

18

Quality Control (QC)

The quality of chemical analyses is of utmost importance in environmental sciences. Chemical analyses are – as are any physical measurement(s) – confounded by uncertainty. Should the users of the analyses have to concern themselves with the quality of the data returned from the laboratory, or is that the laboratory's task? Today most laboratories are nationally accredited, i.e. they follow strict quality control procedures and frequently take part in “round-robin tests” (an analysis of the same samples is performed independently by several laboratories several times). Why should there still be any problems with analytical quality that applied geochemists and environmental scientists, the users of the chemical analyses, should be concerned with? In these times of accredited laboratories and certified analytical methods it is a temptation to take the analytical results and use them without any further concern. However, experience with large geochemical mapping projects during the last two decades has proven that external quality control of the results received from laboratories should, as ever, be an important concern.

For data analysis a good measure of the quality of the data is required; how reliable are they? Would the same results be obtained if the fieldwork was repeated and a second set of samples collected in the same general area as the first survey? Would the same analytical results be obtained if the same sample was re-analysed some weeks later in the same laboratory – or in a different laboratory? Can results of samples collected today be reliably compared with samples to be collected from the same area in ten years time?

To answer such questions, any project in applied geochemistry or environmental sciences should be carried out with its own built-in quality control procedures. It is not sufficient to rely on the accreditation of the laboratory and automatically accept the quality of the values received from a certified laboratory. A project's quality control procedure should include the following steps:

- collection and analysis of field duplicates (frequently, 5–10 per cent);
- randomisation of samples prior to analysis;
- insertion of international reference materials (sparsely, 1 per cent);
- insertion of project standard(s) (frequently, 10–20 per cent); and
- insertion of analytical duplicates of project samples (frequently, 10–20 per cent).

If it is desired to simultaneously quantitatively estimate both the field sampling and the analytical variability it is preferable to prepare the analytical duplicate from the field duplicate.

Although this chapter focuses predominantly on statistical methods for analytical quality control it must be mentioned that quality control of sampling and an adequate sampling protocol for the whole project is at least as important (e.g., Garrett, 1983b; Argyraki *et al.*, 1995; Ramsey, 1997, 2002, 2004a,b).

18.1 Randomised samples

Why should the samples be submitted in a random order to the laboratory? The reason is to spread any temporally systematic bias related to the measurement system(s) randomly over all the samples when they are returned to project order. Thus any time-dependent errors that may occur in the laboratory, such as a slow drift from lower to higher reporting levels, does not appear as a feature in the map.

When submitting randomised samples, the laboratory must be instructed to analyse the samples in the sequence of the submitted sample numbers – otherwise the laboratory may randomise the samples once more. Whether either randomised sample numbers are used during field sampling, or new randomised sample numbers are assigned to samples during sample preparation (easiest because standards and duplicates also need to be inserted so that it is not easy for a laboratory to detect the QC samples), is not important as long as the samples are randomised. For a large project it can be advantageous to wait until all samples are ready to be submitted as a single large batch. Submitting the samples as several batches can create serious problems when QC indicates that clear differences exist between the batches in terms of accuracy and/or precision (see below).

18.2 Trueness

Trueness is the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value for a reference material. In this connection the term “bias” is important. Bias is the difference between the expectation of the test results and an accepted reference value. Bias is the total systematic error. Systematic error is that component of error which, in the course of a number of test results for the same characteristic, remains constant or varies in a predictable way. Systematic error must be contrasted to random error. Random error is the error, which in the course of a number of test results for the same characteristic, varies in an unpredictable way and thus cannot be corrected for. Bias may consist of one or more systematic error components.

It is impossible to absolutely determine the “true” composition of a reference material; it can be approached, but there will always be some uncertainty. The objective is to minimise this uncertainty. The data for international reference materials, analysed by many different laboratories, preferably employing a variety of different methods, are compiled and estimates made of the most likely “true” value. The standard deviation of the repeated analyses provides an estimate of the uncertainty associated with the most likely “true” value.

