								

Figure 7: Example query design for retrieving analytical and site information (using unlinked EA field data table and query shown in Figure 5)

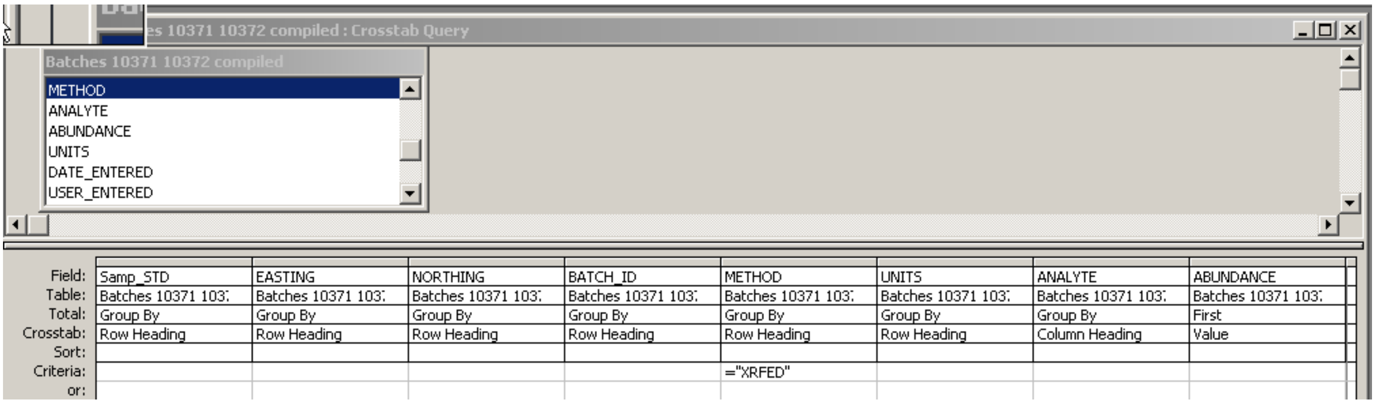


Batches 10371 10372 compiled : Select Query

SITE_NUMBER	SAMP_C	SAMP_A	Samp_STD	EASTING	NORTHING	BATCH_ID	METHOD	ANALYTE	ABUNDANCE	UNITS
1	C			570160	317520	10372	XRFED	Ce	93	ppm
1	C			570160	317520	10372	XRFWD	Cu	16.2	ppm
1	C			570160	317520	10372	XRFWD	Fe2O3	8.42	%
1	C			570160	317520	10372	XRFWD	Hf	130.8	ppm
1	C			570160	317520	10372	XRFWD	K2O	1.01	%
1	C			570160	317520	10372	XRFWD	Nb	19.4	ppm
1	C			570160	317520	10372	XRFWD	Rb	38	ppm
1	C			570160	317520	10372	XRFWD	S	1525	ppm
1	C			570160	317520	10372	XRFWD	Zr	-95	ppm
1	C			570160	317520	10372	XRFWD	SiO2	43.1	%
1	C			570160	317520	10372	XRFED	Sb	0.6	ppm
1	C			570160	317520	10372	XRFWD	As	24	ppm
1	C			570160	317520	10372	XRFWD	Bi	2.6	ppm
1	C			570160	317520	10372	XRFWD	Cr	171.3	ppm
1	C			570160	317520	10372	XRFWD	U	4.6	ppm
1	C			570160	317520	10372	XRFWD	V	114.4	ppm
1	C			570160	317520	10372	XRFWD	Yb	4.6	ppm
1	C			570160	317520	10372	XRFWD	Zn	91.3	ppm
1	C			570160	317520	10372	XRFWD	Sc	8.2	ppm
1	C			570160	317520	10372	XRFED	Te	0.2	ppm
1	C			570160	317520	10372	XRFWD	Y	32.4	ppm
1	C			570160	317520	10372	XRFWD	TiO2	0.618	%
1	C			570160	317520	10372	XRFWD	Mo	-1.3	ppm
1	C			570160	317520	10372	XRFWD	MnO	0.051	%
1	C			570160	317520	10372	XRFWD	Cl	906	ppm
1	C			570160	317520	10372	XRFWD	CaO	5.72	%
1	C			570160	317520	10372	XRFWD	Br	50.7	ppm
1	C			570160	317520	10372	XRFWD	Co	8.8	ppm
1	C			570160	317520	10372	XRFWD	Al2O3	4.2	%
1	C			570160	317520	10372	XRFWD	Cl	0.09	%
1	C			570160	317520	10372	XRFED	La	49	ppm

Record: 31

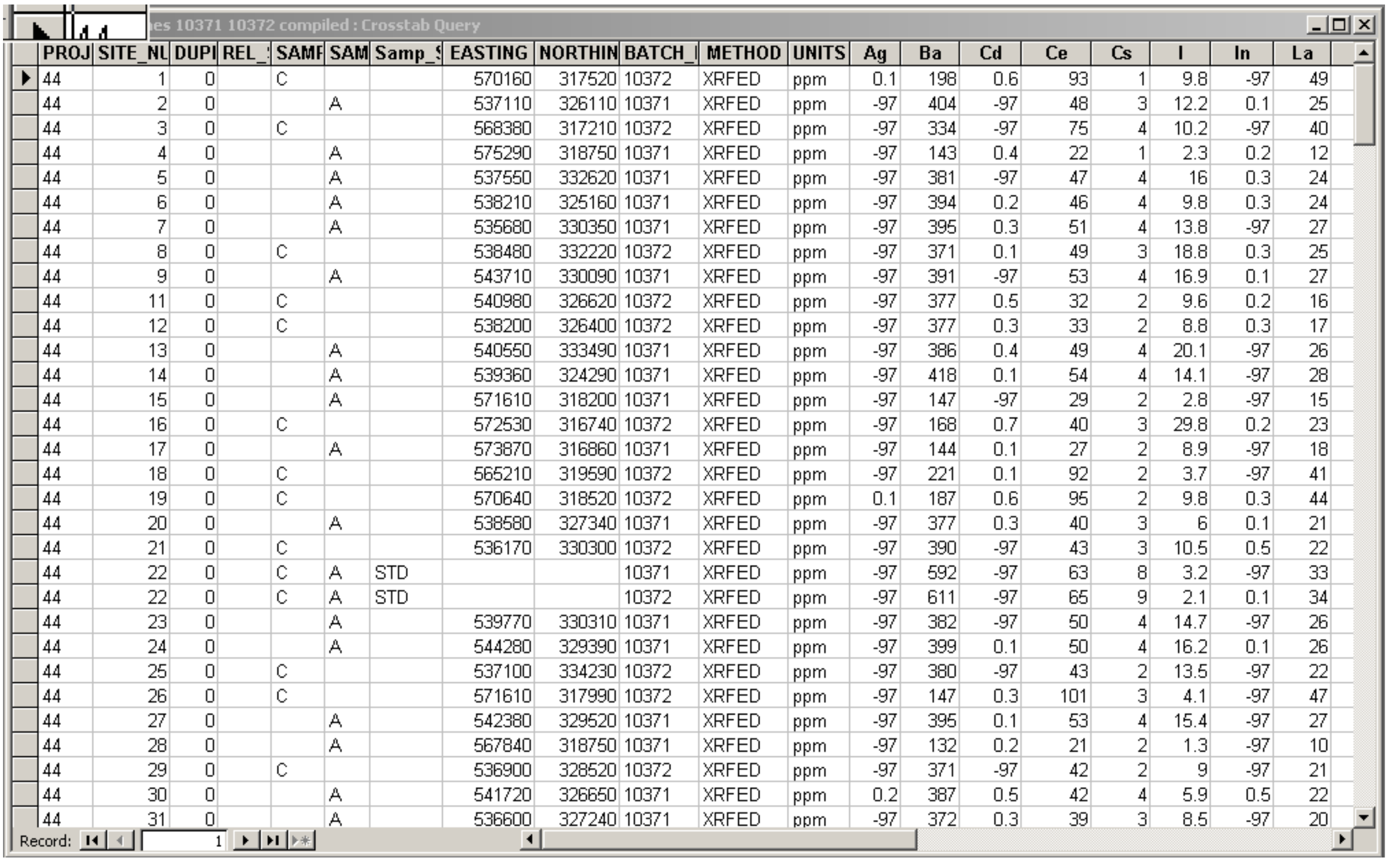
Figure 8: Data retrieved by the query shown in Figure 7



es 10371 10372 compiled : Crosstab Query

Field:	Samp_STD	EASTING	NORTHING	BATCH_ID	METHOD	UNITS	ANALYTE	ABUNDANCE
Table:	Batches 10371 103	Batches 10371 103	Batches 10371 103	Batches 10371 103	Batches 10371 103	Batches 10371 103	Batches 10371 103	Batches 10371 103
Total:	Group By	Group By	Group By	Group By	Group By	Group By	Group By	First
Crosstab:	Row Heading	Row Heading	Row Heading	Row Heading	Row Heading	Row Heading	Column Heading	Value
Sort:								
Criteria:					="XRFED"			
or:								

Figure 9: Design view of a cross tab query for retrieving data in a way that lists analytes in columns and the site number as a single record (listing only analyses by XRFED)



PROJ	SITE_NU	DUPI	REL_	SAMF	SAM	Samp_S	EASTING	NORTHIN	BATCH_	METHOD	UNITS	Ag	Ba	Cd	Ce	Cs	I	In	La
44	1	0		C			570160	317520	10372	XRFED	ppm	0.1	198	0.6	93	1	9.8	-97	49
44	2	0			A		537110	326110	10371	XRFED	ppm	-97	404	-97	48	3	12.2	0.1	25
44	3	0		C			568380	317210	10372	XRFED	ppm	-97	334	-97	75	4	10.2	-97	40
44	4	0			A		575290	318750	10371	XRFED	ppm	-97	143	0.4	22	1	2.3	0.2	12
44	5	0			A		537550	332620	10371	XRFED	ppm	-97	381	-97	47	4	16	0.3	24
44	6	0			A		538210	325160	10371	XRFED	ppm	-97	394	0.2	46	4	9.8	0.3	24
44	7	0			A		535680	330350	10371	XRFED	ppm	-97	395	0.3	51	4	13.8	-97	27
44	8	0		C			538480	332220	10372	XRFED	ppm	-97	371	0.1	49	3	18.8	0.3	25
44	9	0			A		543710	330090	10371	XRFED	ppm	-97	391	-97	53	4	16.9	0.1	27
44	11	0		C			540980	326620	10372	XRFED	ppm	-97	377	0.5	32	2	9.6	0.2	16
44	12	0		C			538200	326400	10372	XRFED	ppm	-97	377	0.3	33	2	8.8	0.3	17
44	13	0			A		540550	333490	10371	XRFED	ppm	-97	386	0.4	49	4	20.1	-97	26
44	14	0			A		539360	324290	10371	XRFED	ppm	-97	418	0.1	54	4	14.1	-97	28
44	15	0			A		571610	318200	10371	XRFED	ppm	-97	147	-97	29	2	2.8	-97	15
44	16	0		C			572530	316740	10372	XRFED	ppm	-97	168	0.7	40	3	29.8	0.2	23
44	17	0			A		573870	316860	10371	XRFED	ppm	-97	144	0.1	27	2	8.9	-97	18
44	18	0		C			565210	319590	10372	XRFED	ppm	-97	221	0.1	92	2	3.7	-97	41
44	19	0		C			570640	318520	10372	XRFED	ppm	0.1	187	0.6	95	2	9.8	0.3	44
44	20	0			A		538580	327340	10371	XRFED	ppm	-97	377	0.3	40	3	6	0.1	21
44	21	0		C			536170	330300	10372	XRFED	ppm	-97	390	-97	43	3	10.5	0.5	22
44	22	0		C	A	STD			10371	XRFED	ppm	-97	592	-97	63	8	3.2	-97	33
44	22	0		C	A	STD			10372	XRFED	ppm	-97	611	-97	65	9	2.1	0.1	34
44	23	0			A		539770	330310	10371	XRFED	ppm	-97	382	-97	50	4	14.7	-97	26
44	24	0			A		544280	329390	10371	XRFED	ppm	-97	399	0.1	50	4	16.2	0.1	26
44	25	0		C			537100	334230	10372	XRFED	ppm	-97	380	-97	43	2	13.5	-97	22
44	26	0		C			571610	317990	10372	XRFED	ppm	-97	147	0.3	101	3	4.1	-97	47
44	27	0			A		542380	329520	10371	XRFED	ppm	-97	395	0.1	53	4	15.4	-97	27
44	28	0			A		567840	318750	10371	XRFED	ppm	-97	132	0.2	21	2	1.3	-97	10
44	29	0		C			536900	328520	10372	XRFED	ppm	-97	371	-97	42	2	9	-97	21
44	30	0			A		541720	326650	10371	XRFED	ppm	0.2	387	0.5	42	4	5.9	0.5	22
44	31	0			A		536600	327240	10371	XRFED	ppm	-97	372	0.3	39	3	8.5	-97	20

Figure 10: Example output from the cross tab query shown in Figure 9

The "Samp_STD' field included in the table shown in Figure 10 can be used to select the different control samples using the codes listed in Table 3.

4 Error Checking and Verification

4.1 INITIAL ERROR CHECKING

The following should be addressed:

- Are analyses received consistent with those requested at the time of sample registration?
- Do the number of results corresponds with the number of samples submitted?
- the analytical results should be merged with the field data to check if there are any samples collected that were not analysed
- the range of analyte concentrations should be checked to see the values correspond to the units that they were reported in and that there are no obviously erroneous data values. It may be necessary to convert some element fields from percentage oxide to mg kg^{-1}, the latter being the unit of concentration for all results in the analyte table of the Geochemistry Database. Table 4 gives the oxide to mg kg^{-1} conversation factors. This table is available as a MS ACCESS data table
- Are a complete set of control samples present and are they correctly identified in the database?
- Do the merged field and analytical data produce a list of samples that all plot within the area sampled?

Any errors found should be systematically listed in a table that should detail how errors may be rectified. This list of errors should be included in the data conditioning report.

Element	Oxide	Conversion Factor
Al	Al2O3	1.889
Ca	CaO	1.399
Fe	Fe2O3	1.430
K	K2O	1.205
Mg	MgO	1.658
Mn	MnO	1.291
Na	Na2O	1.348
P	P2O5	2.291
Ti	TiO2	1.668

Table 4: Table of the most commonly used oxide conversion factors

4.2 DEALING WITH MISSING, SEMI-QUANTITIVE AND UNRELIABLE DATA

Once the initial error checking described above has been completed the control sample results should be prepared for processing and interpretation. As these processes usually involve some statistical analysis or graphical plotting it is necessary to deal with missing, semi-quantitative and unreliable data, particularly replacing the codes summarised in Table 1. Discussions on what to do with such data are to be found in Albert and Horwitz (1995) and AMC (2001). The G-BASE project routinely replaces results recorded as "<" by a value one-half the reported lower limit of detection (LLD).

The minus numeric codes inserted in the results to indicate missing or semi-quantitative results must be removed from the database and substituted with alternative values that will be acceptable to the statistical and plotting processes carried out in the quality control procedures. However, it is important that issues regarding such data are flagged up in the database and this is achieved using the "qualifier" field. In the Oracle analyte abundance table (BGS_DTA_ ANALYTE_DETERMINATIONS) where conditioned results are stored, there is a field called "qualifier" which can be populated with a value from the "Analysis qualifier" domain table (BGS_DOM_ANALYSIS_QUALIFIER). From 1st January 2006 when populating the qualifier field became compulsory, a null entry in this field means there was no known quality issues at the time of data loading.

CODE	TRANSLATION	DEFINED_AS	LIMS Code
>	Probably high	Not determined accurately due to interference. Probably higher.	-96 or -95
<	Probably low	Not determined accurately due to interference. Probably lower.	-97
^	No estimate possible	Not determined due to interference. No estimate possible.	-98
*	Dubious quality	The value has a documented quality control issue that should restrict its use.	
#	Estimated value	Value reported was below the lower detection limit cited by the analyst and has been set to half this detection limit.	
$	Uncertain value	Value as reported was below the lower detection limit cited by the analyst.	
&	Uncertain quality	No information exists regarding quality of this data.	
!	Not available	Applicable, but try as we might, can't find a value.	
?	Not entered	No value has been assigned yet (and it might not be applicable).	
~	Insufficient sample	Insufficient sample for analysis.	-94
D	Uncertain value	Both $ and *apply to this value.	
\	Not applicable	Not requested.	-99
B	Estimated value	Both # and A apply to this value.	
A	Estimated value	Value has been set to zero because data conditioning process created an artificial negative value.	
C	Uncertain value	Both $ and A apply to this value.	

Table 5: Geochemistry Database qualifier codes and their LIMS translations

The following should be undertaken:

- for every analyte field create a corresponding qualifier field
- populate the qualifier fields with direct translations of the LIMS codes as indicated in Table 5
- modify the values of analyte field on the basis of the qualifier and check and deal with reported values that are lower than the reported lower detection limit.

  The above three steps in future could be done by creating a macro in MS ACCESS.

Other quality control statements (e.g. data of dubious quality) can be added once the quality control procedure has been completed. The process of levelling (see Section 5.3) may also necessitate changing the qualifier if an artificial negative value has been created (see A, B and C , Table 5).

The rules of G-BASE data conditioning are summarised in the next sub-section.

Detection limits have changed as analytical methods have improved and this is a very important consideration when using data from the Geochemistry Database. It is important, therefore, to maintain a record of the original data as reported by the laboratory and to use the qualifier field to describe how below detection results have been treated. Historical detection limits for data in the Geochemistry Database are discussed by Johnson et al. (2004) and some guidance on preparing pre-2004 analytical data is given. In the raw laboratory data some results reported are below the element LLD quoted in table BGS_DTA_RAW_ANALYTES (see Appendix 2). These data can often show meaningful variations and so only values with LIMS code of '-97' or < 0 will be reset to one-half the recorded lower detection limit..

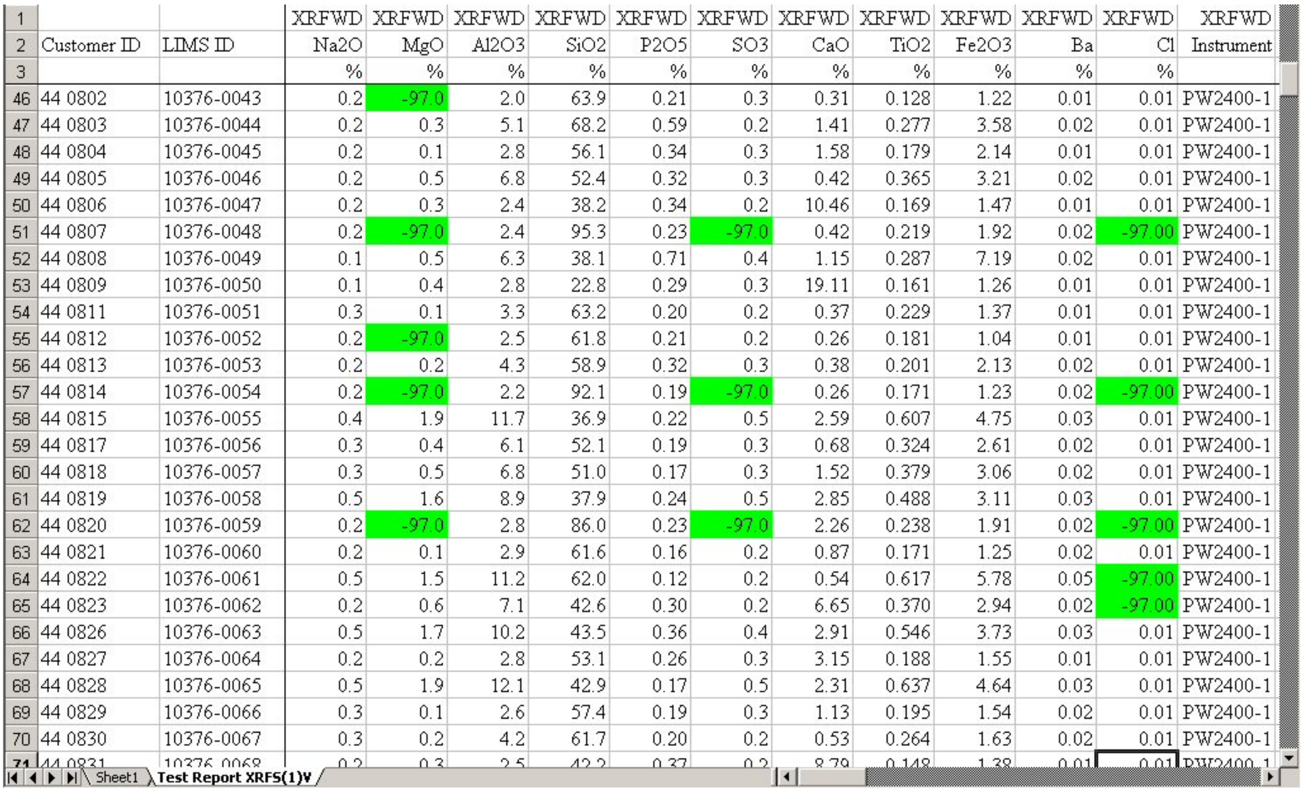
4.3 RULES OF G-BASE DATA CONDITIONING (VERSION 1.0)

1. These rules concern all G-BASE data loaded to the Geochemistry Database from 1st Janaury 2006 and will apply until they are superseded by any revision.
2. Data conditioning is only done on completed data sets signed off by the Laboratory Manager (or their representative) who will also report to the project (via the LIMS system for soil and sediment data) all the associated batch information such as analytical methodology, detection limits and primary reference material analyses.
3. G-BASE data conditioning is the responsibility of the G-BASE Data Manager (Bob Lister for soils and sediments, Louise Ander for stream waters). Loading the conditioned data to the Geochemistry Database is the responsibility of the Geochemistry Database Data Manager (Sue Hobbs).
4. Any data quality issues will be flagged up in the qualifier field associated with an analyte result during the data conditioning process. Users of the data should be aware of the existence of this qualifier field (Table 5) and that other data tables within the Geochemistry Database contain information relevant to the analytical result (e.g. see Appendix 2).

5. All G-BASE geochemical results in the Geochemistry Database may have been subjected to levelling factors, both by the processes of "within-area" and "between-region" levelling (see section 5.3).

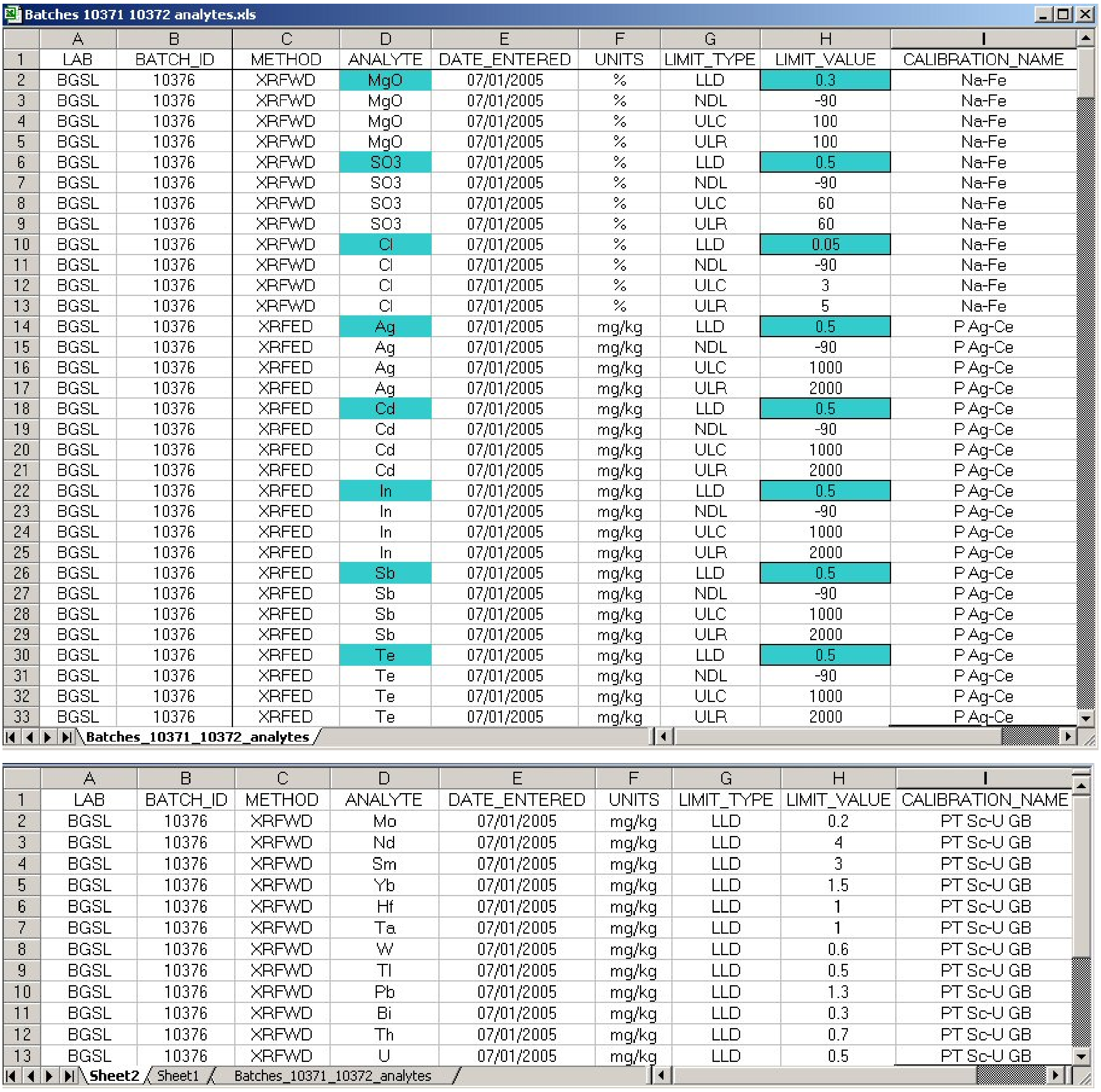
   a. Any value not determined by the analyst should be represented by a "null" value, as should any results considered to be too unsuitable to be included in the database (e.g. -96 code).
   b. Data reported as being less than the lower detection limit represented by a semi-quantitative or qualitative code (e.g. bdl, <5, -97), should have a qualifier ("#") that indicates this fact, and the code will be replaced by a value one-half the reported lower detection limit.
   c. Data reported as being higher than the upper detection limit represented by a semi-quantitative or qualitative code (e.g. > 2000), should have a qualifier (">") that indicates this fact, and the code will be replaced by a value equal to the upper reporting limit.
   d. Results reported by the analyst that are actually below the reported lower limit of detection, and not represented by a semi-quantitative or qualitative code, will be identified by a "$" in the qualifier field. This may include minus or zero values.
   e. All results will be subjected to levelling, including minus, zero and substituted detection limit values. If a result becomes a minus value as a result of the levelling, then it will be reset to zero and "A" inserted in the qualifier field.
   f. The qualifier field can only be represented by a single text character as shown in Table 5. If there is more than one quality issue to be indicated then the most significant issue takes priority in the qualifier field. The exception to this is the use of "B" and "C" which are used to represent a combination of several of the most likely multiple quality issues.