An issue in selecting reference materials is that they should be as similar as possible to the samples collected from the field and under study. The requirement is to match the matrix, the chemistry and mineralogy, of the reference materials as closely as possible to the matrix of the field study samples. Soil standards should be used in soil studies, vegetable materials should be used for moss or other plant materials, rocks for rock analyses, etc. A soil reference

sample prepared from a calcareous soil should not be used for the analyses of soils with a low-carbonate siliceous matrix. Documentation of analytical “quality” for the project samples should involve results obtained from analysing one or more standard reference materials similar to the sample material, be it rock, soil, stream sediment, vegetation or water, analysed a number of times throughout the duration of the analytical project so that values for the MEAN and SD can be computed.

Results obtained when analysing an international reference material together with the project samples are usually summarised in a table showing the most important statistical measures, i.e. MEAN, SD, and Coefficient of Variation (CV, that is the same as Relative Standard Deviation, RSD) of the results and of the agreed certified value of the reference material. Table 18.1

Table 18.1 Results for the nine replicate analyses of international reference material CRM NIST 1547, Peach leaves, carried out while analysing the Kola moss samples; together with the MEAN and SD for the certified values. Results of CRM values in square brackets are not certified. All units mg/kg

Element	Certified value	Kola Project		
	MEAN \pm SD	MEAN	SD	CV %
Al	249 \pm 8	213	8	3.9
As	0.06 \pm 0.018	0.098	0.007	6.7
B	29 \pm 2	27.4	1.9	7.1
Ba	124 \pm 4	126	3.2	2.6
Ca	15600 \pm 200	16210	20	1.3
Cd	0.026 \pm 0.003	0.021	0.003	13.2
Co	[0.07]	0.068	0.004	6.5
Cr	[1.0]	0.88	0.11	12.4
Cu	3.7 \pm 0.4	3.50	0.20	5.7
Fe	218 \pm 14	209	3.4	1.6
Hg	0.031 \pm 0.007	0.035	0.012	33.2
K	24300 \pm 300	24967	524	2.1
La	[9]	9.6	0.35	3.6
Mg	4320 \pm 8	4318	100	2.3
Mn	98 \pm 3	97.9	1.6	1.7
Mo	0.06 \pm 0.008	0.050	0.008	16.5
Na	24 \pm 2	21.8	4.9	22.4
Ni	0.69 \pm 0.09	0.602	0.085	14.1
P	1370 \pm 7	1451	23	1.6
Pb	0.87 \pm 0.03	0.86	0.04	4.6
Rb	19.7 \pm 1.2	19.1	0.5	2.6
S	[2000]	1630	23	1.4
Sb	[0.02]	0.023	0.008	34.7
Si	—	263	19	7.1
Sr	53 \pm 4	57.9	1.7	2.9
Th	[0.05]	0.050	0.008	15.7
Tl	—	0.020	0.0022	11.0
U	[0.015]	0.011	0.0026	24.7
V	0.37 \pm 0.03	0.35	0.02	6.0
Y	[0.2]	3.00	0.09	2.9
Zn	17.9 \pm 0.4	18.1	0.5	2.7

contains the results for the control reference material (CRM) NIST 1547, Peach leaves, analysed nine times together with the Kola moss samples. Such a table can be used to convey information on the trueness of the received analytical results of a project. Other scientists, using the same reference material can now directly judge the comparability of the project data to their data. Furthermore, it will give a first impression as to the overall analytical quality achieved.

Due to the high price of international reference materials, they will usually not be used to monitor for accuracy (see below).

18.3 Accuracy

Accuracy is essentially the absence of bias. Note that analytical results can be highly accurate without being “true”. International reference materials are expensive. In addition they are usually easy for a laboratory to identify when inserted among the project samples. Thus to detect any drift or other changes in analytical results over time, one or more project standards, control reference materials (CRMs), are required that are quite frequently inserted among the field survey samples – probably at least at an overall rate of 1 in 20. These samples should not be different from all other project samples, either in looks or in specific gravity (density), or in the amount submitted for analysis. A common procedure is to collect one or more large samples from the project area, prepare them as for all the field samples of that material, and split them down into the appropriate number of aliquots that can be inserted among the samples prior to submission for laboratory analysis. These samples are then used to monitor for consistency, lack of drift across the duration of the project. The number of CRMs used depends upon the geochemical complexity of the survey area. On one hand the CRMs should reflect the different mineralogical matrices, but there should not be so many CRMs that insufficient analyses are made to be able to monitor the analytical results adequately. Rarely should more than three CRMs be used for a sample material, and in some cases, e.g., mosses with a consistent matrix, only one would be needed.

Immediately after the analytical results are received from the laboratory, the control reference material results should be retrieved from the file. A table for these results can be prepared providing estimates of the repeatability of the measurements (see Section 18.4) at the concentration of the standard sample for each element (Table 18.2). Such a table can be sorted alphabetically for the elements (Table 18.2, left) or in increasing order of the CV to get a rapid overview of the overall quality (Table 18.2, right). The sorted table (Table 18.2, right) shows at one glance that the elements analysed by XRF (recorded as oxides) show an exceptional high repeatability and highlights the elements with possible analytical quality problems at the level of the CRM (poor repeatability could be due to proximity to the DL).