4.4 EXAMPLE TABLES FOR REPLACED MISSING OR SEMI-QUANTITATIVE DATA



1			XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD
2	Customer ID	LIMS ID	Na2O	MgO	Al2O3	SiO2	P2O5	SO3	CaO	TiO2	Fe2O3	Ba	Cl	Instrument
3			%	%	%	%	%	%	%	%	%	%	%	
46	44 0802	10376-0043	0.2	-97.0	2.0	63.9	0.21	0.3	0.31	0.128	1.22	0.01	0.01	PW2400-1
47	44 0803	10376-0044	0.2	0.3	5.1	68.2	0.59	0.2	1.41	0.277	3.58	0.02	0.01	PW2400-1
48	44 0804	10376-0045	0.2	0.1	2.8	56.1	0.34	0.3	1.58	0.179	2.14	0.01	0.01	PW2400-1
49	44 0805	10376-0046	0.2	0.5	6.8	52.4	0.32	0.3	0.42	0.365	3.21	0.02	0.01	PW2400-1
50	44 0806	10376-0047	0.2	0.3	2.4	38.2	0.34	0.2	10.46	0.169	1.47	0.01	0.01	PW2400-1
51	44 0807	10376-0048	0.2	-97.0	2.4	95.3	0.23	-97.0	0.42	0.219	1.92	0.02	-97.00	PW2400-1
52	44 0808	10376-0049	0.1	0.5	6.3	38.1	0.71	0.4	1.15	0.287	7.19	0.02	0.01	PW2400-1
53	44 0809	10376-0050	0.1	0.4	2.8	22.8	0.29	0.3	19.11	0.161	1.26	0.01	0.01	PW2400-1
54	44 0811	10376-0051	0.3	0.1	3.3	63.2	0.20	0.2	0.37	0.229	1.37	0.01	0.01	PW2400-1
55	44 0812	10376-0052	0.2	-97.0	2.5	61.8	0.21	0.2	0.26	0.181	1.04	0.01	0.01	PW2400-1
56	44 0813	10376-0053	0.2	0.2	4.3	58.9	0.32	0.3	0.38	0.201	2.13	0.02	0.01	PW2400-1
57	44 0814	10376-0054	0.2	-97.0	2.2	92.1	0.19	-97.0	0.26	0.171	1.23	0.02	-97.00	PW2400-1
58	44 0815	10376-0055	0.4	1.9	11.7	36.9	0.22	0.5	2.59	0.607	4.75	0.03	0.01	PW2400-1
59	44 0817	10376-0056	0.3	0.4	6.1	52.1	0.19	0.3	0.68	0.324	2.61	0.02	0.01	PW2400-1
60	44 0818	10376-0057	0.3	0.5	6.8	51.0	0.17	0.3	1.52	0.379	3.06	0.02	0.01	PW2400-1
61	44 0819	10376-0058	0.5	1.6	8.9	37.9	0.24	0.5	2.85	0.488	3.11	0.03	0.01	PW2400-1
62	44 0820	10376-0059	0.2	-97.0	2.8	86.0	0.23	-97.0	2.26	0.238	1.91	0.02	-97.00	PW2400-1
63	44 0821	10376-0060	0.2	0.1	2.9	61.6	0.16	0.2	0.87	0.171	1.25	0.02	0.01	PW2400-1
64	44 0822	10376-0061	0.5	1.5	11.2	62.0	0.12	0.2	0.54	0.617	5.78	0.05	-97.00	PW2400-1
65	44 0823	10376-0062	0.2	0.6	7.1	42.6	0.30	0.2	6.65	0.370	2.94	0.02	-97.00	PW2400-1
66	44 0826	10376-0063	0.5	1.7	10.2	43.5	0.36	0.4	2.91	0.546	3.73	0.03	0.01	PW2400-1
67	44 0827	10376-0064	0.2	0.2	2.8	53.1	0.26	0.3	3.15	0.188	1.55	0.01	0.01	PW2400-1
68	44 0828	10376-0065	0.5	1.9	12.1	42.9	0.17	0.5	2.31	0.637	4.64	0.03	0.01	PW2400-1
69	44 0829	10376-0066	0.3	0.1	2.6	57.4	0.19	0.3	1.13	0.195	1.54	0.02	0.01	PW2400-1
70	44 0830	10376-0067	0.3	0.2	4.2	61.7	0.20	0.2	0.53	0.264	1.63	0.02	0.01	PW2400-1
71	44 0831	10376-0068	0.2	0.3	2.5	42.2	0.37	0.2	8.79	0.148	1.38	0.01	0.01	PW2400-1

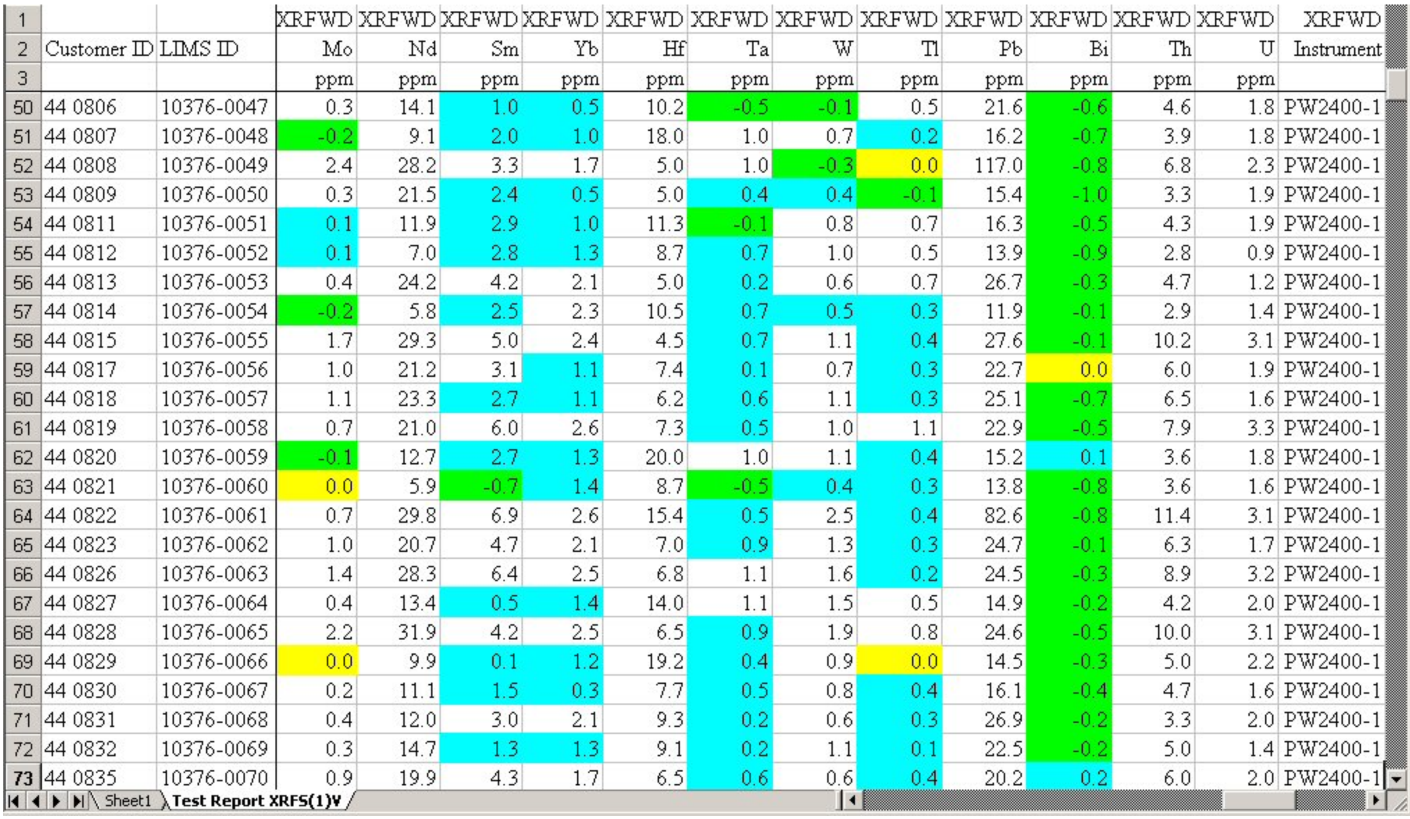
Figure 11: Data from XRFWD Instrument PW2400-1 containing LIMS coded '–97' values



	A	B	C	D	E	F	G	H	I
1	LAB	BATCH_ID	METHOD	ANALYTE	DATE_ENTERED	UNITS	LIMIT_TYPE	LIMIT_VALUE	CALIBRATION_NAME
2	BGSL	10376	XRFWD	MgO	07/01/2005	%	LLD	0.3	Na-Fe
3	BGSL	10376	XRFWD	MgO	07/01/2005	%	NDL	-90	Na-Fe
4	BGSL	10376	XRFWD	MgO	07/01/2005	%	ULC	100	Na-Fe
5	BGSL	10376	XRFWD	MgO	07/01/2005	%	ULR	100	Na-Fe
6	BGSL	10376	XRFWD	SO3	07/01/2005	%	LLD	0.5	Na-Fe
7	BGSL	10376	XRFWD	SO3	07/01/2005	%	NDL	-90	Na-Fe
8	BGSL	10376	XRFWD	SO3	07/01/2005	%	ULC	60	Na-Fe
9	BGSL	10376	XRFWD	SO3	07/01/2005	%	ULR	60	Na-Fe
10	BGSL	10376	XRFWD	Cl	07/01/2005	%	LLD	0.05	Na-Fe
11	BGSL	10376	XRFWD	Cl	07/01/2005	%	NDL	-90	Na-Fe
12	BGSL	10376	XRFWD	Cl	07/01/2005	%	ULC	3	Na-Fe
13	BGSL	10376	XRFWD	Cl	07/01/2005	%	ULR	5	Na-Fe
14	BGSL	10376	XRFED	Ag	07/01/2005	mg/kg	LLD	0.5	P Ag-Ce
15	BGSL	10376	XRFED	Ag	07/01/2005	mg/kg	NDL	-90	P Ag-Ce
16	BGSL	10376	XRFED	Ag	07/01/2005	mg/kg	ULC	1000	P Ag-Ce
17	BGSL	10376	XRFED	Ag	07/01/2005	mg/kg	ULR	2000	P Ag-Ce
18	BGSL	10376	XRFED	Cd	07/01/2005	mg/kg	LLD	0.5	P Ag-Ce
19	BGSL	10376	XRFED	Cd	07/01/2005	mg/kg	NDL	-90	P Ag-Ce
20	BGSL	10376	XRFED	Cd	07/01/2005	mg/kg	ULC	1000	P Ag-Ce
21	BGSL	10376	XRFED	Cd	07/01/2005	mg/kg	ULR	2000	P Ag-Ce
22	BGSL	10376	XRFED	In	07/01/2005	mg/kg	LLD	0.5	P Ag-Ce
23	BGSL	10376	XRFED	In	07/01/2005	mg/kg	NDL	-90	P Ag-Ce
24	BGSL	10376	XRFED	In	07/01/2005	mg/kg	ULC	1000	P Ag-Ce
25	BGSL	10376	XRFED	In	07/01/2005	mg/kg	ULR	2000	P Ag-Ce
26	BGSL	10376	XRFED	Sb	07/01/2005	mg/kg	LLD	0.5	P Ag-Ce
27	BGSL	10376	XRFED	Sb	07/01/2005	mg/kg	NDL	-90	P Ag-Ce
28	BGSL	10376	XRFED	Sb	07/01/2005	mg/kg	ULC	1000	P Ag-Ce
29	BGSL	10376	XRFED	Sb	07/01/2005	mg/kg	ULR	2000	P Ag-Ce
30	BGSL	10376	XRFED	Te	07/01/2005	mg/kg	LLD	0.5	P Ag-Ce
31	BGSL	10376	XRFED	Te	07/01/2005	mg/kg	NDL	-90	P Ag-Ce
32	BGSL	10376	XRFED	Te	07/01/2005	mg/kg	ULC	1000	P Ag-Ce
33	BGSL	10376	XRFED	Te	07/01/2005	mg/kg	ULR	2000	P Ag-Ce

	A	B	C	D	E	F	G	H	I
1	LAB	BATCH_ID	METHOD	ANALYTE	DATE_ENTERED	UNITS	LIMIT_TYPE	LIMIT_VALUE	CALIBRATION_NAME
2	BGSL	10376	XRFWD	Mo	07/01/2005	mg/kg	LLD	0.2	PT Sc-U GB
3	BGSL	10376	XRFWD	Nd	07/01/2005	mg/kg	LLD	4	PT Sc-U GB
4	BGSL	10376	XRFWD	Sm	07/01/2005	mg/kg	LLD	3	PT Sc-U GB
5	BGSL	10376	XRFWD	Yb	07/01/2005	mg/kg	LLD	1.5	PT Sc-U GB
6	BGSL	10376	XRFWD	Hf	07/01/2005	mg/kg	LLD	1	PT Sc-U GB
7	BGSL	10376	XRFWD	Ta	07/01/2005	mg/kg	LLD	1	PT Sc-U GB
8	BGSL	10376	XRFWD	W	07/01/2005	mg/kg	LLD	0.6	PT Sc-U GB
9	BGSL	10376	XRFWD	Tl	07/01/2005	mg/kg	LLD	0.5	PT Sc-U GB
10	BGSL	10376	XRFWD	Pb	07/01/2005	mg/kg	LLD	1.3	PT Sc-U GB
11	BGSL	10376	XRFWD	Bi	07/01/2005	mg/kg	LLD	0.3	PT Sc-U GB
12	BGSL	10376	XRFWD	Th	07/01/2005	mg/kg	LLD	0.7	PT Sc-U GB
13	BGSL	10376	XRFWD	U	07/01/2005	mg/kg	LLD	0.5	PT Sc-U GB

Figure 12: Lower Limit of Detection Values (LLD) from BGS_DTA_RAW_ANALYTES table (shown for various instruments and (in the lower table) LLD only selected)



1			XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD	XRFWD
2	Customer ID	LIMS ID	Mo	Nd	Sm	Yb	Hf	Ta	W	Tl	Pb	Bi	Th	U	Instrument
3			ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm	
50	44 0806	10376-0047	0.3	14.1	1.0	0.5	10.2	-0.5	-0.1	0.5	21.6	-0.6	4.6	1.8	PW2400-1
51	44 0807	10376-0048	-0.2	9.1	2.0	1.0	18.0	1.0	0.7	0.2	16.2	-0.7	3.9	1.8	PW2400-1
52	44 0808	10376-0049	2.4	28.2	3.3	1.7	5.0	1.0	-0.3	0.0	117.0	-0.8	6.8	2.3	PW2400-1
53	44 0809	10376-0050	0.3	21.5	2.4	0.5	5.0	0.4	0.4	-0.1	15.4	-1.0	3.3	1.9	PW2400-1
54	44 0811	10376-0051	0.1	11.9	2.9	1.0	11.3	-0.1	0.8	0.7	16.3	-0.5	4.3	1.9	PW2400-1
55	44 0812	10376-0052	0.1	7.0	2.8	1.3	8.7	0.7	1.0	0.5	13.9	-0.9	2.8	0.9	PW2400-1
56	44 0813	10376-0053	0.4	24.2	4.2	2.1	5.0	0.2	0.6	0.7	26.7	-0.3	4.7	1.2	PW2400-1
57	44 0814	10376-0054	-0.2	5.8	2.5	2.3	10.5	0.7	0.5	0.3	11.9	-0.1	2.9	1.4	PW2400-1
58	44 0815	10376-0055	1.7	29.3	5.0	2.4	4.5	0.7	1.1	0.4	27.6	-0.1	10.2	3.1	PW2400-1
59	44 0817	10376-0056	1.0	21.2	3.1	1.1	7.4	0.1	0.7	0.3	22.7	0.0	6.0	1.9	PW2400-1
60	44 0818	10376-0057	1.1	23.3	2.7	1.1	6.2	0.6	1.1	0.3	25.1	-0.7	6.5	1.6	PW2400-1
61	44 0819	10376-0058	0.7	21.0	6.0	2.6	7.3	0.5	1.0	1.1	22.9	-0.5	7.9	3.3	PW2400-1
62	44 0820	10376-0059	-0.1	12.7	2.7	1.3	20.0	1.0	1.1	0.4	15.2	0.1	3.6	1.8	PW2400-1
63	44 0821	10376-0060	0.0	5.9	-0.7	1.4	8.7	-0.5	0.4	0.3	13.8	-0.8	3.6	1.6	PW2400-1
64	44 0822	10376-0061	0.7	29.8	6.9	2.6	15.4	0.5	2.5	0.4	82.6	-0.8	11.4	3.1	PW2400-1
65	44 0823	10376-0062	1.0	20.7	4.7	2.1	7.0	0.9	1.3	0.3	24.7	-0.1	6.3	1.7	PW2400-1
66	44 0826	10376-0063	1.4	28.3	6.4	2.5	6.8	1.1	1.6	0.2	24.5	-0.3	8.9	3.2	PW2400-1
67	44 0827	10376-0064	0.4	13.4	0.5	1.4	14.0	1.1	1.5	0.5	14.9	-0.2	4.2	2.0	PW2400-1
68	44 0828	10376-0065	2.2	31.9	4.2	2.5	6.5	0.9	1.9	0.8	24.6	-0.5	10.0	3.1	PW2400-1
69	44 0829	10376-0066	0.0	9.9	0.1	1.2	19.2	0.4	0.9	0.0	14.5	-0.3	5.0	2.2	PW2400-1
70	44 0830	10376-0067	0.2	11.1	1.5	0.3	7.7	0.5	0.8	0.4	16.1	-0.4	4.7	1.6	PW2400-1
71	44 0831	10376-0068	0.4	12.0	3.0	2.1	9.3	0.2	0.6	0.3	26.9	-0.2	3.3	2.0	PW2400-1
72	44 0832	10376-0069	0.3	14.7	1.3	1.3	9.1	0.2	1.1	0.1	22.5	-0.2	5.0	1.4	PW2400-1
73	44 0835	10376-0070	0.9	19.9	4.3	1.7	6.5	0.6	0.6	0.4	20.2	0.2	6.0	2.0	PW2400-1

Figure 13: Example of trace element data from XRFWD Instrument PW2400-1 containing sub-detection, zero and negative determinations

4.5 VERIFICATION

Verification that the laboratories are producing data that is reliable and accurate needs to be done as soon after the results are received as possible. With "commercial work" taking precedence over "science work" this is an area in which the G-BASE project has been quite negligent and in some instances it has been several years before data has been verified. A more rapid reporting of analyses and better resourcing of the QA/QC procedures should mean that the project verifies the data within one month of receipt of results.

Procedures for verification basically involve a rapid check of the analytical data with Shewhart control plots and extraction of the duplicate/replicate results. Once the data has passed these procedures it can be considered as "verified data" and would be ready for the final quality control procedures described in the next section.

4.5.1 Shewhart control plots

Secondary reference materials (SRMs) are submitted as normal samples and therefore "blind" to the analyst. When the results are returned the SRM results need to be extracted from the database to check that the analyst is reporting results that are close to the accepted values for each element (see section 4.5.1.1). This is done graphically using Shewhart control plots (Shewhart, 1931) which not only give a rapid visual assessment of the accuracy of the data but also identfies any analytical variance over a period of time. Element data for the SRMs is plotted and if the result falls outside the mean accepted value of ± 3 standard deviations, or two or more values fall outside the ± 2 standard deviation limit, then the analytical data is rejected. The time series data are also monitored for drift, shift and bias (see Glossary of terms). If results for 10 out of 11 consecutive analytical batches fall one side of the mean, or if 8 consecutive results successively rise or fall, necessary quality control procedures need to be implemented (e.g. contacting laboratories for explanation of drift; levelling batch data to correct for shift; reassessment of the accepted standard means).

G-BASE currently uses the simple process of plotting a time series chart in MS Excel (see Figure 14), though the BGS laboratories have a licence to use more sophisticated control plotting software (QI Analyst[1]**)** that has been used on non-G-BASE geochemical surveys (see Figure 15). In 2005 the Tellus geochemical mapping project in N Ireland was using SPC XL2000[2] (Digital Computations Inc.), an add-in Module for MS Excel and this is currently under evaluation for use in the G-BASE project (Figure 16).

If the analytical instruments are performing to the optimum level and have been calibrated/recalibrated correctly, the control graphs should display little variation in element concentration throughout the period of analysis. However, in reality the control graphs will show identifiable shifts in the data. This is often coincident with a recalibration event that has been necessary due to instrument breakdown or after instrument service/overhaul. In the case of Wales and Welsh Borders area samples, several recalibrations of the analytical instruments were carried out during the entire period of analysis of approximately 21,500 samples (Figure 14). A total of 49 individual laboratory numbers constitute the total number of samples analysed for this particular project over a 3-year period. Similar shifts in results seen in the East Anglian data (Figure 16) show relationships to changes in instruments and calibrations.

Shewhart control parts are also used later in the QC process (see Section 5.3) to identify groups of laboratory batches where results need to be levelled with reference to standard materials previously analysed.

---

1 [1] http://www.wonderware.com/products/qianalyst/

2 [2] http://www.sigmazone.com/spcxl.htm

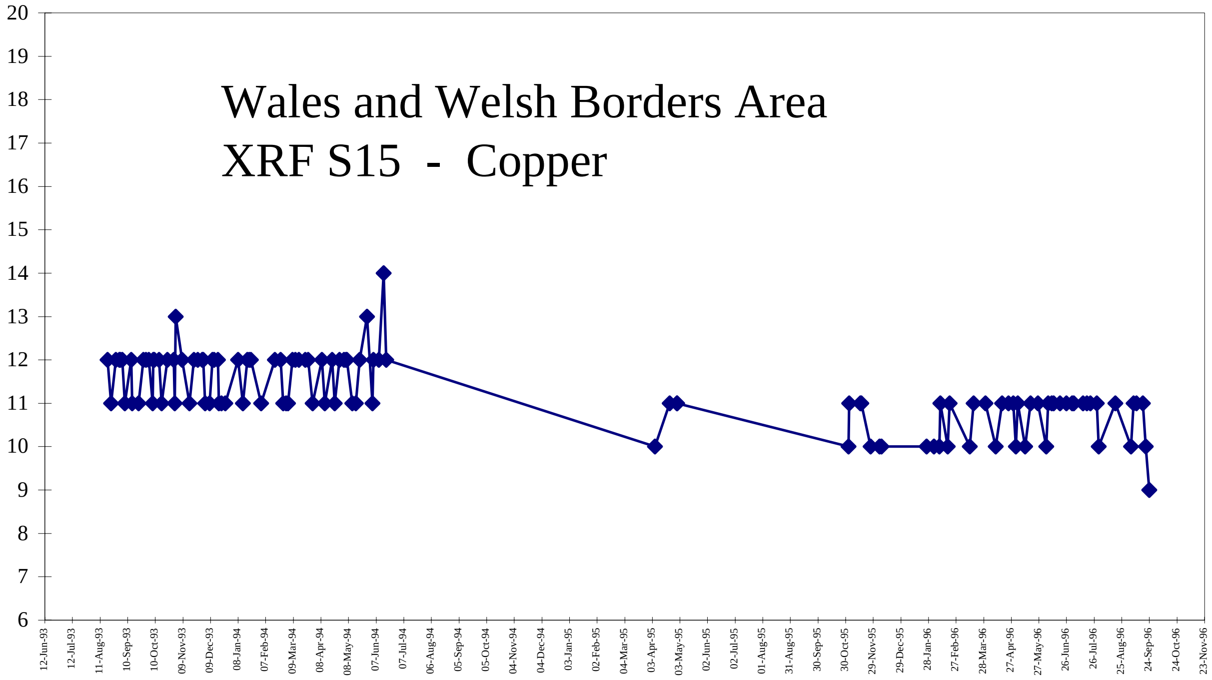
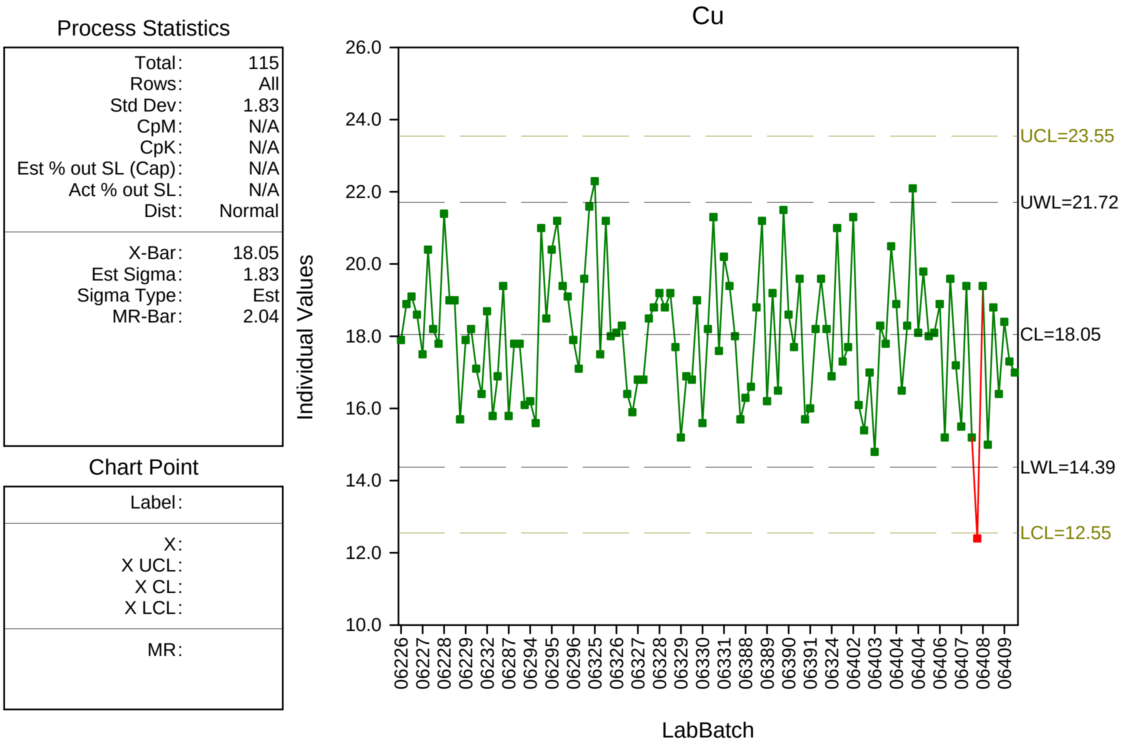


Figure 14: Excel generated control plot for XRF data. S15 – Copper (y-axis in ppm)


Process Statistics

Total: 115
Rows: All
Std Dev: 1.83
CpM: N/A
CpK: N/A
Est % out SL (Cap): N/A
Act % out SL: N/A
Dist: Normal

X-Bar: 18.05
Est Sigma: 1.83
Sigma Type: Est
MR-Bar: 2.04

Chart Point

Label:

X:
X UCL:
X CL:
X LCL:

MR:

Figure 15: Control chart plot using QI Analyst software (Cu in ppm on y-axis)