In addition, the data for each variable are plotted against the sample number. The sample number provides an estimate of “time of analysis” (again, the laboratory must be instructed to analyse the samples in the exact sequence of the submitted sample numbers). In such diagrams, also called “x-Charts”, any trends (e.g., due to instrumental drift) or gross deviations (e.g., due to sample mix-ups) from the average analytical result for the project CRM become immediately visible. It is common practice for laboratories to insert their own control materials, as a result errors due to an instrument malfunction are usually “caught” prior to the data being returned to the project. However, if a problem has remained undetected by the laboratory, x-Charts will likely detect them.

Table 18.2 Average repeatability for selected elements calculated from analyses of the CRM used for the Kola Project C-horizon data ($n = 52$)

Element	Unit	DL	MEAN	SD	CV %	Element	Unit	DL	MEAN	SD	CV %
Ag	mg/kg	0.001	0.0075	0.0014	18	Al ₂ O ₃	wt%	0.05	10	0.200	2.0
Al	mg/kg	10	4015	254	6.3	CaO	wt%	0.007	2.2	0.057	2.6
Al ₂ O ₃	wt%	0.05	10	0.200	2.0	Fe ₂ O ₃	wt%	0.02	3.3	0.098	3.0
As	mg/kg	0.1	0.22	0.078	35	Ni	mg/kg	1	16	0.775	4.9
Ba	mg/kg	0.5	35	2.42	7.0	Fe_INAA	mg/kg	100	24898	1278	5.1
Bi	mg/kg	0.005	0.013	0.0031	24	La	mg/kg	0.5	8.0	0.478	6.0
CaO	wt%	0.007	2.2	0.057	2.6	La_INAA	mg/kg	1	15	0.908	6.1
Cd	mg/kg	0.001	0.010	0.0016	15	Al	mg/kg	10	4015	254	6.3
Co	mg/kg	0.2	5.5	0.666	12	Fe	mg/kg	10	9528	628	6.6
Cr	mg/kg	0.5	42	3.51	8.3	Cu	mg/kg	0.5	12.6	0.84	6.7
Cu	mg/kg	0.5	12.6	0.84	6.7	Ba	mg/kg	0.5	35	2.42	7.0
Fe	mg/kg	10	9528	628	6.6	V	mg/kg	0.5	21	1.54	7.5
Fe_INAA	mg/kg	100	24898	1278	5.1	Zn	mg/kg	0.5	9.1	0.689	7.5
Fe ₂ O ₃	wt%	0.02	3.3	0.098	3.0	Cr	mg/kg	0.5	42	3.51	8.3
La	mg/kg	0.5	8.0	0.478	6.0	Th_INAA	mg/kg	0.2	3.0	0.305	10
La_INAA	mg/kg	1	15	0.908	6.1	Sc_INAA	mg/kg	0.1	9.5	0.972	10
Ni	mg/kg	1	16	0.775	4.9	Co	mg/kg	0.2	5.5	0.666	12
Pb	mg/kg	0.2	0.78	0.096	12	Pb	mg/kg	0.2	0.78	0.096	12
S	mg/kg	5	13	8.07	63	Cd	mg/kg	0.001	0.010	0.0016	15
Sc_INAA	mg/kg	0.1	9.5	0.972	10	Ag	mg/kg	0.001	0.0075	0.0014	18
Se	mg/kg	0.01	0.034	0.030	89	Bi	mg/kg	0.005	0.013	0.0031	24
Te	mg/kg	0.003	0.0048	0.0034	72	As	mg/kg	0.1	0.22	0.078	35
Th_INAA	mg/kg	0.2	3.0	0.305	10	S	mg/kg	5	13	8.07	63
V	mg/kg	0.5	21	1.54	7.5	Te	mg/kg	0.003	0.0048	0.0034	72
Zn	mg/kg	0.5	9.1	0.689	7.5	Se	mg/kg	0.01	0.034	0.030	89

Figure 18.1 shows two such x-Charts constructed with the data obtained from the Kola Project soil CRM, a large soil sample collected within the project area in northern Norway. In such a plot it is desirable to have an indication of whether a sample falls within the range of “normal” variation around a central value, or is an outlier. This can be done by plotting lines for the MEAN, and multiples of the standard deviation (Figure 18.1). This graphical inspection can be used, for example, to decide whether the block of samples containing the CRM result is accepted from the laboratory or rejected and the block needs to be re-analysed. x-Charts will often show time trends or unusual breaks in the data, related to changes in the laboratory – examples displaying pure random variation (the ideal case) are rare. In the examples used in Figure 18.1 the x-Chart for Al₂O₃ indicates the existence of one outlier (ID 510) – this is possibly an indication that while loading the automatic sample changer of the XRF-machine, samples were inadvertently exchanged. Note also the tendency for higher results at IDs > 600. The plot for Sm shows a clear trend towards lower values over time (Figure 18.1). It appears that only for the very last analytical batches (indicated by the last two CRM results) the laboratory became aware of this trend and corrected the calibration. This plot illustrates why samples should be randomised: a time trend like this might otherwise lead to unusual patterns on a geochemical map. Randomisation ensures this drift is distributed as a random component across the map. The plot also

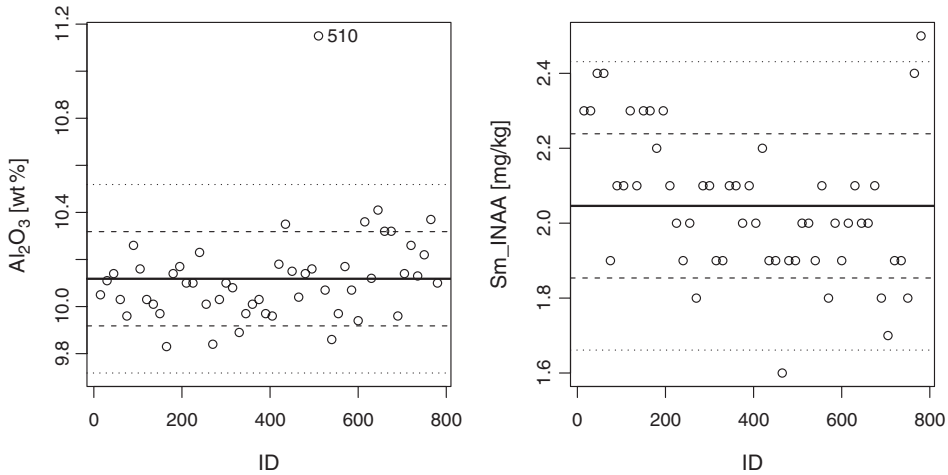


Figure 18.1 x-Chart showing the analytical results of the Kola Project CRM for Al_2O_3 as determined by XRF and for Sm determined by INAA. Dotted lines shown in the plot are $\text{MEAN} \pm 2 \cdot \text{SD}$, dashed lines are drawn at $\text{MEAN} \pm 1 \cdot \text{SD}$, and the solid line at the MEAN

indicates a too-severe rounding, discretisation, of the Sm results (discrete “lines” of results) by the laboratory.

18.4 Precision

Precision is the closeness of agreement between independent test results obtained under stipulated conditions. It depends only on the distribution of random errors and does not relate to the true value. Precision is usually quantitatively expressed in terms of imprecision and estimated through the standard deviation of the test results: a low standard deviation will indicate a high precision. The precision is usually adjusted for the mean and expressed as the CV, or equivalent RSD, both quoted as a percentage (e.g., Massart *et al.*, 1988). When referring to precision, the conditions of measurement must be specified. There is an important difference between repeatability conditions and reproducibility conditions. Repeatability conditions refer to situations where independent test results are obtained using the same method on identical test items (samples) in the same laboratory by the same operator using the same equipment within short intervals of time. Most references to precision found in literature will refer to repeatability conditions. Reproducibility conditions refer to situations where test results are obtained with the same method on identical test items (samples) in different laboratories with different operators using different equipment. It is also important to note that repeatability conditions involve repeated execution of the entire method from the point at which the material for analysis reaches the laboratory, and not just repeat instrumental determinations on prepared extracts. The latter give impressive estimates of precision – but have no relevance to the precision achieved when field samples are analysed in the laboratory as it does not take the natural inhomogeneity of the sample material, an important source of variability, into account.

18.4.1 Analytical duplicates

Precision is routinely estimated via the insertion of duplicates of real project samples, usually at a rate of 1 in 20 (or 1 in 10). It is, for example, straightforward to always reserve position “20” (20, 40, 60, ...) for a duplicate of one of the preceding 18 real samples (plus one project standard). Again the results for all duplicate pairs are retrieved once the laboratory delivers the results. For each pair the squared difference is calculated. The sum of these values divided by the number of samples is a measure of variability. To obtain the standard deviation the square root of this variability measure is taken. The resulting measure of precision as shown in Table 18.3 corresponds to a CV value, because the standard deviation is divided by the overall mean of the samples.

Table 18.3 shows the results of the analytical duplicates as received for the Kola Project C-horizon samples. Again it can be advantageous to show both a table sorted according to the alphabetical sequence of the elements (Table 18.3, left) and the same table sorted according to precision (Table 18.3, right). On average, precision is quite acceptable for most elements. Even the problematic elements in the CRM table (Table 18.2), Se and Te, show a better than expected precision in Table 18.3.

Table 18.3 Precision as calculated for selected elements from the 52 analytical duplicates from the analyses of Kola Project C-horizon soils