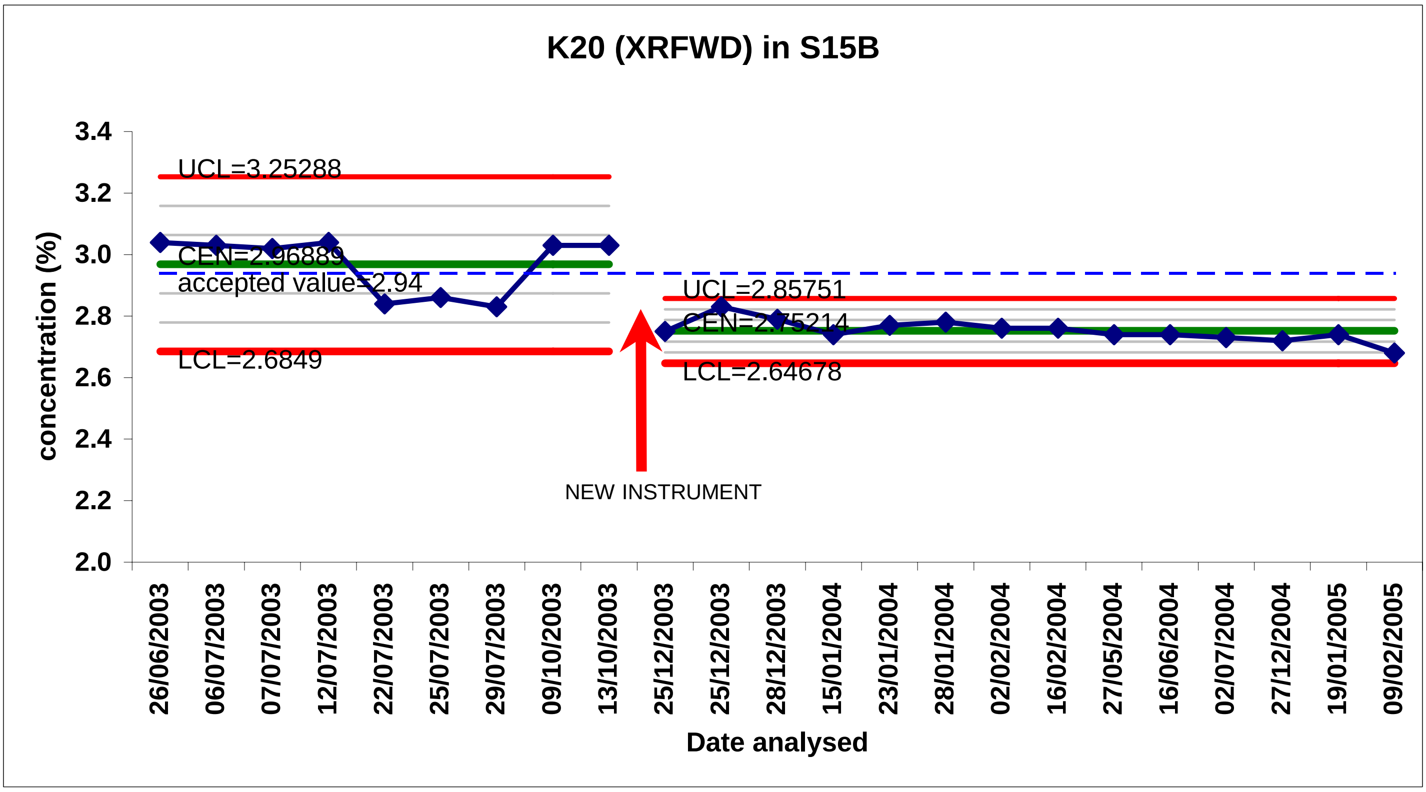
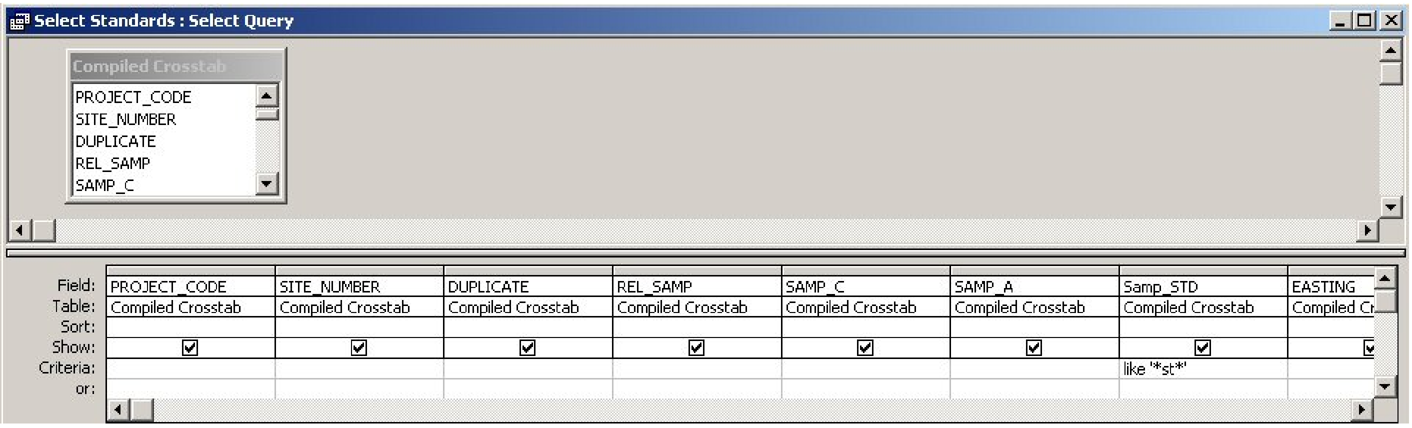


Figure 16: Example of output from the SPC XL control chart plotting MS Excel add-in (from the G-BASE East Anglian SRMs).

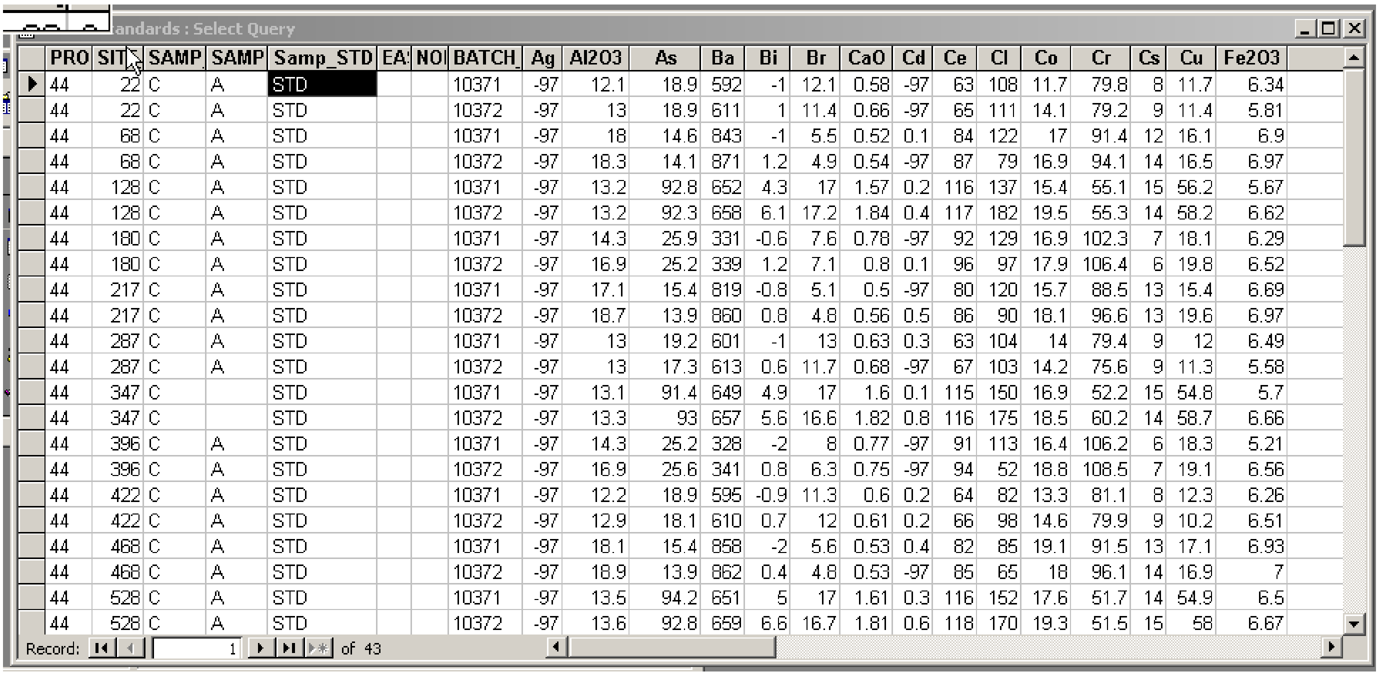
4.5.1.1 SELECTING SRM DATA FOR SHEWHART CONTROL PLOTS

Queries to select secondary standard materials are shown below. The important retrieval criterion here is to select all the samples that have "STD" in the SAMP_STD field.



Field:	PROJECT_CODE	SITE_NUMBER	DUPLICATE	REL_SAMP	SAMP_C	SAMP_A	Samp_STD	EASTING
Table:	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Cr
Sort:								
Show:	☑	☑	☑	☑	☑	☑	☑	☑
Criteria:							like '*st*'	
or:								

Figure 17: Design view select query to select secondary reference materials from the compiled field and analytical data (derived from the query illustrated in Figure 10)



PRO	SIT	SAMP	SAMP	Samp_STD	EA	NO	BATCH_	Ag	Al2O3	As	Ba	Bi	Br	CaO	Cd	Ce	Cl	Co	Cr	Cs	Cu	Fe2O3
44	22	C	A	STD			10371	-97	12.1	18.9	592	-1	12.1	0.58	-97	63	108	11.7	79.8	8	11.7	6.34
44	22	C	A	STD			10372	-97	13	18.9	611	1	11.4	0.66	-97	65	111	14.1	79.2	9	11.4	5.81
44	68	C	A	STD			10371	-97	18	14.6	843	-1	5.5	0.52	0.1	84	122	17	91.4	12	16.1	6.9
44	68	C	A	STD			10372	-97	18.3	14.1	871	1.2	4.9	0.54	-97	87	79	16.9	94.1	14	16.5	6.97
44	128	C	A	STD			10371	-97	13.2	92.8	652	4.3	17	1.57	0.2	116	137	15.4	55.1	15	56.2	5.67
44	128	C	A	STD			10372	-97	13.2	92.3	658	6.1	17.2	1.84	0.4	117	182	19.5	55.3	14	58.2	6.62
44	180	C	A	STD			10371	-97	14.3	25.9	331	-0.6	7.6	0.78	-97	92	129	16.9	102.3	7	18.1	6.29
44	180	C	A	STD			10372	-97	16.9	25.2	339	1.2	7.1	0.8	0.1	96	97	17.9	106.4	6	19.8	6.52
44	217	C	A	STD			10371	-97	17.1	15.4	819	-0.8	5.1	0.5	-97	80	120	15.7	88.5	13	15.4	6.69
44	217	C	A	STD			10372	-97	18.7	13.9	860	0.8	4.8	0.56	0.5	86	90	18.1	96.6	13	19.6	6.97
44	287	C	A	STD			10371	-97	13	19.2	601	-1	13	0.63	0.3	63	104	14	79.4	9	12	6.49
44	287	C	A	STD			10372	-97	13	17.3	613	0.6	11.7	0.68	-97	67	103	14.2	75.6	9	11.3	5.58
44	347	C		STD			10371	-97	13.1	91.4	649	4.9	17	1.6	0.1	115	150	16.9	52.2	15	54.8	5.7
44	347	C		STD			10372	-97	13.3	93	657	5.6	16.6	1.82	0.8	116	175	18.5	60.2	14	58.7	6.66
44	396	C	A	STD			10371	-97	14.3	25.2	328	-2	8	0.77	-97	91	113	16.4	106.2	6	18.3	5.21
44	396	C	A	STD			10372	-97	16.9	25.6	341	0.8	6.3	0.75	-97	94	52	18.8	108.5	7	19.1	6.56
44	422	C	A	STD			10371	-97	12.2	18.9	595	-0.9	11.3	0.6	0.2	64	82	13.3	81.1	8	12.3	6.26
44	422	C	A	STD			10372	-97	12.9	18.1	610	0.7	12	0.61	0.2	66	98	14.6	79.9	9	10.2	6.51
44	468	C	A	STD			10371	-97	18.1	15.4	858	-2	5.6	0.53	0.4	82	85	19.1	91.5	13	17.1	6.93
44	468	C	A	STD			10372	-97	18.9	13.9	862	0.4	4.8	0.53	-97	85	65	18	96.1	14	16.9	7
44	528	C	A	STD			10371	-97	13.5	94.2	651	5	17	1.61	0.3	116	152	17.6	51.7	14	54.9	6.5
44	528	C	A	STD			10372	-97	13.6	92.8	659	6.6	16.7	1.81	0.6	118	170	19.3	51.5	15	58	6.67

Figure 18: Example of data retrieved from the query illustrated in Figure 17

The different standards used can be identified by the list of standard samples maintained as a MS ACCESS table for all batches submitted since 2004 (see Figure 19). Note that until 2003 soils and sediments where submitted as mixed analytical batches and data conditioning was carried out on soils and sediments at the same time. Since 2004 soil and sediment are always submitted as separate batches and are therefore analysed at different times necessitating separate verification.

SITENO	SAMPLE_	SAMP_STD	STANDARD	LabNo
1487	C	STD	S57A	10377
1547	C	STD	S13	10377
1596	C	STD	S13B	10377
1622	C	STD	S15B	10377
1668	C	STD	S23B	10377
1728	C	STD	S57A	10377
1780	C	STD	S13	10377
1817	C	STD	S13B	10377
1887	C	STD	S15B	10377
1947	C	STD	S23B	10377
1996	C	STD	S57A	10377
2022	C	STD	S13	10377
2068	C	STD	S13B	10377
2128	C	STD	S15B	10377
2180	C	STD	S23B	10377
2217	C	STD	S57A	10377
2287	C	STD	S13	10377
2422	C	STD	S15B	10586
2468	C	STD	S23B	10586
2747	C	STD	S23B	10586
2796	C	STD	S57A	10586
2822	C	STD	S13	10586
2868	C	STD	S13B	10586
2928	C	STD	S15B	10586
2980	C	STD	S23B	10586

Figure 19: A sample listing of the site number - STD translation available as MS ACCESS table

4.5.2 Selecting duplicate and replicate results

After completing initial Shewhart plots a further verification of the data is achieved by looking at the duplicate/replicate pairs to see if results are similar. Although there can be significant "within" site variability it should be anticipated that each set of duplicate/replicate analyses will have generally similar results. Before using duplicate-replicate pairs as part of a rigorous quality control procedure to identify sampling or laboratory error, it is important to verify first that the data contains no errors caused by sample mis-numbering or errors in the analytical listing. This can be achieved by a quick inspection of the results for the duplicate-replicate pairs. The relationship between these control samples is illustrated in Figure 20.

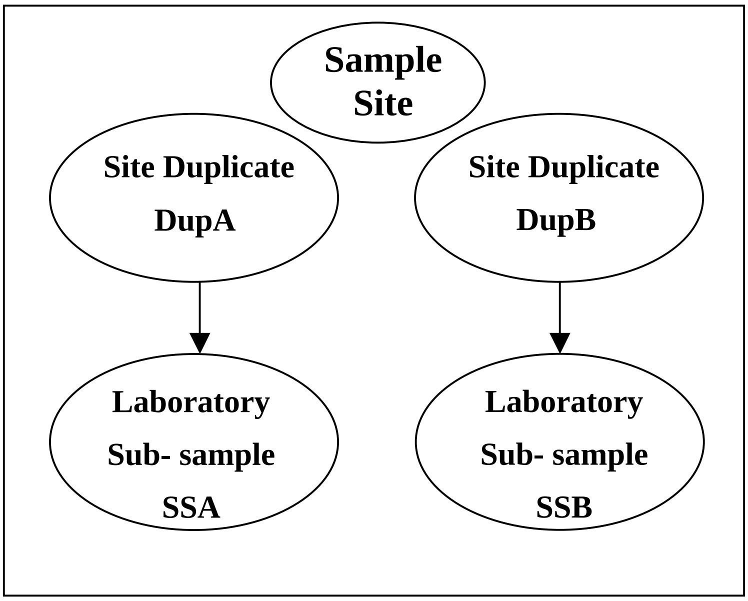
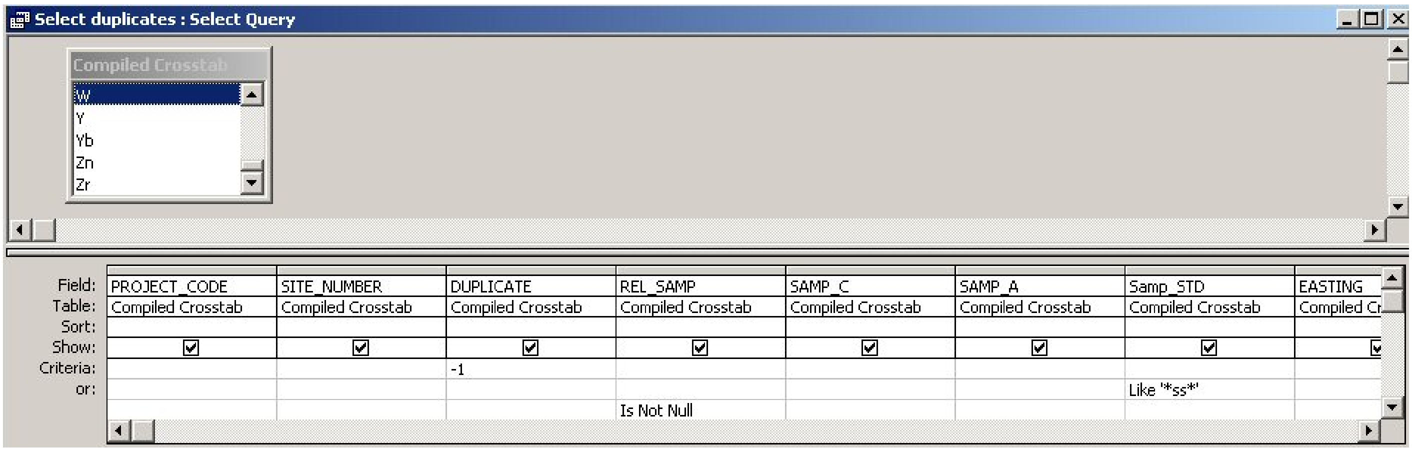


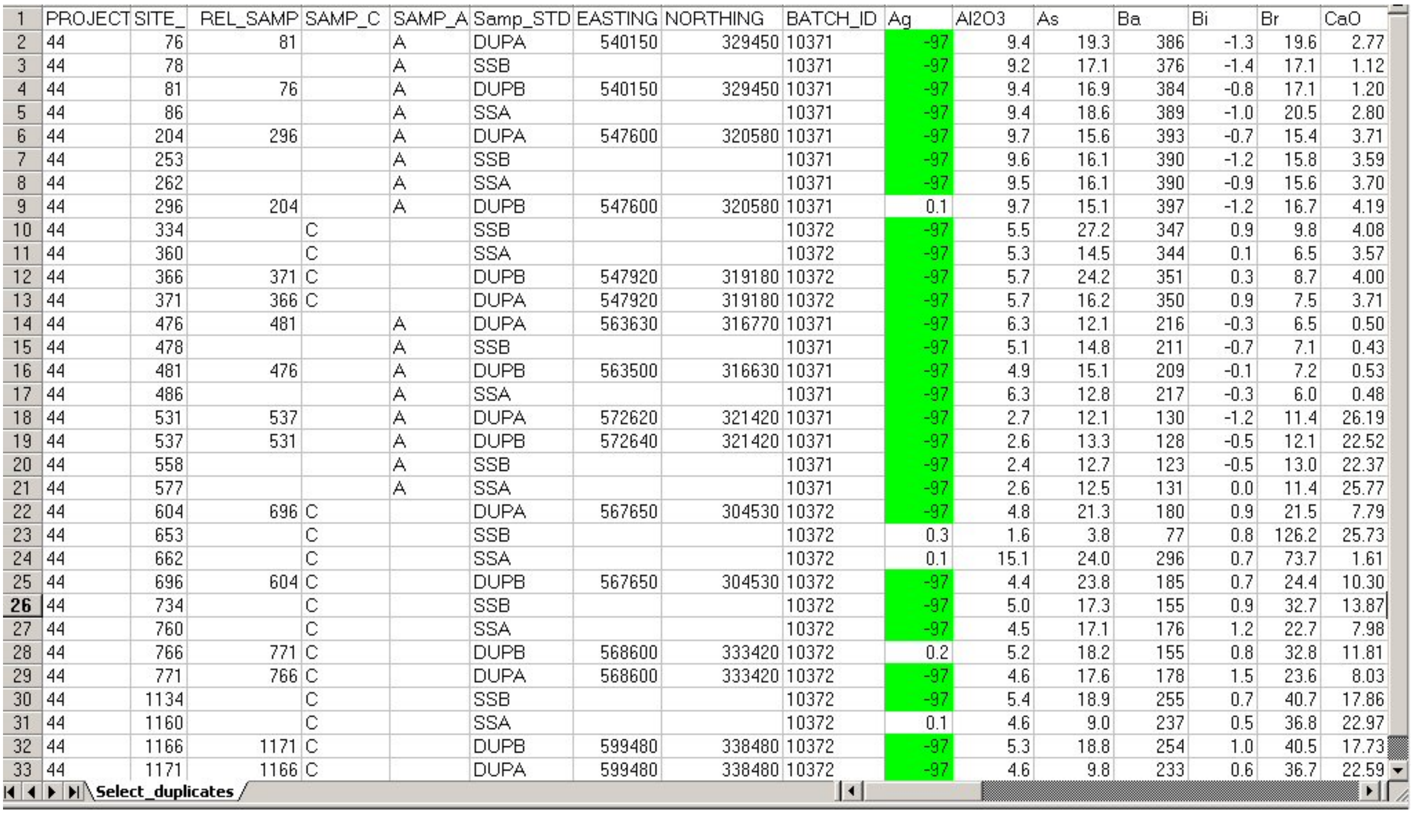
Figure 20: Relationship between site duplicates and laboratory replicates

Duplicate and replicate results can be compared by scanning tabulated data. More detailed analysis is obtained from x-y plots as described in the next section on quality control. Data is retrieved by selecting samples on the basis that the DUPLICATE field = -1, REL_SAMP field is not "" (null) and the SAMP_STD field contains "SS" as shown by the retrieval in Figure 21.



Field:	PROJECT_CODE	SITE_NUMBER	DUPLICATE	REL_SAMP	SAMP_C	SAMP_A	Samp_STD	EASTING
Table:	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Crosstab	Compiled Cr
Sort:								
Show:	☑	☑	☑	☑	☑	☑	☑	☑
Criteria:			-1					
or:							Like "*ss*"	
				Is Not Null				

Figure 21: A design view of a query for selecting duplicate and replicate samples



1	PROJECT	SITE_	REL_SAMP	SAMP_C	SAMP_A	Samp_STD	EASTING	NORTHING	BATCH_ID	Ag	Al2O3	As	Ba	Bi	Br	CaO
2	44	76	81		A	DUPA	540150	329450	10371	-97	9.4	19.3	386	-1.3	19.6	2.77
3	44	78			A	SSB			10371	-97	9.2	17.1	376	-1.4	17.1	1.12
4	44	81	76		A	DUPB	540150	329450	10371	-97	9.4	16.9	384	-0.8	17.1	1.20
5	44	86			A	SSA			10371	-97	9.4	18.6	389	-1.0	20.5	2.80
6	44	204	296		A	DUPA	547600	320580	10371	-97	9.7	15.6	393	-0.7	15.4	3.71
7	44	253			A	SSB			10371	-97	9.6	16.1	390	-1.2	15.8	3.59
8	44	262			A	SSA			10371	-97	9.5	16.1	390	-0.9	15.6	3.70
9	44	296	204		A	DUPB	547600	320580	10371	0.1	9.7	15.1	397	-1.2	16.7	4.19
10	44	334		C		SSB			10372	-97	5.5	27.2	347	0.9	9.8	4.08
11	44	360		C		SSA			10372	-97	5.3	14.5	344	0.1	6.5	3.57
12	44	366	371	C		DUPB	547920	319180	10372	-97	5.7	24.2	351	0.3	8.7	4.00
13	44	371	366	C		DUPA	547920	319180	10372	-97	5.7	16.2	350	0.9	7.5	3.71
14	44	476	481		A	DUPA	563630	316770	10371	-97	6.3	12.1	216	-0.3	6.5	0.50
15	44	478			A	SSB			10371	-97	5.1	14.8	211	-0.7	7.1	0.43
16	44	481	476		A	DUPB	563500	316630	10371	-97	4.9	15.1	209	-0.1	7.2	0.53
17	44	486			A	SSA			10371	-97	6.3	12.8	217	-0.3	6.0	0.48
18	44	531	537		A	DUPA	572620	321420	10371	-97	2.7	12.1	130	-1.2	11.4	26.19
19	44	537	531		A	DUPB	572640	321420	10371	-97	2.6	13.3	128	-0.5	12.1	22.52
20	44	558			A	SSB			10371	-97	2.4	12.7	123	-0.5	13.0	22.37
21	44	577			A	SSA			10371	-97	2.6	12.5	131	0.0	11.4	25.77
22	44	604	696	C		DUPA	567650	304530	10372	-97	4.8	21.3	180	0.9	21.5	7.79
23	44	653		C		SSB			10372	0.3	1.6	3.8	77	0.8	126.2	25.73
24	44	662		C		SSA			10372	0.1	15.1	24.0	296	0.7	73.7	1.61
25	44	696	604	C		DUPB	567650	304530	10372	-97	4.4	23.8	185	0.7	24.4	10.30
26	44	734		C		SSB			10372	-97	5.0	17.3	155	0.9	32.7	13.87
27	44	760		C		SSA			10372	-97	4.5	17.1	176	1.2	22.7	7.98
28	44	766	771	C		DUPB	568600	333420	10372	0.2	5.2	18.2	155	0.8	32.8	11.81
29	44	771	766	C		DUPA	568600	333420	10372	-97	4.6	17.6	178	1.5	23.6	8.03
30	44	1134		C		SSB			10372	-97	5.4	18.9	255	0.7	40.7	17.86
31	44	1160		C		SSA			10372	0.1	4.6	9.0	237	0.5	36.8	22.97
32	44	1166	1171	C		DUPB	599480	338480	10372	-97	5.3	18.8	254	1.0	40.5	17.73
33	44	1171	1166	C		DUPA	599480	338480	10372	-97	4.6	9.8	233	0.6	36.7	22.59

Figure 22: An Excel spread sheet derived from exporting the select query to select site duplicates and replicates (sub-samples) from the compiled field and analytical data (Figure 21).

For a small number of duplicate/replicate pairs it is sufficient for verification just to look at the control groups in a data table to see if analytical results are consistent. For a larger number of duplicate/replicate pairs x-y plots can be done to look at any deviation from a line of gradient 1.

G-BASE uses MS Excel to create x-y plots, though the data that are listed in rows as shown in Figure 22 needs to be reformatted into columns as shown in Figure 23. This can be done automatically using an Excel macro (Appendix 3) originally used in the Morocco Geochemistry Project. However, use of this macro depends on the strict application of G-BASE control procedures by including four duplicate/replicate samples in every field batch of a hundred samples and the reserved control site numbers are used in a consistent way (these numbers are used by the macro to identify which control sample it is).

Li	DupB	RepB	DupA	RepA	LabBatch	FieldBatch
	36	41	46	42	6226	1- 100
	52	52	49	51	6226	101- 200
	11	12	11	15	6226	201- 300
	33	33	40	32	6226	301- 400
	28	28	28	27	6226	401- 500
	42	43	46	44	6227	501- 600
	17	22	22	18	6227	601- 700
	19	18	18	20	6227	701- 800
	30	23	31	31	6227	801- 900
	37	37	37	36	6227	901- 1000
	30	29	31	29	6228	1001- 1100
	23	22	22	23	6228	1101- 1200
	42	44	49	48	6228	1201- 1300
	36	35	34	39	6228	1301- 1400
	26	27	27	26	6228	1401- 1500
	27	29	29	27	6229	1501- 1600
	34	33	32	33	6229	1601- 1700
	37	39	39	37	6229	1701- 1800
	23	22	23	24	6229	1801- 1900
	47	47	45	46	6229	1901- 2000
	31	34	36	29	6232	2001- 2100
	26	27	28	28	6267	2101- 2200
	29	30	31	31	6267	2201- 2300
	37	38	38	38	6267	2301- 2400
	34	33	33	33	6267	2401- 2500
	33	32	32	33	6287	2501- 2600
	25	29	28	26	6287	2601- 2700
	26	25	26	26	6287	2701- 2800
	14	15	14	12	6287	2801- 2900
	46	46	47	47	6287	2901- 3000
	31	20	20	19	6294	3001- 3100
	22	23	21	21	6294	3101- 3200
	37	36	42	37	6294	3201- 3300
	40	39	40	40	6294	3301- 3400
	21	20	20	19	6294	3401- 3500
	29	27	26	31	6295	3501- 3600
	21	22	22	21	6295	3601- 3700
	53	54	54	49	6295	3701- 3800
	30	27	26	29	6295	3801- 3900
	34	35	34	33	6295	3901- 4000
	35	35	36	36	6296	4001- 4100
	38	38	39	41	6296	4101- 4200
	23	18	17	16	6296	4201- 4300
	51	51	49	50	6296	4301- 4400
	33	39	38	32	6296	4401- 4500

Figure 23: Duplicate/replicate results reformatted into columns ready for x-y plots (prepared using the an Excel macro)

4.5.3 Duplicate pair and Duplicate-Replicate plots

A plot of many duplicate/replicate pairs will enable the identification of any systematic errors over a period of time but will require a spread of concentrations for any particular element over a range of different samples. Duplicate/replicate plots require at least five pairs to make plotting the data worthwhile. Plots of DUPA v DUPB, DUPA v SSA and DUPB v SSB can be produced in an automated way using a MS Excel macro (Appendix 3). An example plot is given in **Figure 24**. Ideally the points should lie on a straight line of gradient one and passing through the plot origin. Significant deviations from this line should be investigated and if no suitable explanation can be given for large errors then the analytical batch should be rejected. Any mismatches arising from the duplicate/replicate plots should be resolved prior to doing the ANOVA analysis described in Section 5.1.

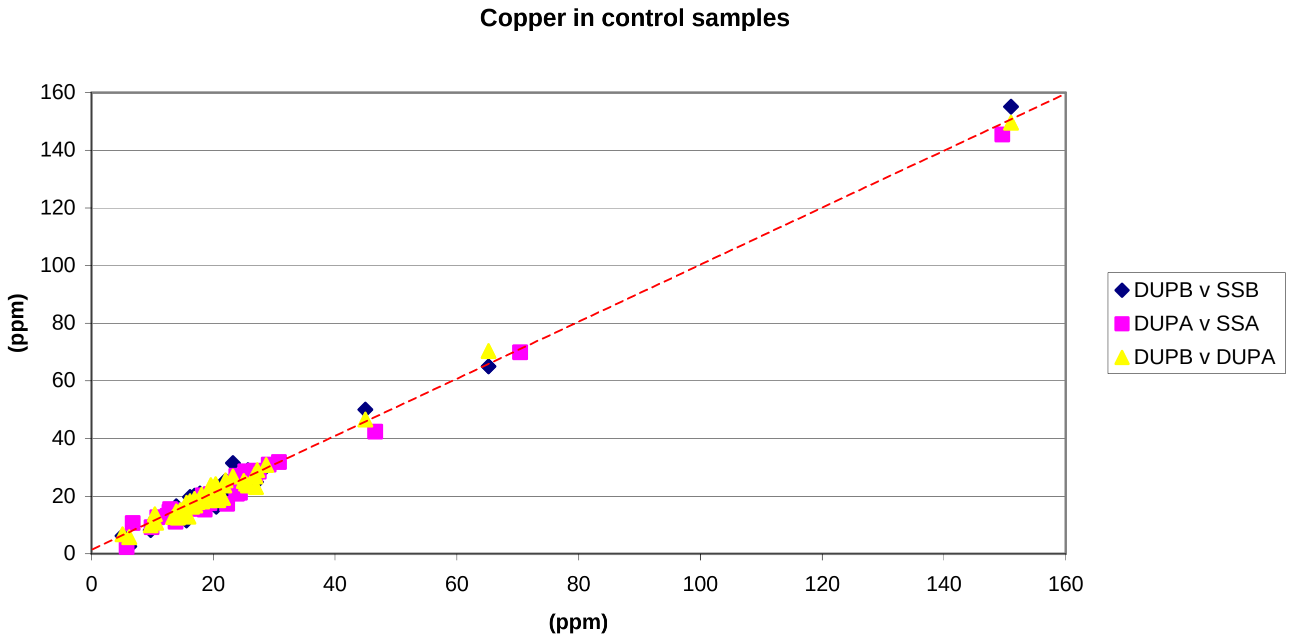


Figure 24: A single plot of DUPA v DUPB, DUPA v SSA and DUPB v SSB

It is possible to work out the variation of standard deviation over a range of element concentrations providing a sufficient number of replicate analyses are done. Thompson and Howarth (1973) discuss the theory behind the estimation and control of precision by duplicate determinations and this work is summarised by Thompson and Howarth (1978). G-BASE does not routinely calculate the analytical precision from the replicate analyses using the graphical methods described in these references.

5 Quality Control

Quality control procedures covered here are those carried out by the G-BASE project using control samples inserted in analytical batches that are "blind" to the analysts, i.e. duplicates, replicates and secondary reference materials. These procedures do not cover the analysis of the international reference standards that are used by the laboratory for its own quality control procedures. Results from the G-BASE quality control measures should be reported to the laboratories as soon after the results are received. Taking measures to implement the results of the quality control are an important part of the quality assurance procedures described in Section 6.

5.1 ANOVA

The duplicate and replicate sample pairs can be used to give an estimate of sampling and analytical errors using statistical analysis known as analysis of variance (ANOVA). A random nested model of ANOVA is selected, since all the analyses form part of a single randomised dataset (Snedecor and Cochran, 1989). This statistical analysis can be carried out using an MS Excel macro as detailed in Johnson (2002) (Appendix 3).

Within-sample variance (represented by multiple components including inhomogeneities introduced during sample handling and preparation, and analytical errors), between-sample variance (representing within-site variation and any variation incorporated during collection of the sample) and between-site variance (the natural distribution of elements) can be estimated. Because the frequency distribution of most elements is multi-modal and does not fit the Gaussian model perfectly, there is an unquantifiable overstatement of the between-site variance - a problem that is inherent in using ANOVA on geochemical data.

Table **6** shows the percentage of variance attributable to each of the three components of variance described above taken from the analysis of the data from eleven urban centres in England and Wales (Lister, 2002b). This gives a good indication as to the integrity of the sampling methodology. If the within sample and between sample variability is greater than the between site variability then the sampling strategy should be considered unsuitable.

5.2 CUMULATIVE PROBABILITY PLOTS

Cumulative probability plots (Sinclair, 1976) are another useful graphical method of studying the variability of geochemical results. In particular, the censoring of reported data and more realistic detection limits can be recognised. However, although they have been used on some international geochemical mapping projects, these are not routinely used in the G-BASE quality control procedures and will not be discussed further.

Element	Between Site (%)	Between Sample (%)	Within Sample (%)
MnO	96.03	3.92	0.05
Fe_2O_3	96.62	3.36	0.01
V	97.85	2.09	0.06
Cr	93.46	5.55	0.99
Co	94.00	5.62	0.38
Ba	97.39	2.56	0.05
Ni	95.96	3.83	0.21
Cu	98.87	1.08	0.06
Zn	92.64	7.34	0.02
Mo	93.59	3.23	3.17
Pb	96.51	3.43	0.06
As	97.87	1.82	0.31
U	76.92	10.99	12.09
Cd	65.44	3.95	30.61
Sn	95.77	2.42	1.81
Sb	87.68	3.05	9.27

Table 6: Percentage of variance in urban sub-surface soil samples attributable to between-site, between-sample and within-sample variance. All data log-transformed with the exception of U and Cd (after Lister, 2002a)

5.3 LEVELLING OF GEOCHEMICAL DATA

In order to compile seamless geochemical images it is necessary to amalgamate discrete data sets produced over a long period of time and determined by different analytical methods. For G-BASE the levelling is a two stage process. Firstly, data from two or three field campaigns has to be levelled to give a discrete data set covering a region such as an atlas sheet area. Secondly, the regional data set (Atlas sheet area) is levelled relative to the national data set held in the BGS corporate database. However, as common controls are used between atlas areas the second phase of levelling is only required following a major change in analytical methodology such as the time when G-BASE changed from DR to XRF analyses. Initially, G-BASE results were levelled relative to the Scottish Borders direct reading (DR) spectrometry calibration. This has subsequently changed and G-BASE data are now levelled relative to the Wales stream sediment data set determined by XRF.

A good discussion of the levelling of geochemical data sets using the mathematical process of normalisation is given in Darnley et al. (1995). This work describes how the term normalisation is used in a mathematical sense, i.e. *"to adjust the representation of a quantity so that this representation lies within a prescribed range (Parker, 1974), or, any process of rescaling a quantity so that a given integral or other functional of the quantity takes on a pre-determined value (Morris, 1991), rather than in the statistical sense, where it connotes a transformation of a data set so that it has a mean of zero and a variance of one"*. Normalisation of the secondary reference material results gives levelling factors that are applied to the data to give, ultimately, a single discrete national G-BASE data set. The levelling factors for the most recent atlas sheet areas are given in Appendix 5. In conjunction with data from the analysis of international reference materials, accepted elemental values for all G-BASE secondary reference materials are determined. It is with respect to these definitive concentrations that normalisation of SRM results are made.

5.3.1 Example of levelling

The procedures carried out during the levelling of G-BASE geochemical data are best demonstrated by way of example. The analytical data reported by the BGS XRF laboratory for K_2O, from the –150 μm fraction of stream sediment samples collected over the East Anglia regional atlas area, provides a typical example.

Analysis of stream sediment samples from East Anglia was carried out between June 2003 and February 2005. In total, 2675 samples, comprising eight discrete batches (including primary and secondary reference materials) were analysed (

Table **7**).

Batch Number	Analysis Start Date	Analysis End Date	Number of Samples
10372	25-Jun-03	02-Jun-04	484
10377	20-Jul-03	27-Oct-03	283
10586	24-Dec-03	17-Feb-04	528
10587	24-Dec-03	31-Dec-03	106
10719	19-May-04	25-Jun-04	493
10721	25-Jun-04	02-Jul-04	116
10993	02-Feb-05	15-Feb-05	157
10994	24-Dec-04	02-Feb-05	508

Table 7: Batches of data reported during analysis of East Anglia stream sediments

Initial inspection of control plots for K_2O showed discontinuity of G-BASE secondary reference materials inserted throughout the duration of analysis. Of particular note were the differences in values reported from batches 10372 and 10377 (Figure 25). The discontinuity is made apparent as, unusually, batch 10377 was analysed during a break in analysis of batch 10372. The difference is most noticeable in values reported for G-BASE secondary reference material S15B.

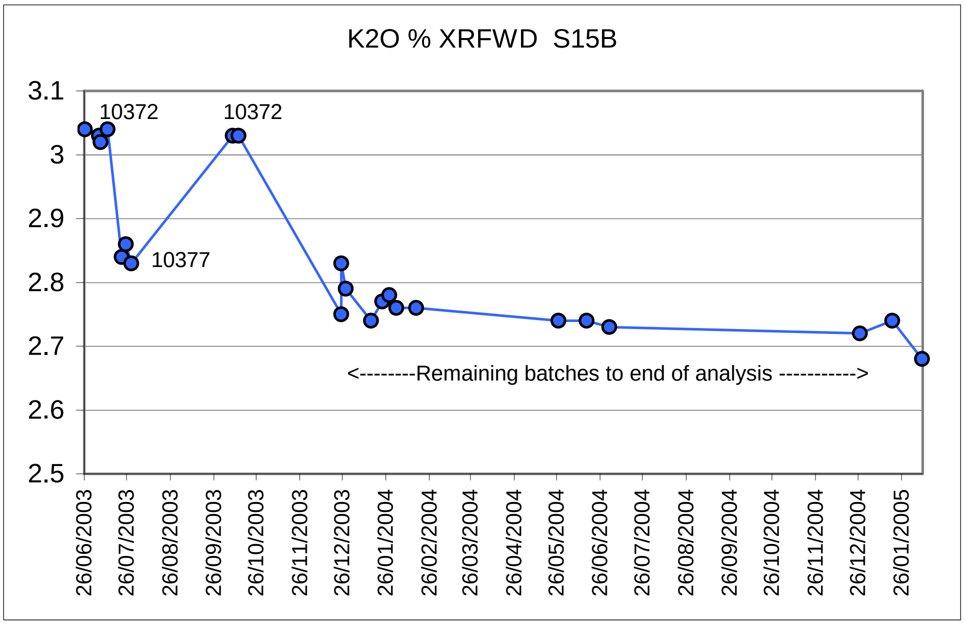


Figure 25: Control plot for G-BASE secondary reference material S15B.

The obvious discontinuities in the K_2O values reported in batches 10372 (red) and 10377 (blue) are highlighted in Table 8. It can also be observed, by reference to the Date field, that analysis of batch 10377 was undertaken during a break in the analysis of batch 10372.

Control plots were also generated for all other G-BASE secondary reference materials included throughout the duration of analysis of the East Anglia stream sediments (Figure 26 - Figure 29).

Mean values for all secondary reference materials analysed within batches 10372, 10377 and the remaining batches (dealt with in this exercise as being continuous) were calculated using Excel. These values were then tabulated along with ‘accepted’ values (see Glossary) for the corresponding secondary reference materials (Table 9 - Table 11).

Sample Number	Standard ID	Batch	Date	K_2O % XRFWD
440022	S15B	10372-0010	26-Jun-03	3.04
440687	S15B	10372-0225	6-Jul-03	3.03
440822	S15B	10372-0260	7-Jul-03	3.02
441087	S15B	10372-0341	12-Jul-03	3.04
441622	S15B	10377-0052	22-Jul-03	2.84
441887	S15B	10377-0109	25-Jul-03	2.86
442128	S15B	10377-0196	29-Jul-03	2.83
440287	S15B	10372-0102	9-Oct-03	3.03
440422	S15B	10372-0150	13-Oct-03	3.03
444022	S15B	10587-0022	25-Dec-03	2.75
442422	S15B	10586-0026	25-Dec-03	2.83
444217	S15B	10587-0064	28-Dec-03	2.79
442928	S15B	10586-0190	15-Jan-04	2.74
443196	S15B	10586-0280	23-Jan-04	2.77
443417	S15B	10586-0346	28-Jan-04	2.78
443668	S15B	10586-0417	2-Feb-04	2.76
443947	S15B	10586-0509	16-Feb-04	2.76
444617	S15B	10719-0109	27-May-04	2.74
445087	S15B	10719-0373	16-Jun-04	2.74
445622	S15B	10721-0097	2-Jul-04	2.73
445828	S15B	10994-0048	27-Dec-04	2.72
446447	S15B	10994-0323	19-Jan-05	2.74
447028	S15B	10993-0074	9-Feb-05	2.68

Table 8: K_2O analytical data for G-BASE secondary reference material S15B

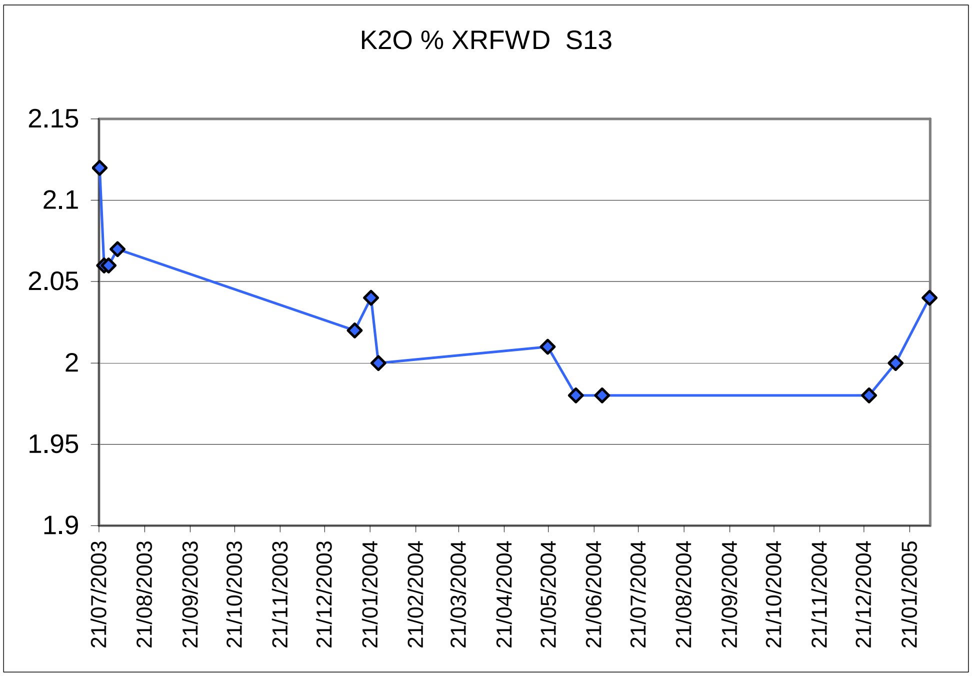


Figure 26: Control plot for G-BASE secondary reference material S13

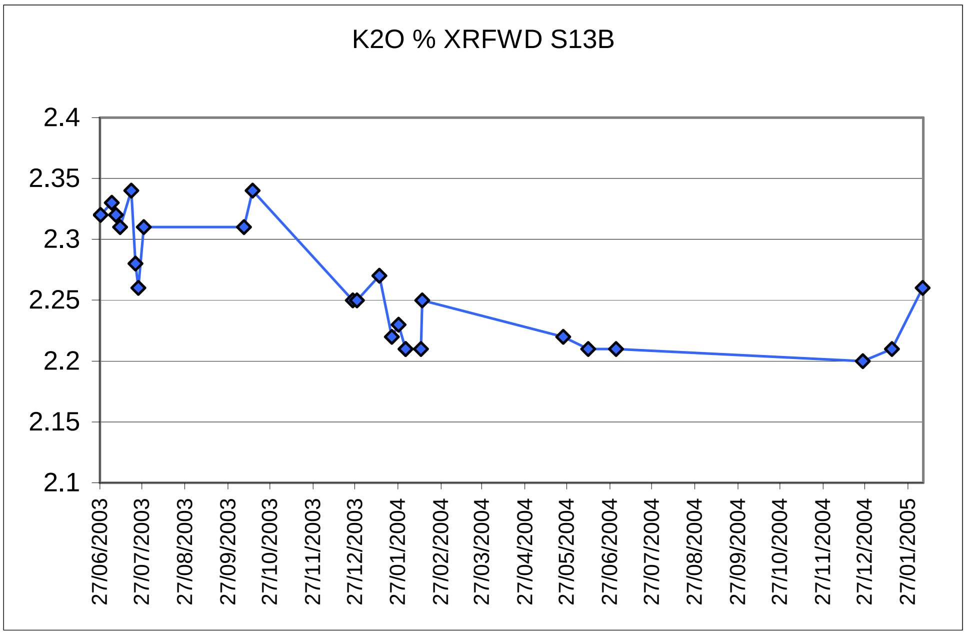


Figure 27: Control plot for G-BASE secondary reference material S13B

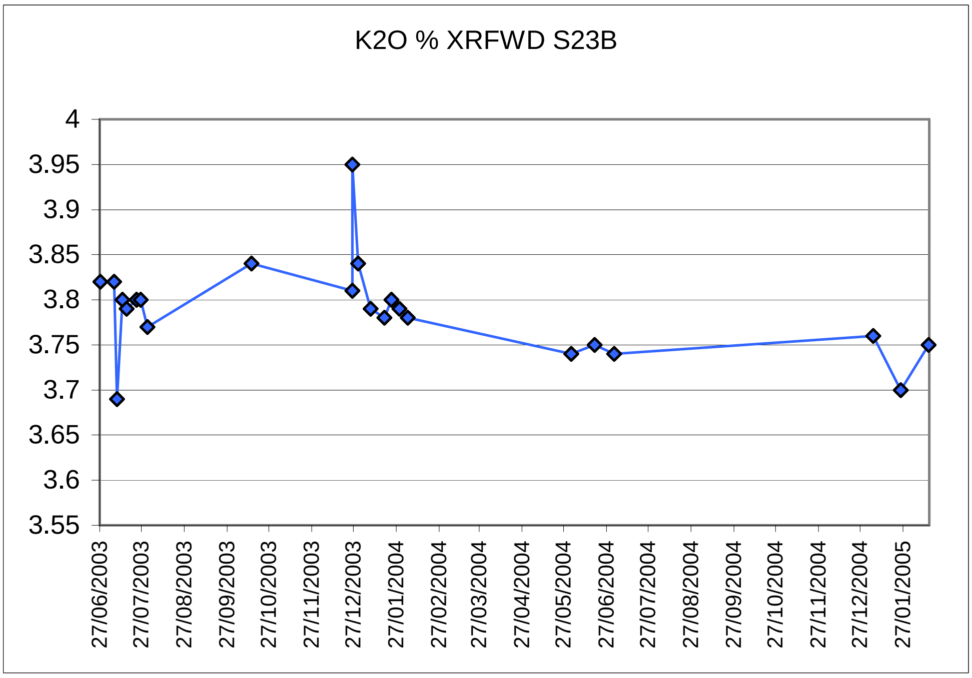


Figure 28: Control plot for G-BASE secondary reference material S23B

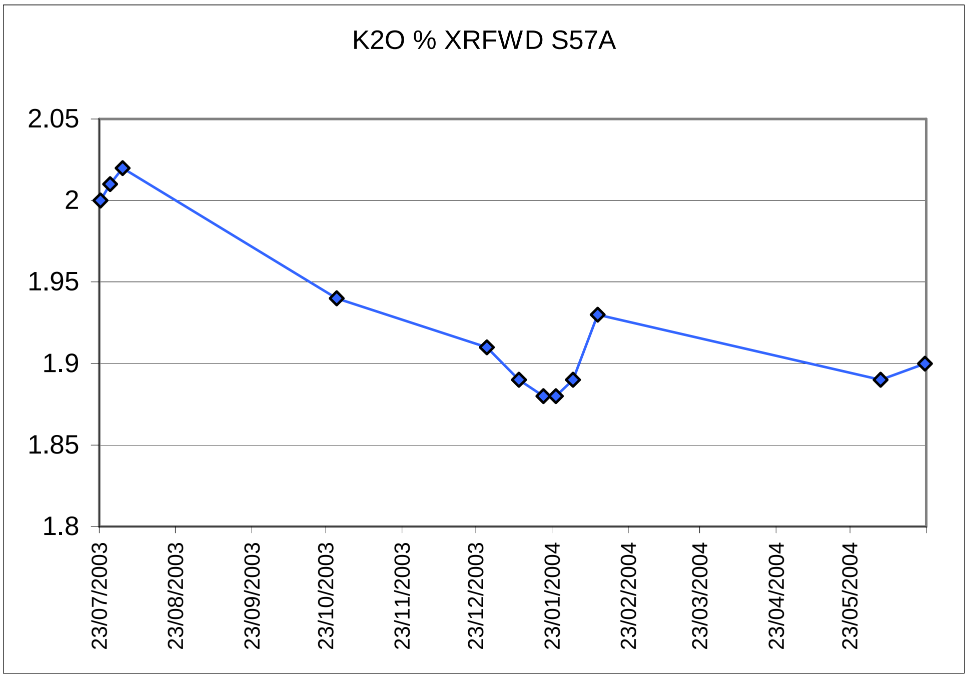


Figure 29: Control plot for G-BASE secondary reference material S57A

Standard ID	Mean Value	K_2O % XRFWD	Accepted K_2O
S13B	Mean 10372	2.32	2.29
S15B	Mean 10372	3.03	2.94
S23B	Mean 10372	3.79	3.86
S58S	Mean 10372	2.14	2.04

Table 9: Mean values for batch 10372

Standard ID	Mean Value	K_2O % XRFWD	Accepted K_2O
S13	Mean 10377	2.08	2.17
S13B	Mean 10377	2.28	2.29
S15B	Mean 10377	2.84	2.94
S23B	Mean 10377	3.79	3.86
S57A	Mean 10377	1.99	1.92

Table 10: Mean values for batch 10377

Standard ID	Mean Value	K_2O % XRFWD	Accepted K_2O
S13	To end of analysis	2.01	2.17
S13B	To end of analysis	2.23	2.29
S15B	To end of analysis	2.75	2.94
S23B	To end of analysis	3.78	3.86
S57A	To end of analysis	1.90	1.92

Table 11: Mean values for all other batches to end of analysis

Linear regression charts were then generated in Excel for batches 10372, 10377 and 'all other' by plotting the reported value against the accepted value (Figure 30 - Figure 32).