Element	Unit	DL	Precision %	Element	Unit	DL	Precision %
Ag	mg/kg	0.001	20	Al ₂ O ₃	wt%	0.05	4
Al	mg/kg	10	5.5	Ni	mg/kg	1	5.2
Al ₂ O ₃	wt%	0.05	4.0	Ba	mg/kg	0.5	5.3
As	mg/kg	0.1	28	Al	mg/kg	10	5.5
Ba	mg/kg	0.5	5.3	Zn	mg/kg	0.5	5.6
Bi	mg/kg	0.005	18	Cu	mg/kg	0.5	6.1
CaO	wt%	0.007	16	Co	mg/kg	0.2	6.7
Cd	mg/kg	0.001	11	Cr	mg/kg	0.5	7.2
Co	mg/kg	0.2	6.7	Fe_INAA	mg/kg	100	7.3
Cr	mg/kg	0.5	7.2	Sc_INAA	mg/kg	0.1	7.5
Cu	mg/kg	0.5	6.1	La_INAA	mg/kg	1	8.3
Fe	mg/kg	10	11	S	mg/kg	5	9.3
Fe_INAA	mg/kg	100	7.3	V	mg/kg	0.5	10
Fe ₂ O ₃	wt%	0.02	15	Cd	mg/kg	0.001	11
La	mg/kg	0.5	18	Fe	mg/kg	10	11
La_INAA	mg/kg	1	8.3	Fe ₂ O ₃	wt%	0.02	15
Ni	mg/kg	1	5.2	CaO	wt%	0.007	16
Pb	mg/kg	0.2	25	Bi	mg/kg	0.005	18
S	mg/kg	5	9.3	La	mg/kg	0.5	18
Sc_INAA	mg/kg	0.1	7.5	Se	mg/kg	0.01	18
Se	mg/kg	0.01	18	Th_INAA	mg/kg	0.2	18
Te	mg/kg	0.003	33	Ag	mg/kg	0.001	20
Th_INAA	mg/kg	0.2	18	Pb	mg/kg	0.2	25
V	mg/kg	0.5	10	As	mg/kg	0.1	28
Zn	mg/kg	0.5	5.6	Te	mg/kg	0.003	33

Again it is advantageous to represent these results in a graphical form. “Thompson and Howarth” plots (Thompson and Howarth, 1978) are a frequently used graphic for this purpose (Figure 18.2). Here the absolute difference between the two analyses, $|D1 - D2|$, is plotted against the mean of the duplicate results $(D1 + D2)/2$, and the overall analytical performance can be grasped at once. These graphs can also be used to estimate the practical detection limit (PDL – see discussion below) for an analytical procedure (the point where precision becomes worse than ± 100 per cent) via regression analysis. It is possible to draw lines for any certain predefined precision (e.g., ± 10 per cent) into these diagrams (in Figure 18.2 at ± 10 per cent and ± 20 per cent). It is then directly recognisable if any, and if so how many, of the samples plot above the line(s). This can support the decision as to whether an analytical batch from the laboratory is accepted or rejected. Figure 18.2 shows that precision is excellent for potassium (K_2O) as analysed by XRF. Four samples that fall above the 20 per cent line are an exception. This is again a strong indication that a number of samples were exchanged during the analytical process. For Zn as analysed by INAA, precision is not so good. The fact that a number of duplicate pairs plot along a straight line to the left in the diagram is caused by a high detection limit of 50 mg/kg. When one of the duplicate samples returns a value above the detection limit and the other a result below the detection limit (set to 25 mg/kg) points on a line result.

In reality precision depends on concentration and follows a curve from poor precision near the detection limit to high precision in the optimal working range of the analytical technique back to poorer precision at the upper limit of concentrations that can be measured with this technique (for an example, see Fauth *et al.*, 1985). Thus there exists both a “lower” limit and an “upper” limit of detection, and fitting just one regression line into these plots is a simplification that has its limits when the duplicates cover a very wide concentration range.