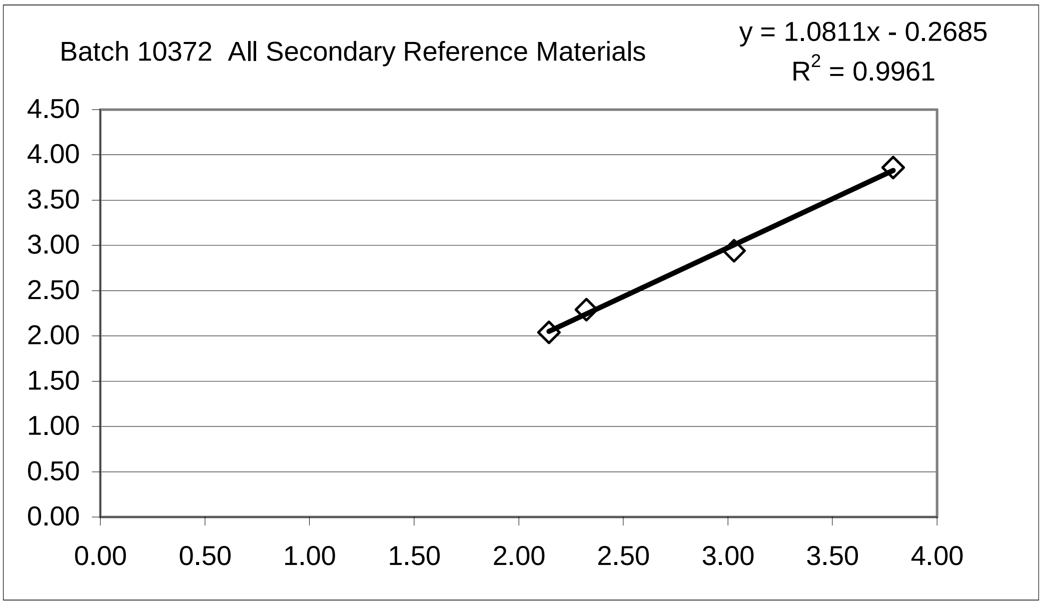


Figure 30: Linear regression plot for batch 10372 (y = accepted value, x = result)

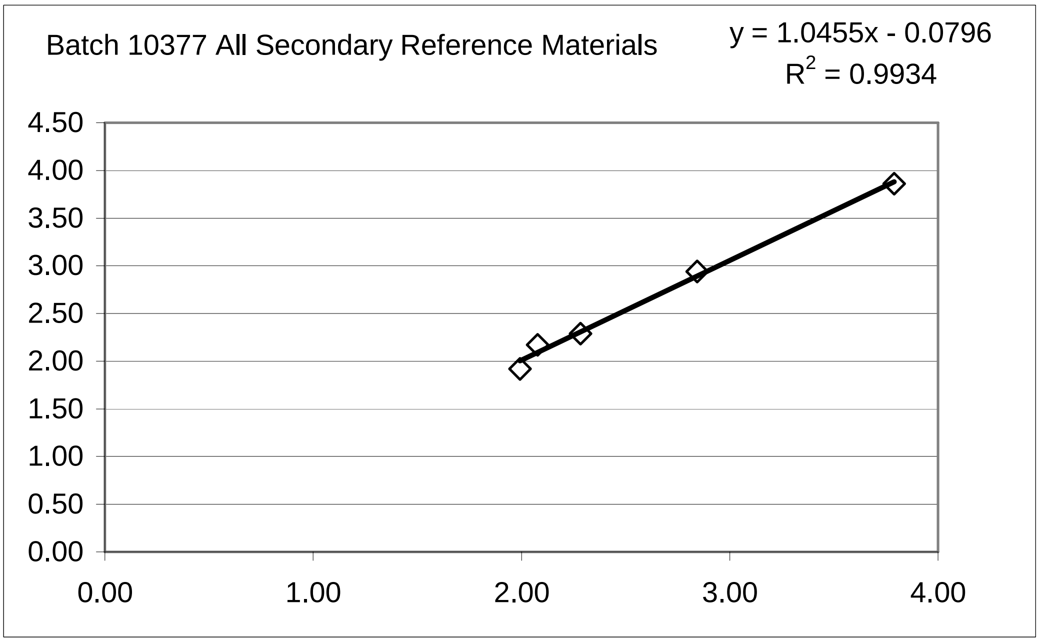


Figure 31: Linear regression plot for batch 10377 (y = accepted value, x = result)

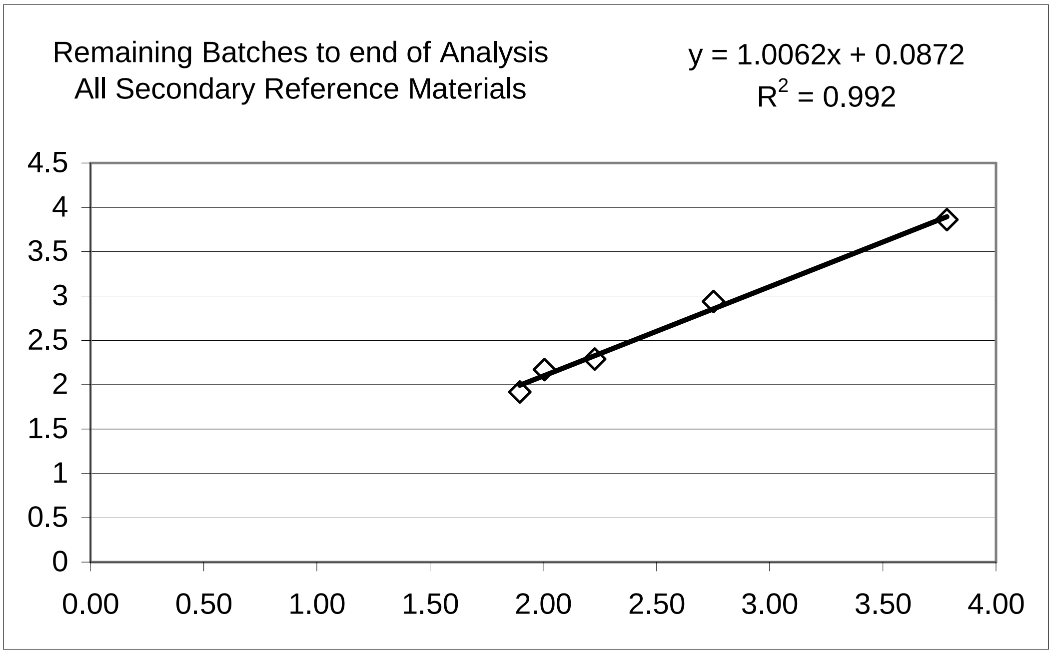


Figure 32: Linear regression plot for all other batches to end of analysis (y = accepted value, x = result)

The 'best fit' linear equations produced in the regression charts were then applied to the corresponding batches of data (Table 12 -

Table **14**). For example, S13B in batch 10372 - applying the regression y = 1.0811x - 0.2685, the result of 2.32 now becomes (1.0811 x 2.32) - 0.2685 = 2.24 (see Table 12).

Standard ID	Mean Value	K2O % XRFWD	Accepted K_2O	Modified K_2O
S13B	Mean 10372	2.32	2.29	2.24
S15B	Mean 10372	3.03	2.94	3.01
S23B	Mean 10372	3.79	3.86	3.83
S58S	Mean 10372	2.14	2.04	2.05

Table 12 : Modified mean values for batch 10372

Standard ID	Mean Value	K2O % XRFWD	Accepted K_2O	Modified K_2O
S13	Mean 10377	2.08	2.17	2.09
S13B	Mean 10377	2.28	2.29	2.31
S15B	Mean 10377	2.84	2.94	2.89
S23B	Mean 10377	3.79	3.86	3.88
S57A	Mean 10377	1.99	1.92	2.00

Table 13: Modified mean values for batch 10377

Standard ID	Mean Value	K2O % XRFWD	Accepted K_2O	Modified K_2O
S13	To end	2.01	2.17	2.11
S13B	To end	2.23	2.29	2.33
S15B	To end	2.75	2.94	2.86
S23B	To end	3.78	3.86	3.89
S57A	To end	1.90	1.92	2.00

Table 14: Modified mean values for all other batches to end of analysis

After applying these factors to the analytical data from all the secondary reference materials within each batch of samples, the control plots were updated. The charts show original ‘raw’ data in blue, and modified data in red (Figure 33 - Figure 37).

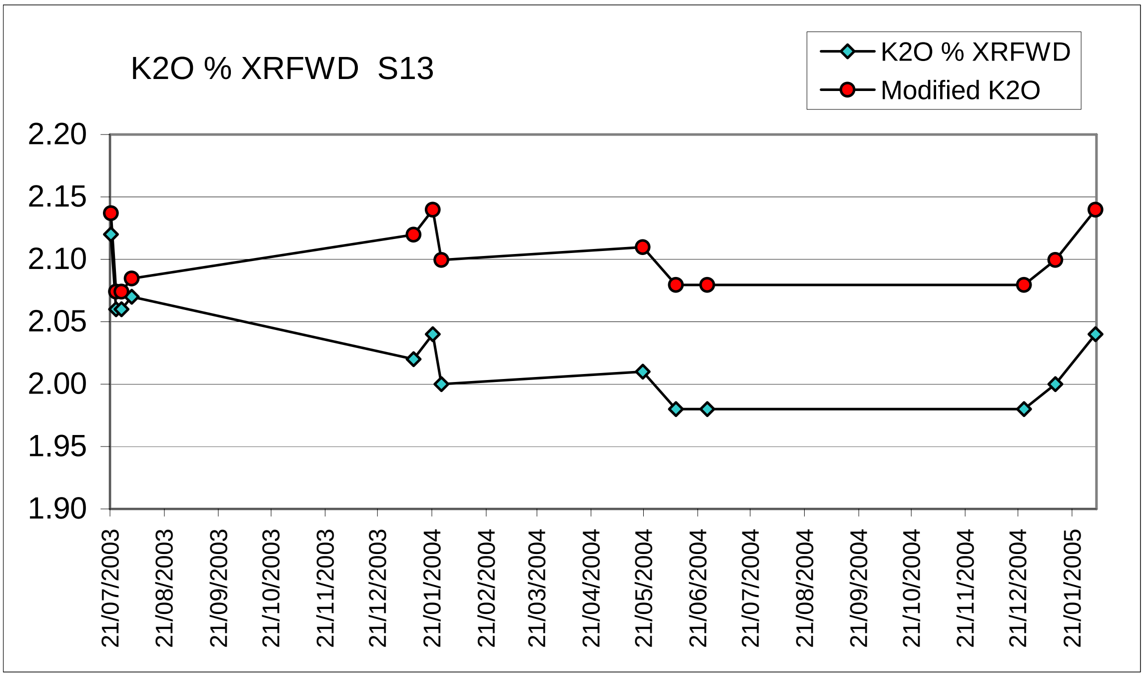


Figure 33: Updated control plot for G-BASE secondary reference material S13

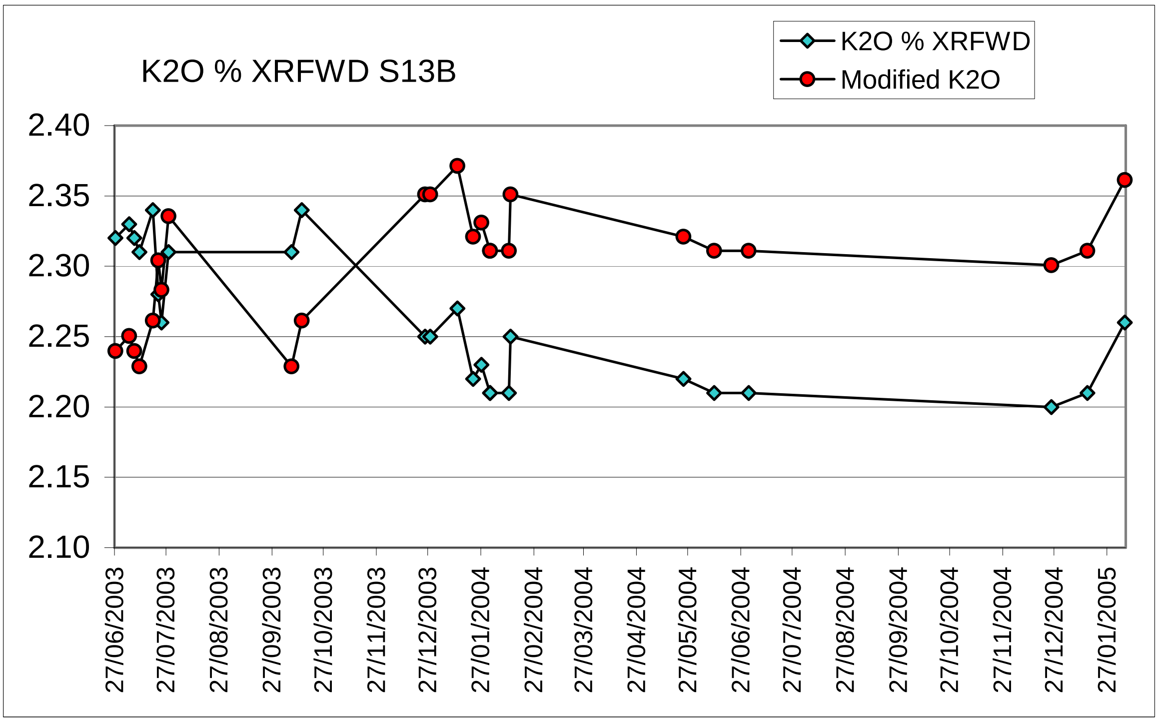


Figure 34: Updated control plot for G-BASE secondary reference material S13B

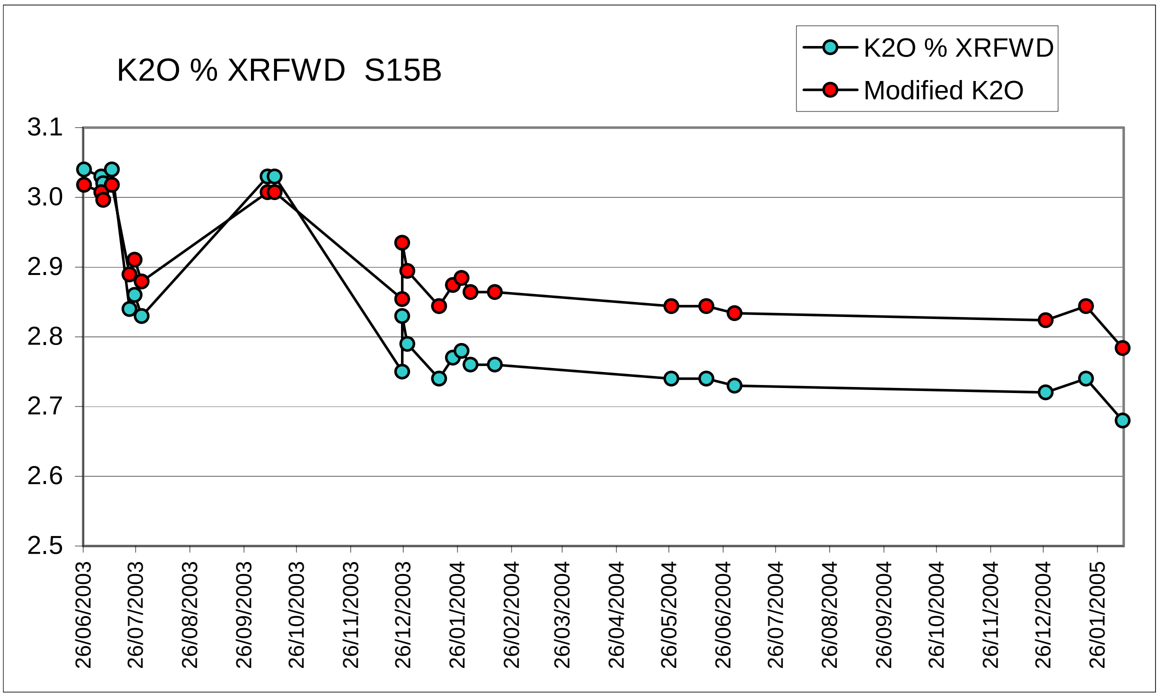


Figure 35: Updated control plot for G-BASE secondary reference material S15B

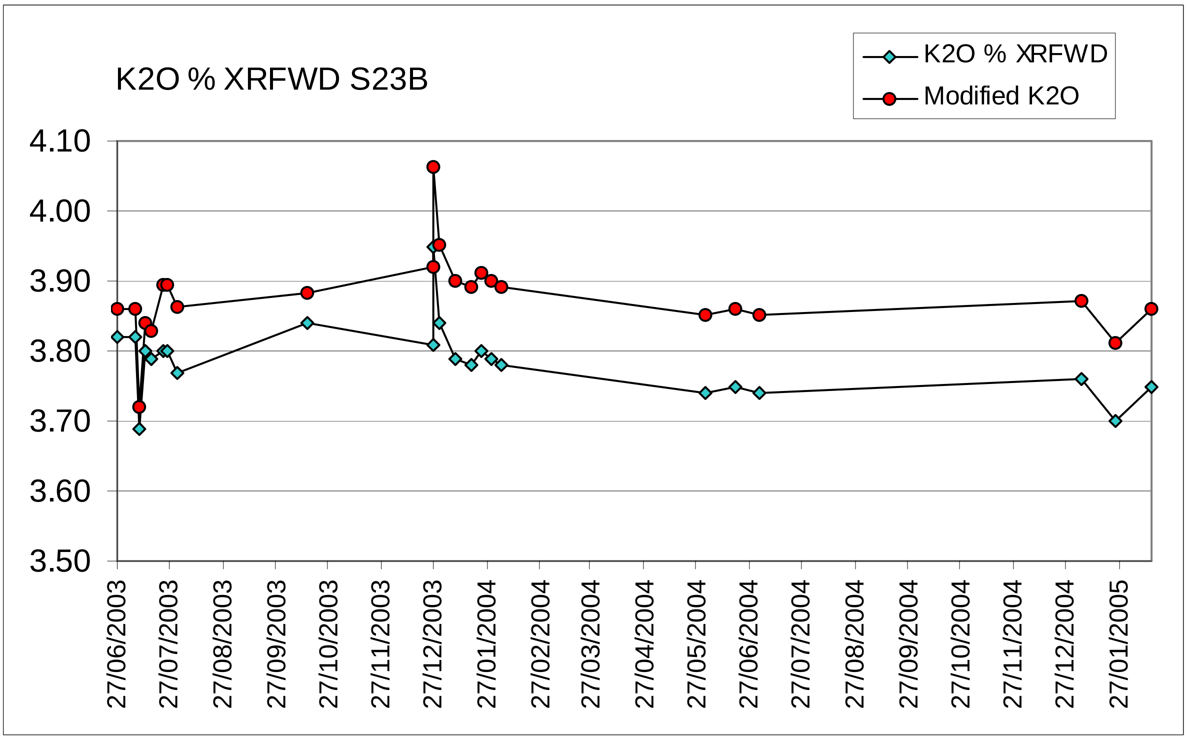


Figure 36: Updated control plot for G-BASE secondary reference material S23B

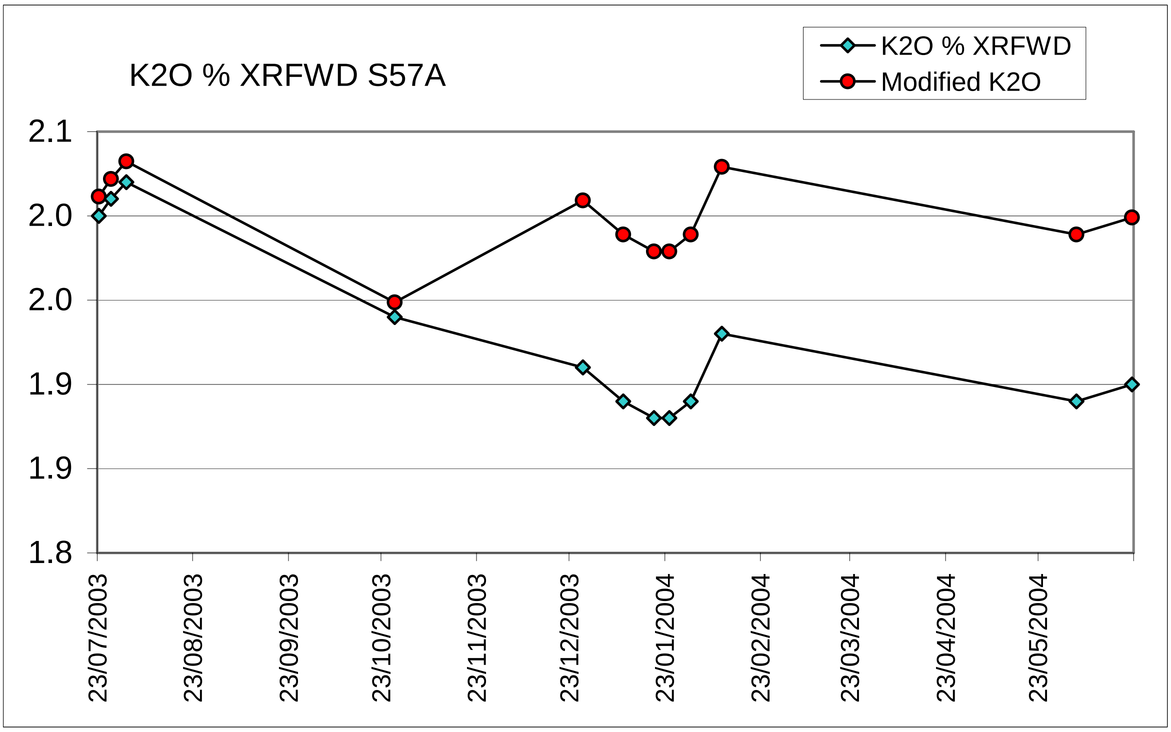


Figure 37: Updated control plot for G-BASE secondary reference material S57A

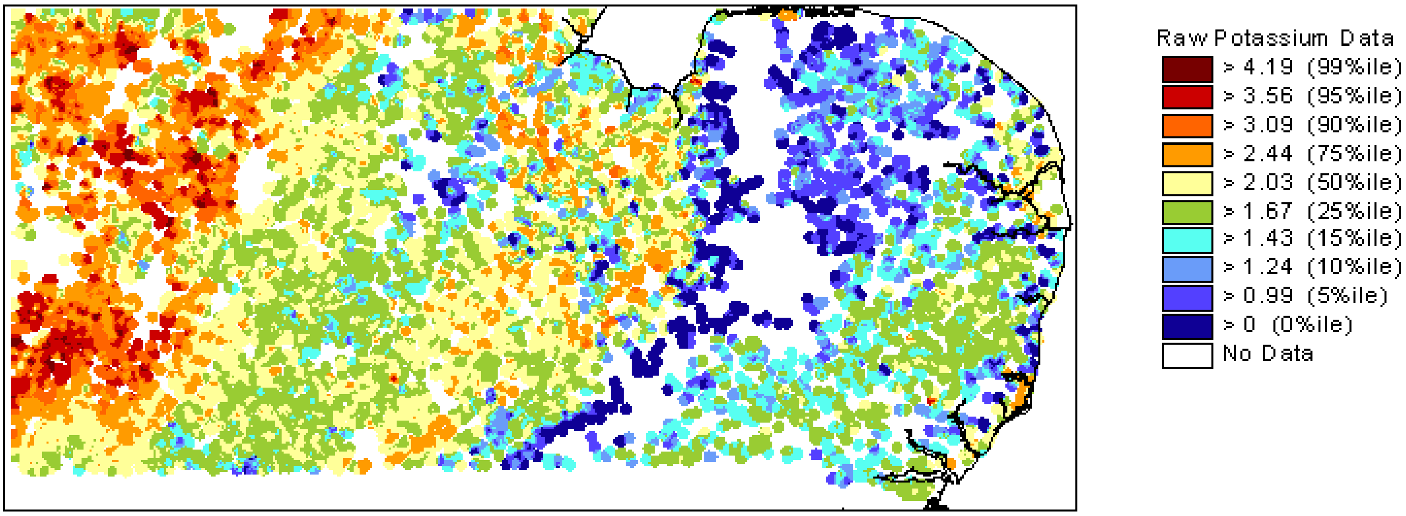
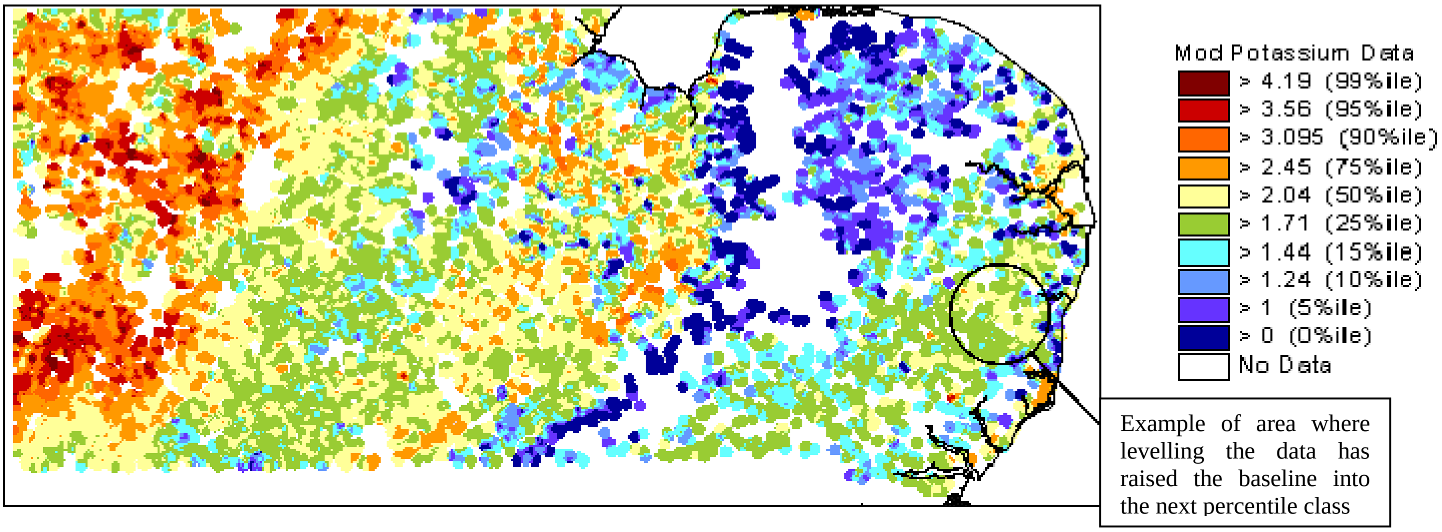


Figure 38: Colour image of K_2O sediment data from Central and Eastern England. Here the raw East Anglia data is plotted along side the conditioned results from the East Midlands.



Example of area where levelling the data has raised the baseline into the next percentile class

Figure 39: Colour image of conditioned K_2O sediment data from Central and Eastern England. Here the conditioned East Anglia data is plotted along side the conditioned data from the East Midlands.

Following re-inspection of the updated control plots, the modified secondary reference material data must be either accepted as being ‘fit for use’ or referred for further conditioning. The effects of applying levelling factors to the results should be checked by examining combined geochemical images. The gridded image for Central and Eastern England is illustrated here before and after levelling factors have been applied to the East Anglia results (Figure 38 and Figure 39). Although the raw data from East Anglia merges with the neighbouring area without any significant "edge" effects (Figure 38), changes in the conditioned data image are very subtle, but can be seen in the differences of percentile classification values, particularly in the 10% - 90% range. The image shown in Figure 39 confirms the levelling factors are suitable and the conditioned data is acceptable.

Plotting the geochemical images is an important part of the iterative process to level the data. It is important to look at the images in the context of other parameters (e.g. sampling campaign boundaries or analytical batches) that may create analytical artefacts. The illustration on the front cover of this report demonstrates how high results of tungsten in stream sediments from the East Midlands atlas area were generated as a consequence of analytical shift between analytical batches. Without comparing the tungsten geochemical images to a plot showing which samples belonged to which analytical batch, an interpretation of the anomalies would most likely have attributed them to geological variations.

Once considered ‘fit for use’ the data levelling factors as applied to the secondary reference material are now those to be finally applied to all the potassium results in the corresponding batches.

In this example:	All batch 10372	(1.0811 x K_2O) – 0.2685
	All batch 10377	(1.0455 x K_2O) – 0.0796
	All other batches	(1.0062 x K_2O) + 0.0872

As shown in the K_2O example for East Anglia, several stages of normalisation may be necessary before data for a particular element satisfy quality control checks. Firstly, shifts in concentration within the period of analysis must be eliminated. This may involve breaking the control plots into a series of segments, within which the concentrations are fairly constant over time. Calculation of the mean concentrations within each segment will provide adequate information to enable ‘within analysis’ correction factors to be applied. An example of this procedure is shown in Appendix 2 of Lister (2002b).

Normalisation of data is undertaken for all elements displaying discontinuous control plots with evident shifts in concentration. Levelling factors for recent a atlas area are given in Appendix 5.

5.4 DATA CONDITIONING REPORT

The concluding part of the data conditioning process is the production of a BGS report in the "Internal Report Series". This report will be a useful reference to the quality control procedures associated with the data set and should be cited whenever the data is presented. The report should document all the results of the work described in Sections 3 to 5 of this report. The data conditioning process will generate hundreds of graphs and data tables and it is not suggested that all the graphs are included in the data conditioning report. Those that show specific issues should be included, others should be archived as digital files and the report should document where such files can be located.

6 Quality Assurance

Quality assurance of G-BASE data is based on reporting and recording of all parts of the data conditioning procedures. This report is an important part of the quality assurance process and by following the procedures documented here the conditioning of G-BASE results can be done in a systematic and consistent manner.

The are four main documents associated with G-BASE quality assurance, namely:

1. the analysts cover note (e.g. Appendix 1)
2. the laboratory batch progress log located in:
   w:\cbh\gbase\ProjectInformation\Laboratories\Sample Analysis\
3. the data conditioning log (Table 15), and
4. the data conditioning report (e.g. Lister, 2002b).

The key stages in quality assurance procedures for G-BASE soil and sediment are as follows (summarised in Table 16):

1. The laboratory manager releases data from the analytical laboratories with a cover note signing that the analyses have been completed satisfactorily. The reported batch of results will have a version number that may be updated should subsequent G-BASE quality control indicate a need for revision. The analytical cover note should include: a description of the sample preparation and analytical methods used; an explanation of upper and lower detection limits; explanation of "null" or not determined entries; and report the analyses of international reference standards as requested.
2. The reporting of results is considered complete once all the required analytical results and batch information have been transferred into the Geochemistry Database via LIMS.
3. The G-BASE data manager or deputy will acknowledge receipt of results by email to the laboratory manager and update the laboratory batch progress log. This log will be monitored monthly by the Project Manager and any issues such as overdue results or payment for results before they are received will be brought to the attention of the laboratory manager. An example of the log is shown in Table 17 regional soils and sediment, stream waters and urban soils being logged on separated worksheets within the EXCEL file.
4. Commencing in 2006 the progress of data conditioning on newly received data will be recorded in a data conditioning log maintained by the data manager or designated deputy. Each data set will have a separate worksheet and an example is shown in Table 15. It is important that any non-conformities or data quality issues are recorded in this. Errors or problems relating to the laboratory analyses should be reported to the laboratory manager immediately. When data error checking and quality control is complete the laboratory will be sent a summary report of the results.
5. The process of data conditioning should be reported in a data conditioning report and any errors or problems recorded. Data is usually conditioned when

an atlas area has been sampled or a group of urban soils has been completed. Commencing financial year 2005-6 it is planned to have an annual turn around of sample collection, analysis and reporting for drainage catchment areas.

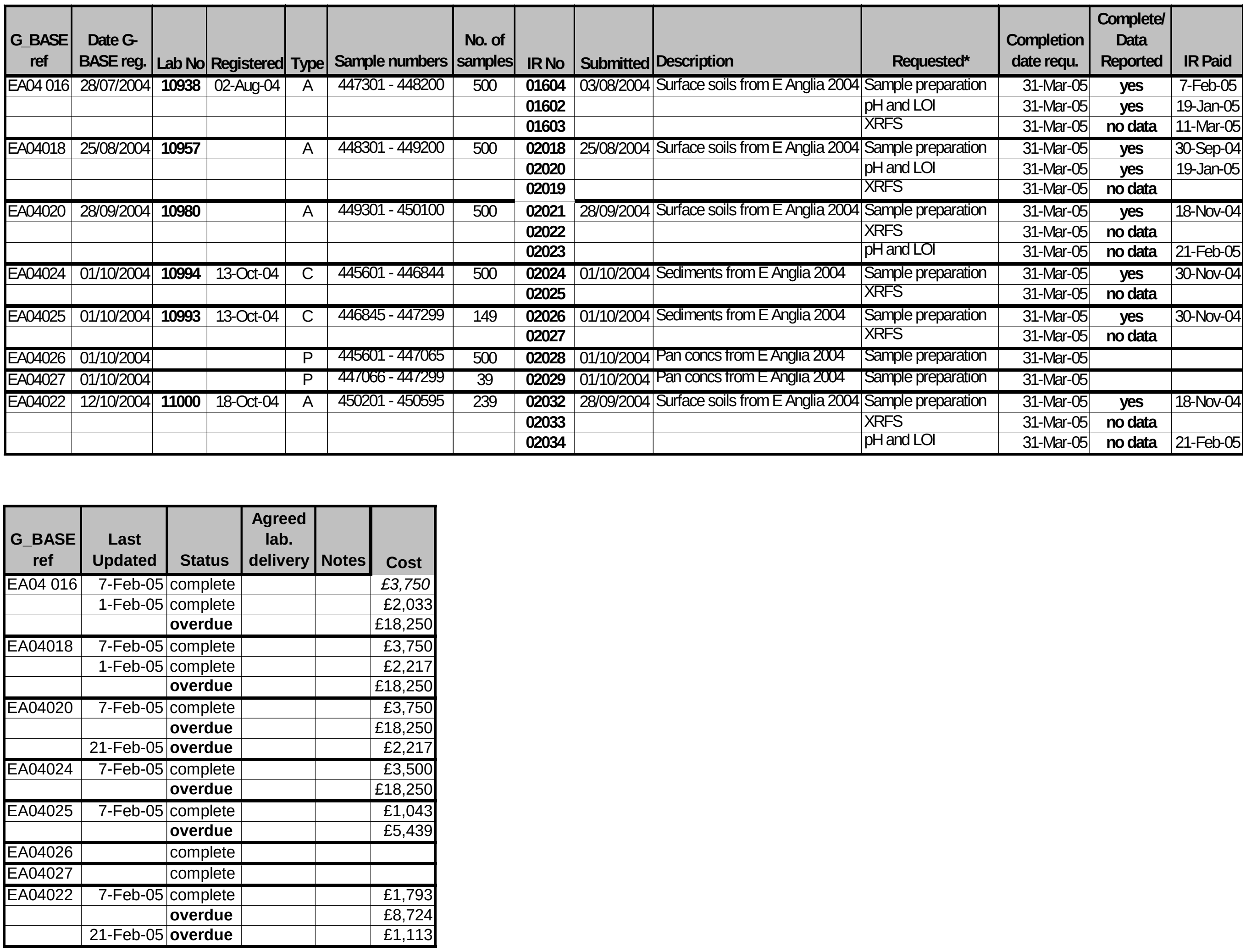
6. On completion of all the control procedures described in this report, data is accepted by the G-BASE Data Manager as being fit for transfer to the BGS corporate Geochemistry Database. The levelled data is passed on to the Geochemistry Data manager for loading to the Oracle data tables. It is important that the analyte qualifier field is populated with the codes shown in Table 5. It is also important that the Geochemistry Database Data Manager receives confirmation from the G-BASE Data Manager that the conditioned data has been loaded to the Geochemistry Database correctly.

Item	Process	Started	Completed	By whom	Notes
1	Raw analytical data loaded to Geochemistry Database	31-Jul-04	31-Mar-05	LIMS	Includes pH and LOI
2	Laboratory batch information loaded to Geochemistry Database	31-Jul-04	31-Mar-05	LIMS	detection limits only
3	Field data loaded to Geochemistry Database	14-Feb-05	25-Mar-05	AS/ACM	using new protocols introduce in 2005
4	Access work database completed	1-Apr-05	1-Apr-05	TRL	EastAngliaSoils.mdb
5	Error checking complete	4-Apr-05	15-Apr-05	SEB	No results for sample C1234, sample lost by labs
6	Control sample tables created	15-Apr-05		TRL	
7	Shewhart plots				
8	Duplicate/replicate x-y plots				
9	ANOVA				
10	International Reference Material comparisons				
11	Summary QC report to labs				
12	Data levelling				
13	Conditioned data loaded to Geochemistry Database				
14	Data conditioning report				

Table 15: Example of data conditioning log file sheet

Item	**Format**	**Created by:**	**Sent to:**
Laboratory Cover Note	MS Word	Laboratory Manager	G-BASE Data Manager and copied to G-BASE Administrator
Laboratory batch progress log	MS EXCEL	G-BASE Administrator	For monthly inspection by G-BASE Project Manager
Data conditioning progress log	MS EXCEL	G-BASE Data Manager	For monthly inspection by G-BASE Project Manager
Summary report of error checking and QC	MS Word	G-BASE Data Manager	G-BASE Project Manager copied to Laboratory Manager
Conditioned analytical results ready for loading to the Geochemistry Database	MS EXCEL	G-BASE Data Manager	Geochemistry Database Manager
Data conditioning report	MS Word (Internal Report Series)	G-BASE Data Manager	BGS Library

Table 16: Summary of Quality Assurance documents produced



G_BASE ref	Date G-BASE reg.	Lab No	Registered	Type	Sample numbers	No. of samples	IR No	Submitted	Description	Requested*	Completion date requ.	Complete/ Data Reported	IR Paid
EA04 016	28/07/2004	**10938**	02-Aug-04	A	447301 - 448200	500	**01604**	03/08/2004	Surface soils from E Anglia 2004	Sample preparation	31-Mar-05	**yes**	7-Feb-05
							01602			pH and LOI	31-Mar-05	**yes**	19-Jan-05
							01603			XRFS	31-Mar-05	**no data**	11-Mar-05
EA04018	25/08/2004	**10957**		A	448301 - 449200	500	**02018**	25/08/2004	Surface soils from E Anglia 2004	Sample preparation	31-Mar-05	**yes**	30-Sep-04
							02020			pH and LOI	31-Mar-05	**yes**	19-Jan-05
							02019			XRFS	31-Mar-05	**no data**	
EA04020	28/09/2004	**10980**		A	449301 - 450100	500	**02021**	28/09/2004	Surface soils from E Anglia 2004	Sample preparation	31-Mar-05	**yes**	18-Nov-04
							02022			XRFS	31-Mar-05	**no data**	
							02023			pH and LOI	31-Mar-05	**no data**	21-Feb-05
EA04024	01/10/2004	**10994**	13-Oct-04	C	445601 - 446844	500	**02024**	01/10/2004	Sediments from E Anglia 2004	Sample preparation	31-Mar-05	**yes**	30-Nov-04
							02025			XRFS	31-Mar-05	**no data**	
EA04025	01/10/2004	**10993**	13-Oct-04	C	446845 - 447299	149	**02026**	01/10/2004	Sediments from E Anglia 2004	Sample preparation	31-Mar-05	**yes**	30-Nov-04
							02027			XRFS	31-Mar-05	**no data**	
EA04026	01/10/2004			P	445601 - 447065	500	**02028**	01/10/2004	Pan concs from E Anglia 2004	Sample preparation	31-Mar-05		
EA04027	01/10/2004			P	447066 - 447299	39	**02029**	01/10/2004	Pan concs from E Anglia 2004	Sample preparation	31-Mar-05		
EA04022	12/10/2004	**11000**	18-Oct-04	A	450201 - 450595	239	**02032**	28/09/2004	Surface soils from E Anglia 2004	Sample preparation	31-Mar-05	**yes**	18-Nov-04
							02033			XRFS	31-Mar-05	**no data**	
							02034			pH and LOI	31-Mar-05	**no data**	21-Feb-05

G_BASE ref	Last Updated	Status	Agreed lab. delivery	Notes	Cost
EA04 016	7-Feb-05	complete			*£3,750*
	1-Feb-05	complete			£2,033
		overdue			£18,250
EA04018	7-Feb-05	complete			£3,750
	1-Feb-05	complete			£2,217
		overdue			£18,250
EA04020	7-Feb-05	complete			£3,750
		overdue			£18,250
	21-Feb-05	**overdue**			£2,217
EA04024	7-Feb-05	complete			£3,500
		overdue			£18,250
EA04025	7-Feb-05	complete			£1,043
		overdue			£5,439
EA04026		complete			
EA04027		complete			
EA04022	7-Feb-05	complete			£1,793
		overdue			£8,724
	21-Feb-05	**overdue**			£1,113

Table 17: Example of sheet from the laboratory batch progress log

7 Concluding remarks

1. Data conditioning is a time consuming but very necessary task. The G-BASE project now analyses three different media types (sediments, soils and waters), and for each media in excess of forty inorganic elements are determined resulting in hundreds of plots and statistical analyses. At the rate of processing two or three elements a day, a rough estimate of twenty-five staff days would be required to process a G-BASE atlas area. If any major data quality issues are identified then further time is also required to address them.
2. Many of the quality issues relating to G-BASE data in the BGS corporate Geochemistry Database arise not from the sampling and analyses of the G-BASE samples but from the process of databasing the data. Results that were passed on from the G-BASE project for loading into the Geochemistry Database have in many instances been degraded because of inappropriate procedures in loading the data such as loading null values as zeros, truncating decimal places and not populating the analyte qualifier field correctly. These issues of quality control and assurance are not covered by this report.
3. Whilst the data conditioning process can be prescribed as a number of routine statistical and plotting procedures, the interpretative skills of a geochemist are still required to make decisions on the significance of any quality issues identified. An inexperienced interpreter will inevitably waste time on issues that are insignificant in terms of what the data is to be used for.
4. It is very important that the data conditioning carried out on each data set is well documented preferably in the form of a BGS Internal report. In such reports it is not necessary to publish every plot and statistical analysis carried out. Reports should concentrate on any quality issues found with the data. Related digital maps and files used in the data conditioning process must be archived in the appropriate manner. It is intended that from 2006 onwards the data conditioning process can be completed and reported in the same financial year that samples are collected.
5. The task of data conditioning is made much more difficult by delays in the laboratory analyses of sample batches which, in turn, may lead to different batches from the same field campaign being analysed over one or more laboratory instrument or calibrations changes. The laboratories should appreciate how this impacts on work and every effort should be made to analyse samples from the same field campaign rapidly and without interruption.
6. The results of the G-BASE data conditioning process should be made more accessible to the BGS laboratories that in the past would have only known about failings in quality control rather than the more positive successes. Staff from the laboratories should be encouraged to be involved in the sampling process and must be given more feedback from our quality control procedures, that by necessity, initially remain "blind" to the analysts. This will be achieved by more systematic reporting of the QC/QA procedures and more regular

meetings between the analysts and geochemists at which data conditioning work can be presented and discussed.

7. The creation and maintenance of secondary reference materials is a vital part of the control process. It is estimated that these materials cost G-BASE between £300 - £400 per kilogram to collect and prepare. Budget must be put aside in the G-BASE finances to maintain a stock of adequate secondary reference materials. G-BASE and the Sample Preparation Section should also consider preparing stocks of such material for sale and use by commercial projects that have in the past relied on G-BASE to provide them (usually free-of-charge).

Appendix 1 : Example of Analysis Report Cover Note

ANALYTICAL GEOCHEMISTRY LABORATORIES

ANALYSIS REPORT COVER NOTE

This report consists of a 85 page Analysis Report Cover Note and 58 pages of test data

Report Number:	06206/2	Customer Ref/Order No:	IR18718
Report Date:	4 July 2003	Sample(s) received on:	15 July 1999
Issue Status:	Complete	Analysis commenced on:	13 August 1999

Sample Details

All samples were received in good condition. Samples 420263 and 420426 were not delivered for analysis.

Unless previously agreed otherwise in writing, samples will be retained for three months from the date of issue of this report prior to disposal. Please contact the Laboratory if you wish to make alternative arrangements.

Analysis Details

Determinands	Test Method	Notes
Major and trace elements	XRFS pressed powder pellets	

Because of limitations with the current software used for reporting data, the number of significant figures quoted in the attached table may not be representative of the actual uncertainty. Data should be considered accurate to no more than three significant figures.

Data for some elements are released beneath the usual reporting limits and to additional decimal places at the Client’s request.
Samples containing Cs, Ba, La or Ce >1000 ppm will possibly effect the data in this report. Iodine is not present in our QC control standards and therefore its QC status is inferred from tellurium, a spectrographically adjacent element.

N.D. represents not determined due to unspecified interferences that have not been corrected for.
Samples with Zn >1000 ppm will effect Sc, V, Cr, Co, Ni, Cu, Zn, Ga, Ba, Hf, Ta, W data and possibly other elements, as the Zn concentrations are above the trace element program calibration limit. The specified trace element data have not been corrected for this effect.

Samples with Sr >1000 ppm will effect Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Y, Zr, Hf, Ta, W, Tl, Pb, Bi, Th, U data and possibly other elements, as the Sr concentrations are above the trace element program calibration limit. The specified trace element data have not been corrected for this effect.

Samples with Zr >1000 ppm will effect Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Y, Zr, Nb, Mo, Hf, Ta, W, Tl, Pb, Bi, Th, U data and possibly other elements, as the Zr concentrations are above the trace element program calibration limit. The specified trace element data have not been corrected for this effect.

Samples with Ba >1000 ppm will effect Sc, V, Cr, Ba data and possibly other elements, as the Ba concentrations are above the trace element program calibration limit. The specified trace element data have not been corrected for this effect.

Samples with Pb >1000 ppm will effect Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Y, Zr, Hf, Ta, W, Tl, Pb, Bi, Th, U data and possibly other elements, as the Pb concentrations are above the trace element program calibration limit. The specified trace element data have not been corrected for this effect.

The client should note that trace elements, whose characteristic x-ray lines lie on the long wavelength side of the iron absorption edge, i.e. Sc, V, Cr, Co, Cs, Ba, La, Ce, Nd and Sm, are effected by absorption from major elements which are not corrected for by this calibration method. Therefore, these trace elements are not as accurate as others determined by X-ray Fluorescence Spectrometry (XRFS); measurement by other techniques, e.g. ICP-MS are likely to be more reliable. The XRFS calibration lines were established using numerous reference materials (RMs) and by placing the slope to give best fit through the average of the predominantly ‘silicate’ RM matrices. Thus, if the sample matrix differs widely from this average it may produce erroneous results.

The client should also note that the data for major elements in pressed powder pellets will not be as accurate as those produced by XRFS on fused glass beads because they too are not corrected for matrix effects. However, over 200 RMs were measured for the calibration, followed by a limited validation exercise using stream and lake sediment RMs and International Soil Exchange (ISE) proficiency testing samples.

This report is issued under complete status. All analyses requested have been completed and results are issued with full compliance of data verification subject to the statements above. The report supersedes report 06206/1 previously issued under intermediate status on the 10 July 2002. Ag, Cd, Sn, Sb, Te, I, Cs, La and Ce data have been included since the last issue. No other data have changed since the intermediate issue.

Report authorised by: ………………………………………. Date: ……………………….

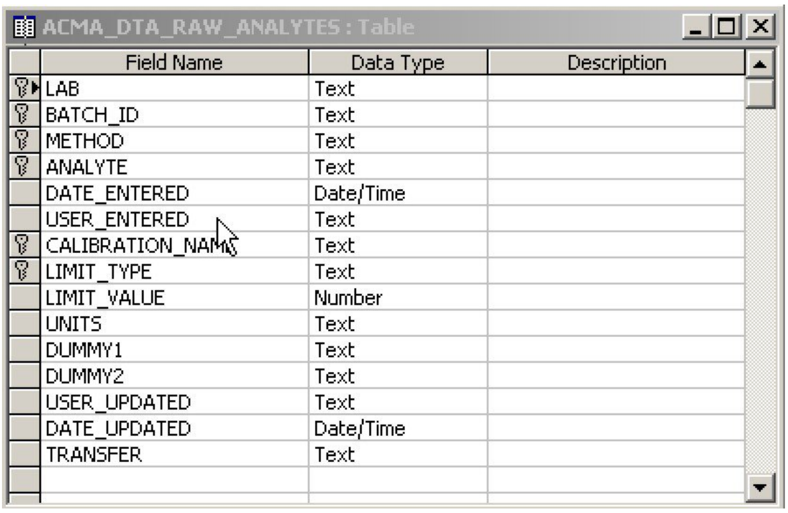
Dr Charles J B Gowing
Deputy X-Ray Fluorescence Spectrometry Section Manager

on behalf of Mark N Ingham, X-Ray Fluorescence Spectrometry Section Manager

Appendix 2 :Oracle data tables populated during the transfer of raw analytical data from the BGS LIMS

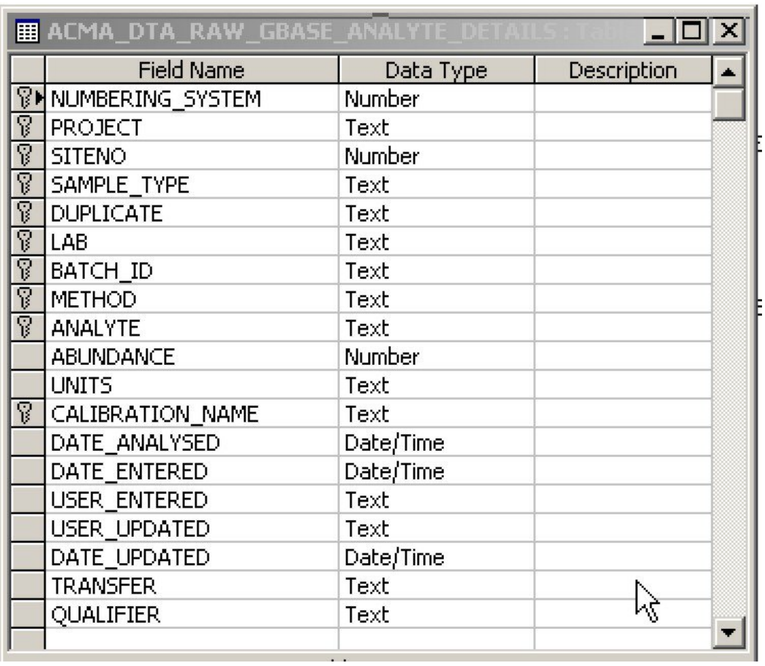
The following Oracle data tables are populated during the transfer of raw analytical data via the BGS LIMS. Project Batch, Batch and Analysis are described in Harris and Coats, 1992. The analyte and analyte determinations tables are specifically for G-BASE raw (unconditioned) results and were tables created by Alan Mackenzie in 2004.

BGS_DTA_RAW_ANALYTES



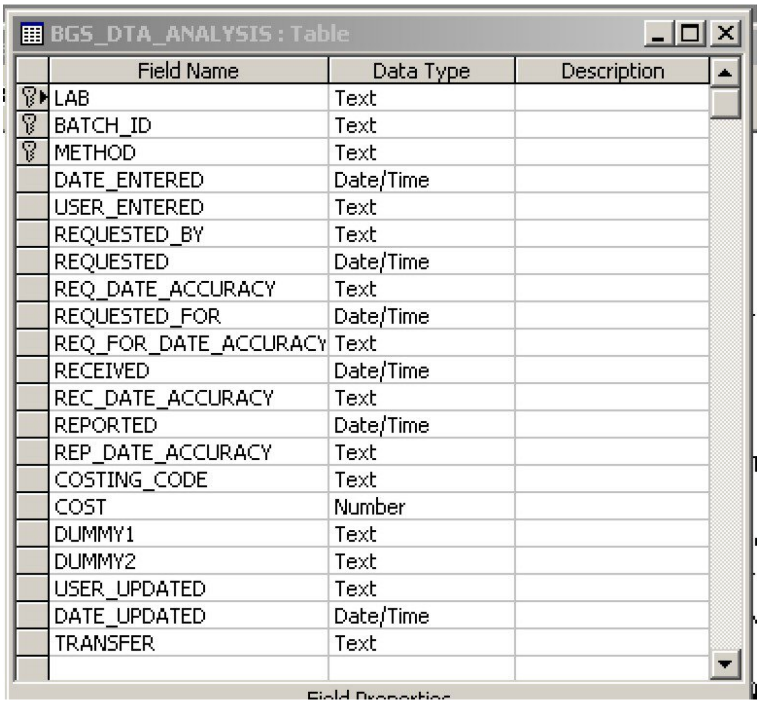
	Field Name	Data Type	Description
🔑	LAB	Text	
🔑	BATCH_ID	Text	
🔑	METHOD	Text	
🔑	ANALYTE	Text	
	DATE_ENTERED	Date/Time	
	USER_ENTERED	Text	
🔑	CALIBRATION_NAME	Text	
🔑	LIMIT_TYPE	Text	
	LIMIT_VALUE	Number	
	UNITS	Text	
	DUMMY1	Text	
	DUMMY2	Text	
	USER_UPDATED	Text	
	DATE_UPDATED	Date/Time	
	TRANSFER	Text	

BGS_DTA_RAW_GBASE_ANALYTE_DETAILS



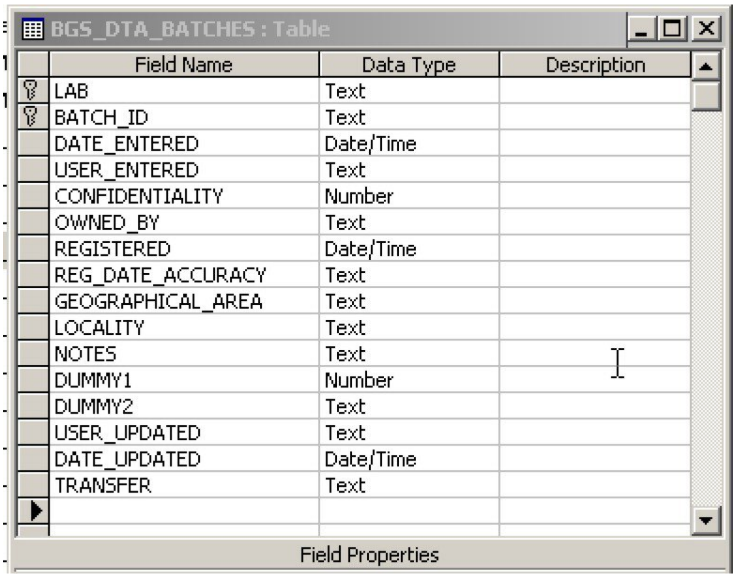
	Field Name	Data Type	Description
🔑	NUMBERING_SYSTEM	Number	
🔑	PROJECT	Text	
🔑	SITENO	Number	
🔑	SAMPLE_TYPE	Text	
🔑	DUPLICATE	Text	
🔑	LAB	Text	
🔑	BATCH_ID	Text	
🔑	METHOD	Text	
🔑	ANALYTE	Text	
	ABUNDANCE	Number	
	UNITS	Text	
🔑	CALIBRATION_NAME	Text	
	DATE_ANALYSED	Date/Time	
	DATE_ENTERED	Date/Time	
	USER_ENTERED	Text	
	USER_UPDATED	Text	
	DATE_UPDATED	Date/Time	
	TRANSFER	Text	
	QUALIFIER	Text	

BGS_DTA_ANALYSIS



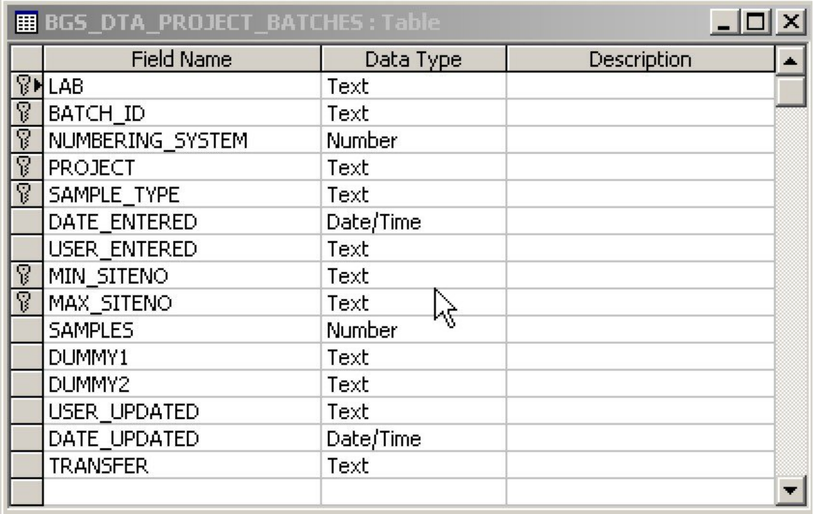
Field Name	Data Type	Description
LAB	Text	
BATCH_ID	Text	
METHOD	Text	
DATE_ENTERED	Date/Time	
USER_ENTERED	Text	
REQUESTED_BY	Text	
REQUESTED	Date/Time	
REQ_DATE_ACCURACY	Text	
REQUESTED_FOR	Date/Time	
REQ_FOR_DATE_ACCURACY	Text	
RECEIVED	Date/Time	
REC_DATE_ACCURACY	Text	
REPORTED	Date/Time	
REP_DATE_ACCURACY	Text	
COSTING_CODE	Text	
COST	Number	
DUMMY1	Text	
DUMMY2	Text	
USER_UPDATED	Text	
DATE_UPDATED	Date/Time	
TRANSFER	Text	

BGS_DTA_BATCHES



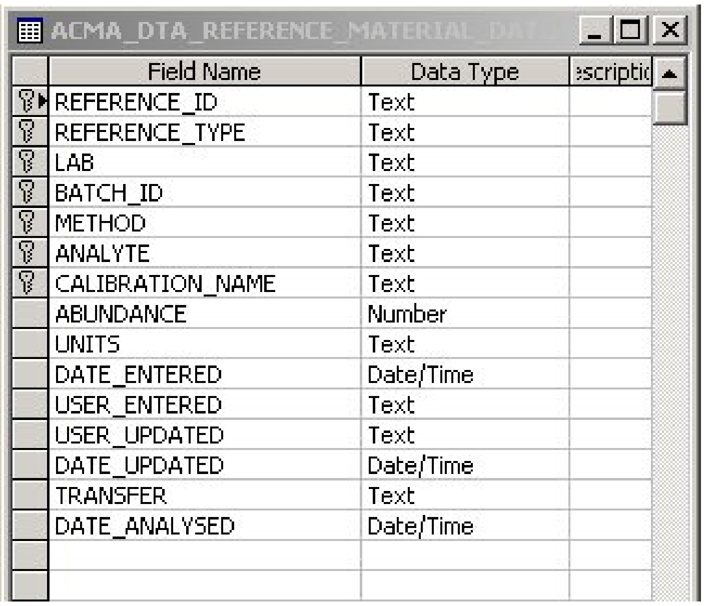
Field Name	Data Type	Description
LAB	Text	
BATCH_ID	Text	
DATE_ENTERED	Date/Time	
USER_ENTERED	Text	
CONFIDENTIALITY	Number	
OWNED_BY	Text	
REGISTERED	Date/Time	
REG_DATE_ACCURACY	Text	
GEOGRAPHICAL_AREA	Text	
LOCALITY	Text	
NOTES	Text	
DUMMY1	Number	
DUMMY2	Text	
USER_UPDATED	Text	
DATE_UPDATED	Date/Time	
TRANSFER	Text	

BGS_DTA_PROJECT_BATCHES



	Field Name	Data Type	Description
🔑	LAB	Text	
🔑	BATCH_ID	Text	
🔑	NUMBERING_SYSTEM	Number	
🔑	PROJECT	Text	
🔑	SAMPLE_TYPE	Text	
	DATE_ENTERED	Date/Time	
	USER_ENTERED	Text	
🔑	MIN_SITENO	Text	
🔑	MAX_SITENO	Text	
	SAMPLES	Number	
	DUMMY1	Text	
	DUMMY2	Text	
	USER_UPDATED	Text	
	DATE_UPDATED	Date/Time	
	TRANSFER	Text	

BGS_DTA_REFERENCE_MATERIAL_DATA



	Field Name	Data Type	escriptic
🔑	REFERENCE_ID	Text	
🔑	REFERENCE_TYPE	Text	
🔑	LAB	Text	
🔑	BATCH_ID	Text	
🔑	METHOD	Text	
🔑	ANALYTE	Text	
🔑	CALIBRATION_NAME	Text	
	ABUNDANCE	Number	
	UNITS	Text	
	DATE_ENTERED	Date/Time	
	USER_ENTERED	Text	
	USER_UPDATED	Text	
	DATE_UPDATED	Date/Time	
	TRANSFER	Text	
	DATE_ANALYSED	Date/Time	

Appendix 3 : Excel macros used in the QC process

There are several MS Excel macros that were written for the Morocco Geochemical Mapping Project that can be used in the G-BASE data conditioning process. The advantage of using these macros is MS Excel is widely used all over the world so the data conditioning process does not have a dependence on specialist statistical software.