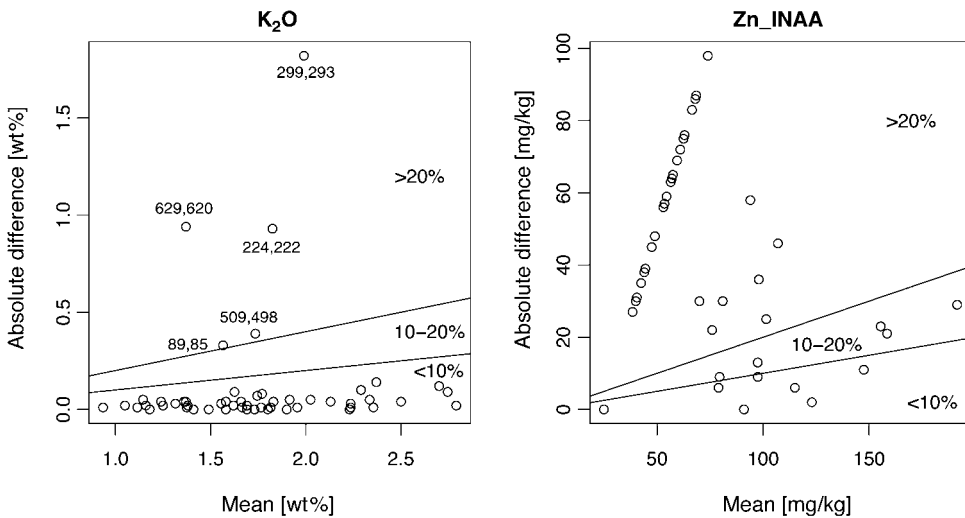


Figure 18.2 Thompson and Howarth plots for analytical duplicate results of K_2O (XRF) and Zn (INAA) determined in Kola Project C-horizon soils. Number pairs annotating samples in the left plot identify analytical duplicates with unusually poor precision for follow-up

18.4.2 Field duplicates

In a comprehensive quality control program field duplicates will have been collected at a number of randomly selected sites (usually 5–10 per cent). These samples are used to estimate the variation introduced by sampling and to answer the question of whether more or less the same analytical results would be obtained if undertaking the survey a second time at approximately the same sample sites. An estimate of the field variability is especially important in monitoring programs, i.e. when the sampling exercise is to be repeated after a number of years to detect any changes over time.

The precision of field duplicates can be estimated in the same way as was the precision of analytical duplicates (Table 18.4), and even Thompson and Howarth plots could be drawn for field duplicates (Figure 18.3). The precision determined from field duplicates includes variability due to both sampling and analysis. Table 18.4 and the Thompson and Howarth plots for the field duplicates (Figure 18.3) can be directly compared to the same results from the analytical duplicates to get an initial impression of the relative magnitude of the sampling error in relation to the analytical error. Again, sorting Table 18.4 according to decreasing precision (right half) can aid interpretation and help to detect any problematic elements.

Table 18.4 Combined sampling and analytical precision calculated for selected elements from the 49 field duplicates of Kola Project C-horizon soils

Element	Unit	DL	Precision %	Element	Unit	DL	Precision %
Ag	mg/kg	0.001	57	Al ₂ O ₃	wt%	0.05	4.2
Al	mg/kg	10	16	Fe_INAA	mg/kg	100	8.1
Al ₂ O ₃	wt%	0.05	4.2	Fe	mg/kg	10	9.3
As	mg/kg	0.1	37	V	mg/kg	0.5	11
Ba	mg/kg	0.5	21	CaO	wt%	0.007	12
Bi	mg/kg	0.005	26	Fe ₂ O ₃	wt%	0.02	13
CaO	wt%	0.007	12	Co	mg/kg	0.2	14
Cd	mg/kg	0.001	71	La_INAA	mg/kg	1	14
Co	mg/kg	0.2	14	Sc_INAA	mg/kg	0.1	14
Cr	mg/kg	0.5	23	Zn	mg/kg	0.5	14
Cu	mg/kg	0.5	30	Al	mg/kg	10	16
Fe	mg/kg	10	9.3	La	mg/kg	0.5	16
Fe_INAA	mg/kg	100	8.1	Th_INAA	mg/kg	0.2	17
Fe ₂ O ₃	wt%	0.02	13	Ba	mg/kg	0.5	21
La	mg/kg	0.5	16	Cr	mg/kg	0.5	23
La_INAA	mg/kg	1	14	Bi	mg/kg	0.005	26
Ni	mg/kg	1	33	S	mg/kg	5	27
Pb	mg/kg	0.2	44	Cu	mg/kg	0.5	30
S	mg/kg	5	27	Ni	mg/kg	1	33
Sc_INAA	mg/kg	0.1	14	As	mg/kg	0.1	37
Se	mg/kg	0.01	39	Se	mg/kg	0.01	39
Te	mg/kg	0.003	170	Pb	mg/kg	0.2	44
Th_INAA	mg/kg	0.2	17	Ag	mg/kg	0.001	57
V	mg/kg	0.5	11	Cd	mg/kg	0.001	71
Zn	mg/kg	0.5	14	Te	mg/kg	0.003	170