The macros will process duplicate-replicate data formatted in a specified manner to give duplicate-replicate x-y plots for every element field. The same data sheet with minor editing can then be used to do nested-ANOVA analysis. The ANOVA analysis by an Excel macro (available in nested_Anova_blank.xls) has already been documented in Johnson (2002) and will only be briefly described here. The x-y plotting macro is available in a blank Excel workbook called duprepplot_blank.xls.

REQUIREMENTS TO RUN MACROS

- a PC running MS Excel 2000 or later
- a simple knowledge of Excel macros
- a set of duplicate-replicate results reported in the format described below (a minimum of six duplicate-replicate pairs are recommended)

DUPLICATE-REPLICATE PLOTS (duprepplot_blank.xls)

Summary - The user creates a worksheet containing the duplicate-replicate results. This is copied into the duprepplot_blank.xls workbook and the block of data to be plotted is selected. On execution the macro checks that the reported results contain a valid number of duplicate/replicate pairs and then reformats the data into columns, element by element. The reformatted data is inserted after the last row of results of the selected data. Finally x-y plots are generated for each element and inserted as a new chart (named with a label taken from the header row of the original data sheet). Three plots are superimposed on each chart, namely, DUPB v SSB, DUPA v SSA, DUPB v DUPA.

SAMPLE_NO	Type	Sampnumb	Labnumb	Na2O	MgO	Al2O3	SiO2	
420076	DUPA	76	1234	0.3	4.3	12.2	52.4	……
420086	SSA	86	1234	0.3	4.3	12.2	54	……
420081	DUPB	81	1234	0.3	4.7	13.6	54	……
420078	SSB	78	1234	0.3	4.8	13.8	55.3	……
420131	DUPA	131	1234	0.6	5.2	14.7	53.9	……
420177	SSA	177	1234	0.6	4.9	14.2	54.9	……
420137	DUPB	137	1234	0.6	5	14.1	56.1	……
420158	SSB	158	1234	0.6	5	14.2	57	……
420476	DUPA	476	1234	0.4	1.4	18.5	58	……
420486	SSA	486	1234	0.4	1.4	18.4	58.5	……
420481	DUPB	481	1234	0.5	1.3	18.3	60.3	……
420478	SSB	478	1234	0.5	1.3	18.6	60.6	……
……	……	……	……	……	……	……	……	……

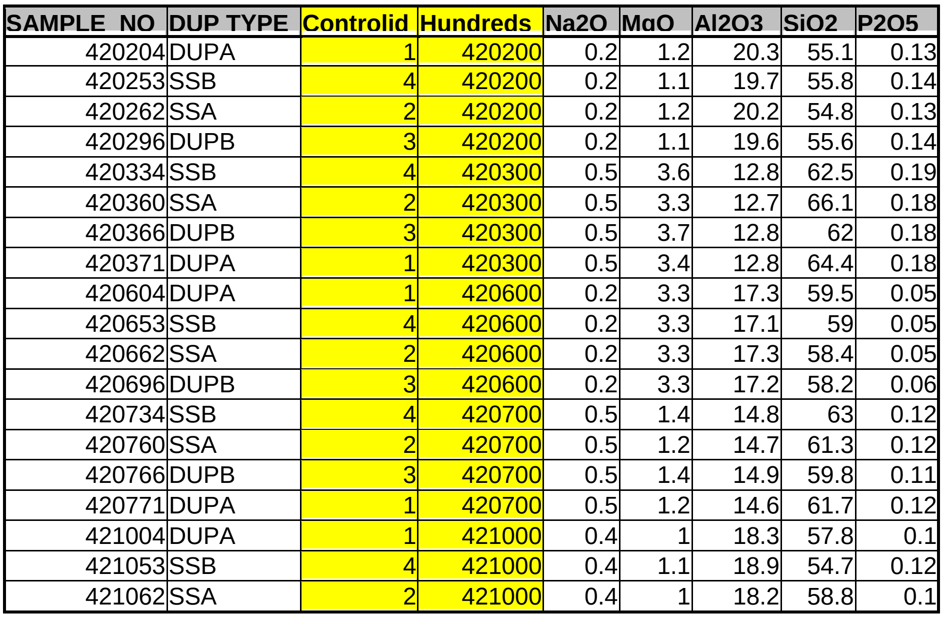
Figure 40: Figure illustrating the format of the worksheet required for the x-y plotting macro. (All element columns and data rows not illustrated)

Procedure:

1. Prepare a worksheet containing data in the exact format as that shown in Figure 40. The initial data is selected as described earlier in this report (see Figure 21 and Figure 22).
    - The first row of data should be a header row, the first four columns should be (in order) (i) a unique sample number in numeric format; (ii) the control sample id (as designated in Table 3), i.e. DUPA, SSA, DUPB, SSB. (iii) an unspecified column (i.e. the content of the column does not matter but it is important that it is there); and (iv) the laboratory batch number (as a string expression)
    - columns five onwards should contain the element results, the header row will be used to label the plots and included in summary information tables
    - The data should be sorted so that the duplicate-replicate results are grouped together in sets of four, in increasing sample number. For every set the DUPA should be reported first, SSA second, DUPB third and SSB fourth (see Figure 40 and note below).
    - there should be no zeros or minus results in the data
    - paste the worksheet into the duprepplot_blank.xls workbook and name the worksheet "GBASEDUPS". It is important that the worksheet is named exactly this as the macro searches for data in this named worksheet.
    - ****** SAVE THE WORKBOOK AS A NEW FILE ****** (i.e. keep dupreplot_blank.xls as a blank template for others to use)
    - the "GBASEDUPS " worksheet should be copied into the nested_Anova_blank.xls workbook and this file also saved with a new filename.

Note

In order to sort the data into DUPA, SSA, DUPB and SSB order it is best to create two new columns to carry out this procedure. Firstly a column (called "Hundreds") to include the hundred to which the control sample belongs and secondly a column (called "controlid" which replaces DUPA, SSA, DUPB and SSB with 1, 2, 3 and 4 respectively) - see Figure 41. Using the sort function of Excel, sort first on the "Hundreds" column then the "controlid" column and the data will then be in the required order. The "Hundreds" and "controlid" columns can then be deleted.



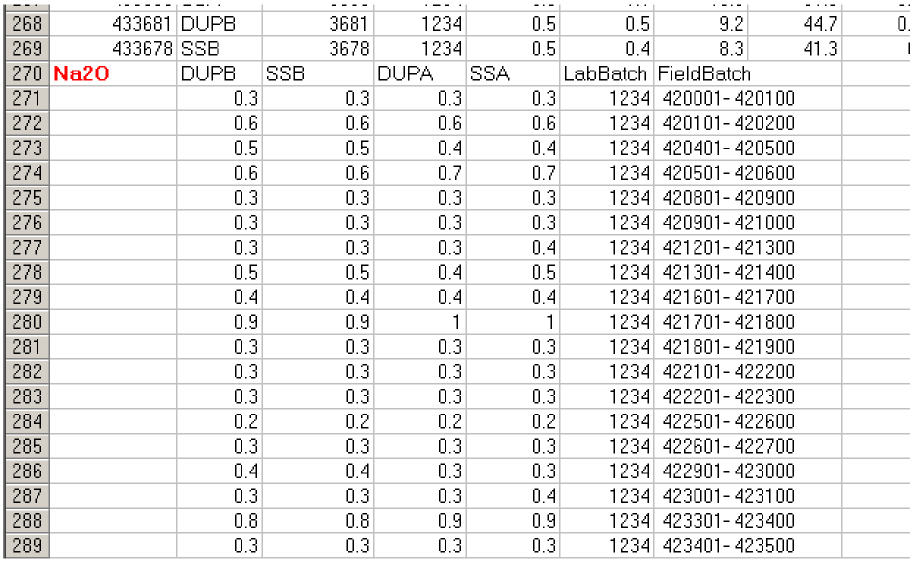
SAMPLE NO	DUP TYPE	Controlid	Hundreds	Na2O	MgO	Al2O3	SiO2	P2O5
420204	DUPA	1	420200	0.2	1.2	20.3	55.1	0.13
420253	SSB	4	420200	0.2	1.1	19.7	55.8	0.14
420262	SSA	2	420200	0.2	1.2	20.2	54.8	0.13
420296	DUPB	3	420200	0.2	1.1	19.6	55.6	0.14
420334	SSB	4	420300	0.5	3.6	12.8	62.5	0.19
420360	SSA	2	420300	0.5	3.3	12.7	66.1	0.18
420366	DUPB	3	420300	0.5	3.7	12.8	62	0.18
420371	DUPA	1	420300	0.5	3.4	12.8	64.4	0.18
420604	DUPA	1	420600	0.2	3.3	17.3	59.5	0.05
420653	SSB	4	420600	0.2	3.3	17.1	59	0.05
420662	SSA	2	420600	0.2	3.3	17.3	58.4	0.05
420696	DUPB	3	420600	0.2	3.3	17.2	58.2	0.06
420734	SSB	4	420700	0.5	1.4	14.8	63	0.12
420760	SSA	2	420700	0.5	1.2	14.7	61.3	0.12
420766	DUPB	3	420700	0.5	1.4	14.9	59.8	0.11
420771	DUPA	1	420700	0.5	1.2	14.6	61.7	0.12
421004	DUPA	1	421000	0.4	1	18.3	57.8	0.1
421053	SSB	4	421000	0.4	1.1	18.9	54.7	0.12
421062	SSA	2	421000	0.4	1	18.2	58.8	0.1

Figure 41: An example illustrating how to sort the duplicate/replicate data into the correct order

The "Hundreds" column contains the formula =INT({sample_No cell ref}/100)*100

Running the macro

- first select the data to be used. This should include the header row, all data rows and columns
- activate the macro with ctrl+q
- data will be reformatted as shown in Figure 42
- a chart will be created for every element column selected (e.g. Figure 43)



268	433681	DUPB	3681	1234	0.5	0.5	9.2	44.7	0.
269	433678	SSB	3678	1234	0.5	0.4	8.3	41.3	
270	**Na2O**	DUPB	SSB	DUPA	SSA	LabBatch	FieldBatch		
271		0.3	0.3	0.3	0.3	1234	420001- 420100		
272		0.6	0.6	0.6	0.6	1234	420101- 420200		
273		0.5	0.5	0.4	0.4	1234	420401- 420500		
274		0.6	0.6	0.7	0.7	1234	420501- 420600		
275		0.3	0.3	0.3	0.3	1234	420801- 420900		
276		0.3	0.3	0.3	0.3	1234	420901- 421000		
277		0.3	0.3	0.3	0.4	1234	421201- 421300		
278		0.5	0.5	0.4	0.5	1234	421301- 421400		
279		0.4	0.4	0.4	0.4	1234	421601- 421700		
280		0.9	0.9	1	1	1234	421701- 421800		
281		0.3	0.3	0.3	0.3	1234	421801- 421900		
282		0.3	0.3	0.3	0.3	1234	422101- 422200		
283		0.3	0.3	0.3	0.3	1234	422201- 422300		
284		0.2	0.2	0.2	0.2	1234	422501- 422600		
285		0.3	0.3	0.3	0.3	1234	422601- 422700		
286		0.4	0.4	0.3	0.3	1234	422901- 423000		
287		0.3	0.3	0.3	0.4	1234	423001- 423100		
288		0.8	0.8	0.9	0.9	1234	423301- 423400		
289		0.3	0.3	0.3	0.3	1234	423401- 423500		

Figure 42: Example of reformatted duplicate-replicate data added to the worksheet by the macro

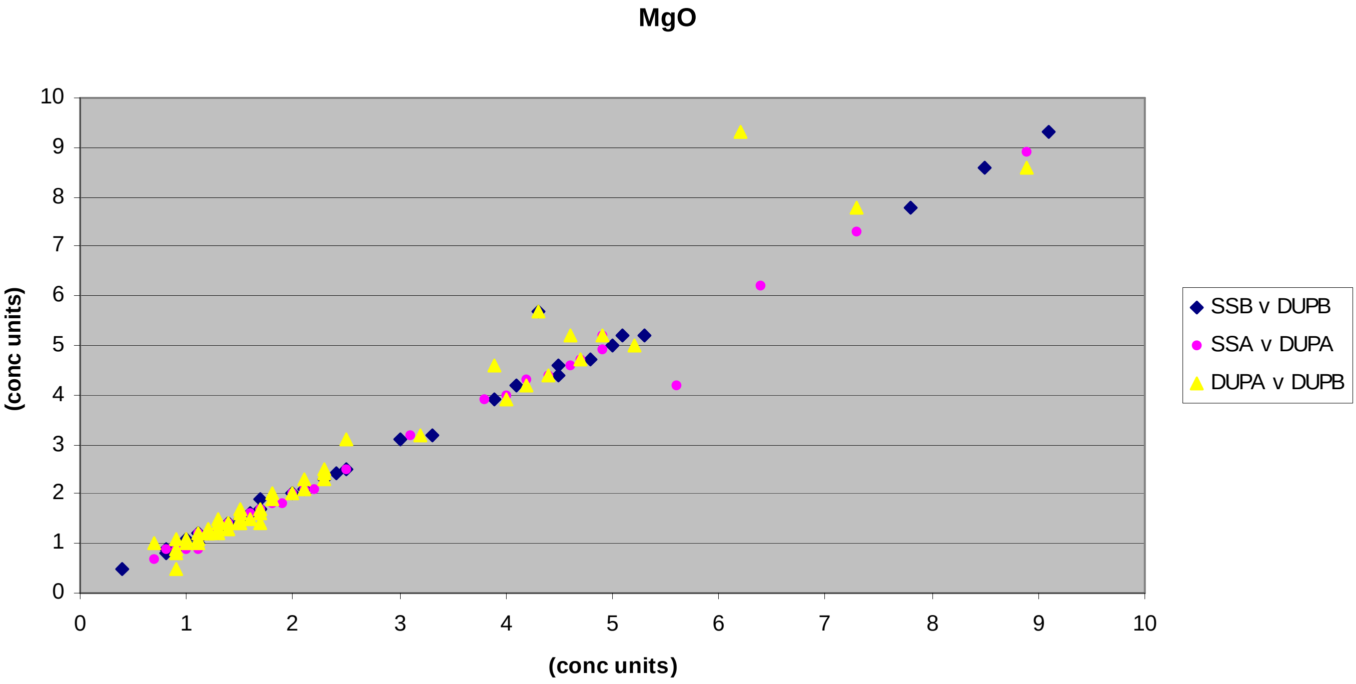


Figure 43: An example chart plotted by the macro

Listing of xyplot macro

```
Sub reformatter()
' reformatter macro to reformat Maroc dup and rep samples and plot x-y plots
' chris johnson 23/11/99
' updated for use with G-BASE project 10/11/05

Dim norows, nocols, ro, col As Integer
Dim Sampnumb(4)  As String
Dim LabBatch(4) As String

' user must select area first
' count the number of rows in the selection
norows = Selection.Rows.Count
nocols = Selection.Columns.Count

If (norows - 1) / 4 <> Int((norows - 1) / 4) Then 'not a complete set of dups and reps
   MsgBox "You do not have the correct number of rows for a complete set of control
samples. YOU MUST HAVE A DUPS AND REPS IN SETS OF FOUR PLUS A ROW OF
COLUMN HEADERS"
   End
Else ' say the selection is OK
   MsgBox "The selection contains " & _
      norows & " rows and " & _
      nocols - 4 & " element columns which is a valid area of selection"
End If
nocontrols = (norows - 1) / 4
noelements = nocols - 4
'use column headers to name blocks of data

For col = 5 To nocols
     xtra = nocontrols + 1
     x = (norows + 1) + ((col - 5) * xtra)
' set headings for reformatted element data
      element = ActiveSheet.Cells(1, col).Value
      ActiveSheet.Cells(x, 1).Value = element
      ActiveSheet.Cells(x, 1).Font.Color = RGB(255, 0, 0)
      ActiveSheet.Cells(x, 1).Font.Bold = True
      ActiveSheet.Cells(x, 2).Value = "DUPB"
      ActiveSheet.Cells(x, 3).Value = "SSB"
      ActiveSheet.Cells(x, 4).Value = "DUPA"
      ActiveSheet.Cells(x, 5).Value = "SSA"
      ActiveSheet.Cells(x, 6).Value = "LabBatch"
      ActiveSheet.Cells(x, 7).Value = "FieldBatch"
Next col

' start on 2nd row of sheet and reformat data in blocks of 4
For ro = 2 To norows Step 4

   For n = 1 To 4
                 Sampnumb(n) = ActiveSheet.Cells(ro + n - 1, 1).Value
         LabBatch(n) = ActiveSheet.Cells(ro + n - 1, 4).Value
' determine what type of sample it is
' this procedure is based on standard GBASE Checklists
' the following should be edited if different checklists are used
' Details of control sample numbers
'**********************************

      Select Case Val(Right(Sampnumb(n), 2))
        Case 53, 58, 78 'SSB
           y = 3
```

```
        Case 62, 77, 86 'SSA
          y = 5
        Case 37, 81, 96 'DUPB
          y = 2
        Case 4, 31, 76  'DUPA
          y = 4
'************************************
        Case Else   'not a regognised dup or rep number
          MsgBox "The sample number " + Sampnumb(n) + " is not a recognised control
sample number. " + Val(Sampnumb(n)) - (((ro - 2) / 4) * 100)
          End
        End Select
' place element result from original cell into a new location on the worksheet
        For col = 5 To nocols
        xtra = ((norows - 1) / 4) + 1
        x = (norows + 2 + ((ro - 2) / 4)) + ((col - 5) * xtra)
        ActiveSheet.Cells(x, y).Value = ActiveSheet.Cells(ro + (n - 1), col).Value
        ActiveSheet.Cells(x, 6).Value = LabBatch(n)
' work out which batch of 100 samples
          If n = 4 Then
          n1 = (Int(Sampnumb(n) / 100) * 100) + 1
          n2 = n1 + 99
          hundreds = Str(n1) + "-" + Str(n2)
          ActiveSheet.Cells(x, 7).Value = hundreds
          End If
        Next col

    Next n

Next ro

'start x-y plots

 startrow = norows + 1
 endrow = startrow + nocontrols

 ' 3 plots on each graph, namely, SSB v DUPB, SSA v DUPA and DUPB v DUPA

  For n = 1 To noelements
   elementrange = "B" + LTrim(Str(startrow)) + ":" + "E" + LTrim(Str(endrow))
   series1xv = "R" + LTrim(Str(startrow + 1)) + "C3:R" + LTrim(Str(endrow)) + "C3"
   series1v = "R" + LTrim(Str(startrow + 1)) + "C2:R" + LTrim(Str(endrow)) + "C2"
   series2xv = "R" + LTrim(Str(startrow + 1)) + "C5:R" + LTrim(Str(endrow)) + "C5"
   series2v = "R" + LTrim(Str(startrow + 1)) + "C4:R" + LTrim(Str(endrow)) + "C4"
   series3xv = "R" + LTrim(Str(startrow + 1)) + "C4:R" + LTrim(Str(endrow)) + "C4"
   series3v = "R" + LTrim(Str(startrow + 1)) + "C2:R" + LTrim(Str(endrow)) + "C2"
   If n <> 1 Then
   locname = "A" + LTrim(Str(startrow - 1))
   Else
   locname = "A" + LTrim(Str(startrow))
   End If

   elementname = Worksheets("GBASEDUPS").Range(locname).Value

   Charts.Add
   ActiveChart.ChartType = xlXYScatter
   ActiveChart.SetSourceData Source:=Sheets("GBASEDUPS").Range(elementrange), _
     PlotBy:=xlColumns

   ActiveChart.SeriesCollection(1).Name = "=""SSB v DUPB"""
   ActiveChart.SeriesCollection(1).XValues = "=GBASEDUPS!" + series1xv
```

```
    ActiveChart.SeriesCollection(1).Values = "=GBASEDUPS!" + series1v
    ActiveChart.SeriesCollection(2).XValues = "=GBASEDUPS!" + series2xv
    ActiveChart.SeriesCollection(2).Values = "=GBASEDUPS!" + series2v
    ActiveChart.SeriesCollection(2).Name = "=""SSA v DUPA"""
    ActiveChart.SeriesCollection(3).XValues = "=GBASEDUPS!" + series3xv
    ActiveChart.SeriesCollection(3).Values = "=GBASEDUPS!" + series3v
    ActiveChart.SeriesCollection(3).Name = "=""DUPA v DUPB"""
    ActiveChart.Location Where:=xlLocationAsNewSheet, Name:=elementname
    With ActiveChart
        .HasTitle = True
        .ChartTitle.Characters.Text = elementname
        .Axes(xlCategory, xlPrimary).HasTitle = True
        .Axes(xlCategory, xlPrimary).AxisTitle.Characters.Text = "(conc units)"
        .Axes(xlValue, xlPrimary).HasTitle = True
        .Axes(xlValue, xlPrimary).AxisTitle.Characters.Text = "(conc units)"
    End With
    ActiveChart.PlotArea.Select
    With Selection.Border
        .ColorIndex = 16
        .Weight = xlThin
        .LineStyle = xlContinuous
    End With
    With Selection.Interior
    .ColorIndex = 15
        .PatternColorIndex = 1
        .Pattern = xlSolid
    End With
    ActiveChart.SeriesCollection(2).Select
    With Selection.Border
        .Weight = xlHairline
        .LineStyle = xlNone
    End With
    With Selection
        .MarkerBackgroundColorIndex = xlAutomatic
        .MarkerForegroundColorIndex = xlAutomatic
        .MarkerStyle = xlCircle
        .Smooth = False
        .MarkerSize = 3
        .Shadow = False
    End With

    startrow = endrow + 2
 endrow = startrow + nocontrols - 1

Next n
End Sub
```

NESTED ANOVA ANALYSIS

Summary: This macro is described in a BGS Internal Report by Johnson (2002) and users should refer to this more detailed report for further information. The theory behind the analysis and formulae used in the macro were originally based and tested on the nested ANOVA example given by Sinclair (1983) (taking into account typographical errors found in Table 3-VII of this work). In order to make the data have a more "normal" distribution (see discussion in Section 5.1) the macro log-transforms the data during before producing the final output table.

The workbook containing the ANOVA macro and worksheet created in the procedure described above is opened and minor modifications made to the layout of the results. On executing the macro (crtl + shift + r) nested ANOVA analysis is carried out on the duplicate-replicate pairs. The calculations for each element are written onto a sheet for each element and a summary table is made that attributes the variance found for each element (Table 18).

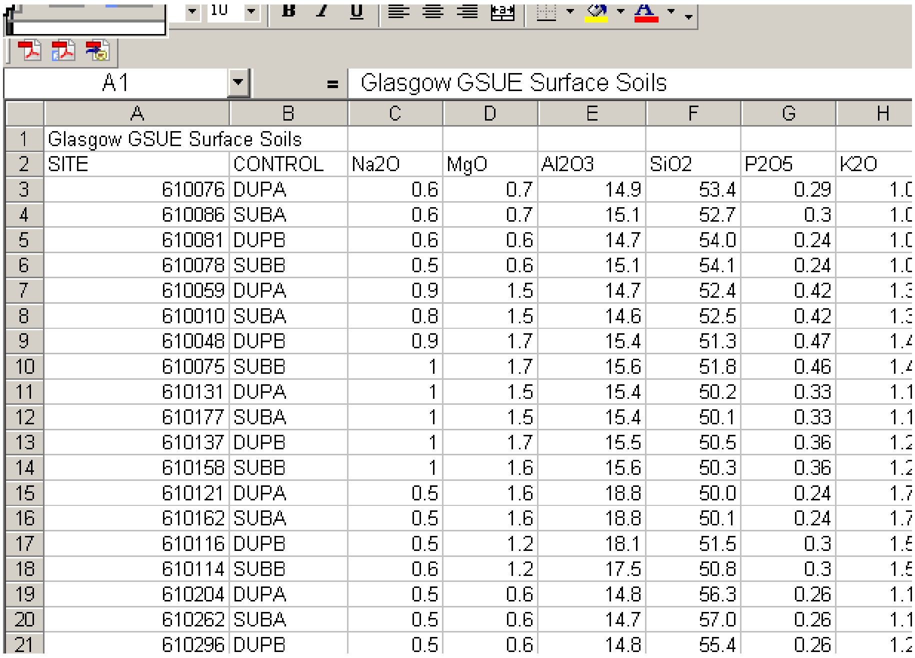
Procedure:

- in the ANOVA workbook, data worksheet, insert a title on line 1 and delete columns 3 and 4 of the data so the format is exactly as shown in Figure 44, remembering that the set of four related controls must be in the order DUPA, SSA, DUPB and SSB
- **The worksheet containing the results must be named “data”.** It is important that the worksheet is named exactly this as the macro searches for data in this named worksheet.
- activate the macro with CTRL+shift+r

Glasgow GSUE Surface Soils

Element	Between Site %	Between Sample %	Within Sample %	Total %
Na2O	97.33	1.81	0.86	100
MgO	91.68	7.35	0.98	100
Al2O3	95.18	4.18	0.64	100
SiO2	91.36	4.04	4.60	100
P2O5	90.38	9.06	0.56	100
K2O	94.96	4.64	0.40	100
CaO	90.31	9.46	0.23	100
TiO2	97.19	2.51	0.30	100
MnO	94.26	5.23	0.51	100
Fe2O3	95.97	3.90	0.13	100
Sc	87.79	6.39	5.82	100
V	96.41	3.12	0.47	100
Cr	90.03	5.91	4.06	100
Co	87.98	4.34	7.68	100
Ba	88.23	8.71	3.06	100
Ni	94.42	5.21	0.37	100
Cu	77.75	21.96	0.30	100
Zn	62.15	37.23	0.63	100
Ga	91.91	6.28	1.81	100
Ge	93.39	3.70	2.91	100
As	90.74	5.46	3.80	100
Se	80.28	7.50	12.22	100
Br	95.23	4.18	0.59	100
Rb	93.00	6.56	0.44	100
Sr	92.25	7.56	0.20	100
Y	93.14	4.67	2.19	100
Zr	94.08	4.01	1.91	100
Nb	96.73	2.81	0.46	100
Mo	84.75	5.83	9.42	100
Hf	58.44	0.13	41.43	100
Ta	14.40	9.99	75.60	100
W	63.23	7.64	29.13	100
Tl	31.49	16.48	52.02	100
Pb	88.61	10.39	1.00	100
Bi	-2.73	86.16	16.56	100
Th	88.43	5.56	6.01	100
U	70.03	-4.07	34.04	100
Ag	62.95	7.32	29.74	100
Cd	59.16	0.00	40.84	100
Sn	63.59	22.46	13.96	100
Sb	77.64	6.28	16.08	100
Te	754.84	-1590.32	935.48	100
I	90.17	5.57	4.25	100
Cs	76.40	4.44	19.16	100
La	92.22	4.33	3.45	100
Ce	92.70	4.03	3.27	100

Table 18: Summary table attributing element variance produced by the nested ANOVA macro



A1 = Glasgow GSUE Surface Soils

	A	B	C	D	E	F	G	H
1	Glasgow GSUE Surface Soils							
2	SITE	CONTROL	Na2O	MgO	Al2O3	SiO2	P2O5	K2O
3	610076	DUPA	0.6	0.7	14.9	53.4	0.29	1.0
4	610086	SUBA	0.6	0.7	15.1	52.7	0.3	1.0
5	610081	DUPB	0.6	0.6	14.7	54.0	0.24	1.0
6	610078	SUBB	0.5	0.6	15.1	54.1	0.24	1.0
7	610059	DUPA	0.9	1.5	14.7	52.4	0.42	1.3
8	610010	SUBA	0.8	1.5	14.6	52.5	0.42	1.3
9	610048	DUPB	0.9	1.7	15.4	51.3	0.47	1.4
10	610075	SUBB	1	1.7	15.6	51.8	0.46	1.4
11	610131	DUPA	1	1.5	15.4	50.2	0.33	1.1
12	610177	SUBA	1	1.5	15.4	50.1	0.33	1.1
13	610137	DUPB	1	1.7	15.5	50.5	0.36	1.2
14	610158	SUBB	1	1.6	15.6	50.3	0.36	1.2
15	610121	DUPA	0.5	1.6	18.8	50.0	0.24	1.7
16	610162	SUBA	0.5	1.6	18.8	50.1	0.24	1.7
17	610116	DUPB	0.5	1.2	18.1	51.5	0.3	1.5
18	610114	SUBB	0.6	1.2	17.5	50.8	0.3	1.5
19	610204	DUPA	0.5	0.6	14.8	56.3	0.26	1.1
20	610262	SUBA	0.5	0.6	14.7	57.0	0.26	1.1
21	610296	DUPB	0.5	0.6	14.8	55.4	0.26	1.2

Figure 44: Figure showing the required layout for results in the data worksheet for the ANOVA macro

Appendix 4 :Accepted values for G-BASE secondary reference materials (SRMs)

Sample ID	Na2O %	MgO %	Al2O3 %	SiO2 %	P2O5 %	K2O %	CaO %	TiO2 %	MnO %	Fe2O3 %	Ag ppm	Cd ppm	Sn ppm	Analysis Date
S13	0.3	1.2	20.0	57.1	0.1	2.17	0.35	0.82	0.128	6.88	0.6	0.7	3.0	May 2001 - July 2002
S13B	0.3	1.3	20.7	56.8	0.1	2.29	0.57	0.81	0.108	7.46	0.9	0.5	2.9	May 2001 - July 2002
S15	0.7	0.6	8.1	64.4	0.1	2.27	0.20	0.39	0.082	1.88	0.9	0.6	4.6	May 2001 - July 2002
S15B	0.5	1.7	14.9	60.2	0.1	2.94	0.71	0.67	0.162	6.96	0.8	0.3	3.4	May 2001 - July 2002
S22B	0.8	2.3	14.1	47.4	0.2	2.48	1.05	1.090	0.968	10.16	1.0	7.8	14.1	June 2004
S23B	1.3	2.0	15.2	52.3	0.3	3.86	2.09	0.95	0.393	7.06	0.9	1.1	6.2	May 2001 - July 2002
S24	0.9	1.1	25.7	48.2	0.2	3.40	0.32	1.12	0.458	10.22	4.6	3.0	5.5	May 2001 - July 2002
S24B	0.9	1.4	22.7	51.0	0.2	3.22	0.37	1.101	0.550	9.73	2.5	1.6	4.6	June 2004
S57A	0.5	0.4	8.4	76.9	0.30	1.92	0.51	0.474	0.058	2.63	n/d	0.3	2.9	June 2003 - January 2005
S58S	0.1	1.3	15.4	50.4	0.07	2.04	0.79	0.867	0.078	6.32	n/d	n/d	2.6	June 2003 - October 2003
Sample ID	Sb ppm	Te ppm	I ppm	Cs ppm	Ba ppm	La ppm	Ce ppm	S ppm	Cl ppm	Sc ppm	V ppm	Cr ppm	Co ppm	Analysis Date
S13	0.2	0.5	1	n/d	1703	51	102	n/d	n/d	12	97	98	28.5	May 2001 - July 2002
S13B	n/d	0.5	1	4	744	48	100	n/d	n/d	14	99	100	24.6	May 2001 - July 2002
S15	1.1	0.6	2	0	291	19	38	n/d	n/d	4	35	41	8.5	May 2001 - July 2002
S15B	3.1	0.5	1	3	609	36	74	n/d	n/d	11	93	85	19.4	May 2001 - July 2002
S22B	9.7	2.5	23.4	43	523	59	91	855	129	18.5	148.4	39.5	43.4	June 2004
S23B	4.2	0.5	4	6	667	64	137	n/d	n/d	12	115	60	24.4	May 2001 - July 2002
S24	7.9	0.5	8	n/d	983	69	141	n/d	n/d	21	140	123	97.2	May 2001 - July 2002
S24B	13.7	n/d	19.4	11	1189	71	137	945	70	19.0	136.1	116.1	119.5	June 2004
S57A	1.1	0.3	3.0	4	300	28	53	472	76	5.0	37.6	45.5	6.9	June 2003 - January 2005
S58S	0.3	0.2	5.4	7	334	46	92	750	133	14.3	152.9	107.3	18.0	June 2003 - October 2003
Sample ID	Ni ppm	Cu ppm	Zn ppm	Ga ppm	Ge ppm	As ppm	Se ppm	Br ppm	Rb ppm	Sr ppm	Y ppm	Zr ppm	Nb ppm	Analysis Date
S13	36.4	16.5	113.1	20.5	1.3	15.2	0.5	6.3	96.3	120.3	20.8	233.8	13.5	May 2001 - July 2002
S13B	45.7	17.1	99.7	21.9	1.5	14.0	0.4	5.0	111.4	122.0	22.6	159.6	14.0	May 2001 - July 2002
S15	12.1	5.6	29.9	5.9	0.5	9.0	0.1	8.2	66.3	56.3	14.3	529.9	7.1	May 2001 - July 2002
S15B	33.3	13.0	63.7	14.0	1.5	18.7	0.6	11.6	106.2	86.8	26.1	549.1	12.8	May 2001 - July 2002
S22B	38.2	270.8	1098.4	24.0	n/d	3404.8	n/d	87.4	196.0	47.4	75.9	379.4	19.3	June 2004
S23B	20.9	59.2	115.7	22.7	1.2	92.7	0.1	17.3	178.3	191.6	26.2	453.7	22.4	May 2001 - July 2002
S24	45.2	64.0	387.3	36.3	3.7	123.6	0.8	42.9	163.0	138.0	35.7	147.8	21.7	May 2001 - July 2002
S24B	55.5	90.4	575.6	31.2	3.4	151.6	1.2	70.5	157.7	138.1	39.7	163.6	20.4	June 2004
S57A	13.9	16.3	58.3	8.0	1.1	10.8	0.3	7.3	78.8	69.2	18.5	539.6	8.9	June 2003 - January 2005
S58S	50.6	19.0	71.0	18.1	1.4	25.0	0.4	7.2	95.3	59.3	29.0	351.4	17.8	June 2003 - October 2003
Sample ID	Mo ppm	Nd ppm	Sm ppm	Yb ppm	Hf ppm	Ta ppm	W ppm	Tl ppm	Pb ppm	Bi ppm	Th ppm	U ppm		Analysis Date
S13	1.6	n/d	n/d	n/d	6.2	0.8	1.5	0.6	109.1	0.2	9.9	2.5		May 2001 - July 2002
S13B	1.5	n/d	n/d	n/d	4.5	0.9	1.2	0.7	62.9	0.5	10.6	2.8		May 2001 - July 2002
S15	0.7	n/d	n/d	n/d	11.6	0.8	2.1	0.3	24.4	0.8	4.7	1.2		May 2001 - July 2002
S15B	0.6	n/d	n/d	n/d	12.4	0.9	2.6	0.4	81.8	0.5	9.6	2.4		May 2001 - July 2002
S22B	11.2	72.5	13.2	9.5	7.8	n/d	1147.8	n/d	439.7	95.4	6.1	7.8		June 2004
S23B	30.8	n/d	n/d	n/d	11.4	1.2	36.0	1.0	112.4	5.6	26.3	37.7		May 2001 - July 2002
S24	1.9	n/d	n/d	n/d	4.2	1.1	2.6	2.2	1069.6	n/d	16.2	1.7		May 2001 - July 2002
S24B	1.1	56.3	7.6	2.3	3.1	1.6	5.1	2.0	1929.1	1.6	17.4	n/d		June 2004
S57A	0.7	19.8	4.3	2.3	12.1	0.6	2.0	0.9	37.3	0.5	8.4	3.1		June 2003 - January 2005
S58S	0.5	40.8	7.0	3.0	9.2	1.2	2.9	0.7	23.8	0.8	13.6	3.6		June 2003 - October 2003

Appendix 5 : Levelling factors applied to "within" Atlas Areas

EAST MIDLANDS

Specific batch corrections required for W and U

Analyte	Levelling factor
W	Lab Number 6206 (W x 1.1341) – 2.3899
	Lab Number 6323 (W x 1.1832) + 0.4756
	Lab Number 6399 (W x 1.1258) – 0.7944
	Lab Number 6487 (W x 1.2162) + 0.7881
	Lab Number 6661 (W x 1.2053) + 0.5452
	Lab Number 6662 (W x 1.1959) + 0.2359
	Lab Number 6663 (W x 1.1838) + 0.6528
	Lab Number 6670 (W x 1.1857) – 1.2015
	Lab Number 6671 (W x 1.1826) – 1.0816
	Lab Number 6672 (W x 1.1629) – 1.336
	Lab Number 6673 (W x 1.1729) – 1.1172
	Lab Number 6681 (W x 1.1536) + 0.2843
	Lab Number 6682 (W x 1.1391) + 0.6405
	Lab Number 6683 (W x 1.1431) + 0.2179
	Lab Number 6725 (W x 1.1485) + 0.8631
	Lab Number 6726 (W x 1.1355) + 0.3678
	Lab Number 6727 (W x 1.1436) + 0.6371
	Lab Number 6728 (W x 1.1339) + 0.3048
	Lab Number 6729 (W x 1.147) + 0.4059
	Lab Number 6730 (W x 1.1458) + 0.407
	Lab Number 6761 (W x 1.1492) – 2.2638
U	Lab Number 6206 (U x 0.9617) + 1.061
	Lab Number 6323 (U x 0.9986) + 0.3529
	Lab Number 6399 (U x 0.9882) – 0.5555
	Lab Number 6487 (U x 1.0029) – 0.0078
	Lab Number 6661 (U x 1.0086) + 0.3807
	Lab Number 6662 (U x 1.0208) – 0.1947
	Lab Number 6663 (U x 1.0308) – 0.6466
	Lab Number 6670 (U x 0.9829) + 0.147
	Lab Number 6671 (U x 0.9846) – 0.6001
	Lab Number 6672 (U x 1.0144) – 0.6451
	Lab Number 6673 (U x 1.0086) – 0.4258
	Lab Number 6681 (U x 0.9872) + 0.4789
	Lab Number 6682 (U x 0.9915) + 0.4196
	Lab Number 6683 (U x 0.9971) + 0.3075
	Lab Number 6725 (U x 1.0159) + 0.5171
	Lab Number 6726 (U x 0.9983) – 0.242
	Lab Number 6727 (U x 0.979) + 0.105
	Lab Number 6728, Sample numbers 432001 – 432100 (U x 1.0055) + 1.189
	Lab Number 6728, All other samples (U x 1.0055) + 0.189
	Lab Number 6729 (U x 1.0042) + 0.1413
	Lab Number 6730 U + 1.5
	Lab Number 6761, Sample numbers 431102 – 431543 (U x 0.99) + 1.4729
	Lab Number 6761, All other samples (U x 0.99) + 0.4729

Levelling applied to other elements

Analyte	**Levelling factor**
MgO	(MgO x 1.0089) – 0.1956
Al_2O_3	(Al_2O_3 x 1.0011) + 0.2285
K_2O	(K_2O x 1.0019) – 0.2258
MnO	(MnO x 0.9783) + 0.0086
Fe_2O_3	(Fe_2O_3 x 1.0286) + 0.3602
Cr	(Cr x 0.8071) + 5.5999
Co	(Co x 1.1452) + 6.3301
Cu	(Cu x 0.9607) + 0.4018
Zn	(Zn x 0.8951) + 0.7901
Rb	(Rb x 0.99) – 2.7947
Sr	(Sr x 0.9163) + 2.1896
Y	(Y x 0.9855) + 2.3335
Zr	(Zr x 1.1602) – 39.745
Mo	Mo +1
Hf	(Hf x 1.1241) – 0.2392
Ta	(Ta x 0.4666) + 0.4814
Tl	(Tl x 1.3745) + 0.0454
Th	(Th x 1.0055) –1.4612
Ag	(Ag x 2.532) + 0.032
Cd	(Cd x 1.4985) – 0.0946
Sn	(Sn x 0.9541) – 0.0658
Sb	(Sb x 0.9395) – 0.9506
Cs	(Cs x 1.1596) – 3.0665
La	(La x 1.2646) – 8.0621
Ce	Lab Number 6761 (Ce x 1.084) + 7.1835
	Then ALL (Ce x 1.1993) – 7.1401

Glossary

The following terms are used in this report with respect to their usage by geochemists conditioning of G-BASE geochemical data. Words defined elsewhere in the glossary are highlighted in red.

Accepted value	It is never possible to determine the exact concentration (true value) of an element in a sample due to limitations of the analytical method. The result determined by one or more analyses is the measured value. After repeated analyses an accepted value can be computed and the sample can be certified as having that element concentration. Different analytical methodologies will have differing accepted values so the accepted value should always be quoted in the context of the analytical method used.
Accuracy	Accuracy measures how close to a true or accepted value a measurement lies. This can be seen graphically on a Shewhart plot.
ANOVA	Analysis of Variance (ANOVA) is a statistical procedure dependent on the fact that the total variability in a data set can be attributed to various sources. With the use of duplicate and replicate samples random nested analysis of variance can be performed to attribute variability to within a site, "between" and "within" samples.
Blind sample	A blind sample is a control sample that has been submitted for analysis presented in such a way that its identity is unknown to the analyst.
Bias	Bias is the tendency to favour one analytical value over another. Analytical errors fall into two major categories: bias (systematic error) and variability (random error). Bias causes consistently positive or negative deviation in the results from the accepted value. Repeated measurement of SRMs over time provide evidence of both inter- and intra-batch systematic bias and random variability in the laboratory analytical procedures.
Censored data	When results (usually reported as semi-quantitative values) are replaced by a substitute value, the data can be described as censored data.

Conditioned data	Conditioned data are raw data that have been subjected to data conditioning.
Conditioning	Conditioning is the process of making data fit for the purpose for which it is to be used following the QA procedures documented in this report It can represent the accumulation of error checking, verification, quality control, quality assurance and levelling processes. It is important that conditioned data is accompanied by a statement as to what processes it has been subjected to.
Control chart	A control chart is a graphical representation (plot) showing how the value of a sample varies over a period of time in relation to an accepted value or range of values as defined by a control sample. A particular type of control chart showing how a value varies over time within an envelope of mean ± n standard deviations is referred to as a Shewhart plot. Both these can also be more generally referred to as a time-series plot.
Control sample	A control sample is one that is inserted into a batch of samples during the process of sampling or analysis for the purpose monitoring error, precision and accuracy. Examples of control samples are duplicates, replicates, SRMs and PRMs.
Drift	Drift is the gradual systematic change over a period of time relative to the accepted value.
Duplicate sample	A duplicate sample is collected from the same site as another sample as defined in the protocol for collecting duplicate samples (see Johnson, 2005). It is a control sample that can be used to show the variability in results that can be attributed to the process of sampling by collecting two samples from the same location. A duplicate sample collected in the field is sometimes also referred to as a "field duplicate".
Error	Deviation from what is believed to be correct, right, or true is the error of a measurement, i.e. the measured result minus the true value.
G-BASE	Geochemical Baseline Survey of the Environment project. A British Geological Survey Project tasked with making geochemical maps of the surface environment for Great Britain and Northern Ireland.

Geochemistry Database This is the British Geological Survey (BGS) corporate database in which the majority of BGS geochemical data is stored. This Oracle™ database is described by Coats (2004).

International reference standard This is a control sample for which there are internationally certified values. All geochemical data reported in peer-reviewed publications should report reference standard results in order to give readers comprehension of the accuracy of the results. These samples are commercially available at costs £600 - £6000 per kilogram and are also known as Primary Reference Materials (PRMs) or Certified Reference Materials (CRMs).

Laboratory batch When samples are submitted for analyses they are grouped together to form a laboratory batch. The number of samples in each batch will vary according to the sample type and analytical method. Each batch is assigned a unique laboratory batch number which must be associated with metadata such as date of analysis, analytical calibration, detection limits, etc..

Levelling This is the process whereby disparate data sets are combined to form a single discrete data set. In G-BASE this is achieved using normalisation of results using SRMs that are repeatedly analysed in each laboratory batch.

Levelling factor This is a mathematical function that is applied to results in order to combine data sets into a single discrete data set and is determined during the process of data conditioning by normalisation of the results from SRMs repeatedly determined in each analytical batch over a period of time.

LIMS Laboratory Information Management System

Lower Limit of detection In general terms this is the concentration at which the analyst's instrument gives a significantly different signal to that of a "background" or blank signal. Different analytical methods will define it in different ways and it can be quite arbitrary. A value that is above the lower limit of detection can be regarded as quantifiable and reproducible. It is important that a value is quotable for each laboratory batch of samples analysed.

Lower Limit of reporting This is a quantitative value representing the lowest measurement that the analyst is prepared to report to the client.

Measured value	A quantitative result reported by the analyst following analysis of a sample.
Normalisation	The G-BASE data conditioning process uses the term normalisation in a mathematical sense, i.e. *"to adjust the representation of a quantity so that this representation lies within a prescribed range (Parker, 1974), or, any process of rescaling a quantity so that a given integral or other functional of the quantity takes on a pre-determined value (Morris, 1991)"* rather than in the statistical sense, where it denotes a transformation of a data set so that it has a mean of zero and a variance of one. Normalisation is a process to determine levelling factors carried out using control samples.
Precision	Precision is a measurement of how closely the analytical results can be reproduced. It should not be confused with the term accuracy. Results can have a good precision (i.e. consistently fall at or near a specified value), yet the mean of these results may be a long way off the accepted value.
QA	This is an abbreviation for the term "Quality Assurance". QA equates to the overall G-BASE data conditioning procedure. It is a system of protocols, checks, audits, and corrective actions to ensure that all analytical results prepared for the Geochemistry Database are of high and consistent quality.
QC	This is an abbreviation for the term "Quality Control" a process that is part of the overall G-BASE data conditioning exercise. It is a collection of documented procedures applied to the raw data to continuously assess whether the laboratory is producing results of acceptable quality as assessed by the inclusion of control samples in all procedures from sampling through to analysis.
PRM	This is an abbreviation for Primary Reference Material. This is a control sample, usually an international reference standard, which the project requests that the laboratory includes for analysis at the start and end of every laboratory batch.
Raw data	The results as received from the analyst.
Replicate Sample	This is a control sample created in the laboratory by dividing a sample into two identical parts according to a well-defined protocol. It is used to help define laboratory error. In the G-BASE field database this sample is also referred to as a "sub-sample".

Sensitivity	Analytical sensitivity is the lowest concentration that can be distinguished from background noise or more correctly called the assay's detection limit. When a technique is described as being sensitive the implication is it has a low limit of detection.
Shewhart plot	This a control chart or time series plot with defined quality limits named after the person who first documented their use (Shewhart, 1931).
Shift	Shift is a significant sudden change in a measured value compared with the previous measurement of the same measurand. This can be seen graphically on a time-series or Shewhart plot and is usually the consequence of an instrument recalibration. It differs from drift in that the change in measurement is large and sudden.
SRM	This is an abbreviation for Secondary Reference Material, a control sample of prime importance in the G-BASE data conditioning procedure. It differs from a PRM in that it is not internationally certified (and therefore less expensive to use), it is submitted blind to the analyst and is more likely to have a matrix/mineralogical composition similar to the routine samples being analysed. In other contexts, SRM is an abbreviation of Standard Reference Material, a more generic term for reference samples
Time-series plot	When a measurement is repeatedly determined over a time period, a plot of quantity against time graphically shows how the result is varying over time. A Shewhart plot is an example of a time-series plot.
Variability	Variability is a random error that affects the ability to reproduce results (see bias and precision).
Verification	Verification is that part of the G-BASE data conditioning procedure that checks that the laboratory has analysed and reported all the samples submitted to the specifications of the analytical request form (or in the case of an external laboratory, as detailed in the contract).

References

ALBERT, R H AND HORWITZ W. 1995. Incomplete Data Sets: Coping with Inadequate Databases. Journal of Association of Analytical Chemists (AOAC) International, 78(6), 1513-1515.

AMC (Analytical Methods Committee). 2001. Measurement of near zero concentration: recording and reporting results that fall close to or below the detection limit. Analyst, 126, 256-259.

ANDER, E L. In Prep. G-BASE quality control and quality assurance procedures for stream water analyses. *British Geological Survey*, BGS Internal Report Number IR/06/XXX.

COATS, J S. 2004. The BGS Geochemistry Database: history, design and current usage. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/04/033.

DARNLEY, A G, BJORKLUND, A, BOLVIKEN, B, GUSTAVSSON, N, KOVAL, P V, PLANT, J A, STEENFELT, A, TAUCHID, M, AND XUEJING, XIE. 1995. A global geochemical database for environmental and resource management. 19, UNESCO publishing, Paris.

JOHNSON, C C. 2002. Within site and between site nested analysis of variance (ANOVA) for Geochemical Surveys using MS EXCEL. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/02/043.

JOHNSON, C C, COATS, J S, BREWARD, N, ANDER, E L AND MACKENZIE, A C. 2004. geochemical data as a standard reference data set for the SIGMA project. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/04/026.

JOHNSON, C C. 2005. 2005 G-BASE Field Procedures Manual. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/05/097.

LISTER, T R. 2002a. Analysis of Variance (ANOVA) of G-BASE sub-surface soil data from 11 urban centres in England and Wales. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/02/009.

LISTER, T R. 2002b. Quality control of G-BASE data. Procedures followed for XRF data from Wales and Welsh Borders area stream sediment samples. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/02/105.

LISTER, T R, FLIGHT, D M A, BROWN, S E, JOHNSON, C C AND MACKENZIE, A C. 2005. The G-BASE field database. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/05/001.

MORRIS, C.G. 1991. Academic Press dictionary of science and technology. Academic press, San Diego, 2432p.

PARKER, S.B. 1974. McGraw-Hill dictionary of scientific and technical terms. McGraw-Hill, New York, 2088p.

PLANT, J, JEFFREY, K, GILL, E AND FAGE, C. 1975. The systematic determination of accuracy and precision in geochemical exploration data. Journal of Geochemical Exploration, 4, 467-486.

RAWLINS, B G, O'DONNELL, K E, and INGHAM, M. 2003. Geochemical survey of the Tamar catchment (south-west England). *British Geological Survey, Keyworth, Nottingham, UK*, Commissioned Report CR/03/027.

REIMANN, C. 2005. Sub-continental-scale geochemical mapping: sampling, quality control and data analysis issues. Geochemistry: Exploration, Environment, Analysis. 5(4), 311-323.

SHEWHART, W A. 1931. Economic Control of Quality of Manufactured Product. Van Nostrad, Washington DC.

SINCLAIR, A J. 1976. Applications of probability graphs in mineral exploration. The Association of Exploration Geochemists. Spec. Vol. No. 4.

SINCLAIR, A J. 1983. Univariate Analysis. Chapter 3, pages 59-81. In: Statistics and Data Analysis in Geochemical Prospecting. Handbook of Exploration Geochemistry Volume 2. R J Howarth Editor. Elsevier Scientific Publishing, Amsterdam-Oxford_New York.

SNEDECOR, G W AND COCHRAN, W G. 1989. *Statistical Methods*. (Iowa State University Press.)

THOMPSON, M AND HOWARTH, R J. 1973. The rapid estimation and control of precision by duplicate determinations. Analyst, 98, 153-160.

THOMPSON, M AND HOWARTH, R J. 1978. A new approach to the estimation of analytical precision. Journal of Geochemical Exploration, 9, 23-30.