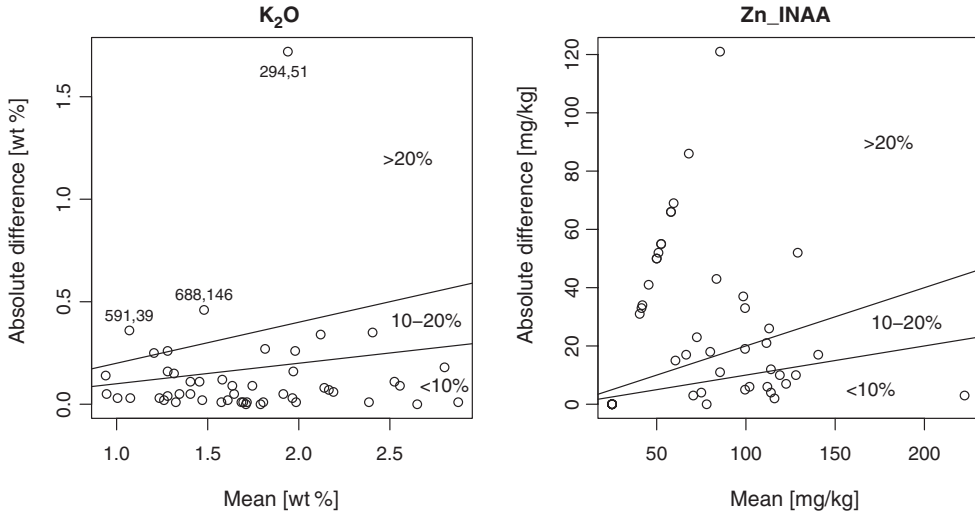


Figure 18.3 “Thompson and Howarth” plots for field duplicate results of K₂O (XRF) and Zn (INAA) determined in Kola Project C-horizon soils. Number pairs annotating samples in the left plot identify analytical duplicates with unusually poor precision for follow-up

18.5 Analysis of variance (ANOVA)

In a more formalised approach, results of all samples, field and analytical duplicates can be used to carry out an unbalanced Analysis of Variance (ANOVA) (see, e.g., Scheffé, 1999). By “unbalanced” is meant that unequal numbers of analyses occur at each level of the design. A “balanced” design for studying field and analytical variability would require that each sample collected at a field duplicate site would be split and analysed twice. In an unbalanced design only one of the field duplicate pairs is split and analysed twice. The same information can be extracted from both designs, however, the unbalanced design makes a more efficient use of resources. The results of the ANOVA include estimates of the proportion of the total variability due to each of sampling and analysis. Whether or not the analytical variability is significantly smaller than the sampling variability can be determined by a formal statistical test. However, this formal analysis can only be undertaken when the analytical duplicate is prepared from the field duplicate. This was not done for the Kola Project. In its place two independent sets of duplicates were obtained: one to estimate the combined sampling and analytical variability from field duplicates; and the second to estimate the analytical variability from a set of duplicates prepared from the field samples. The “at-site” sampling variability should be considerably larger than the analytical variability. However, this depends on how the field duplicates were collected. Were they just a split of the field sample (not a “true” field duplicate!); were they, for example, taken from opposite walls of the same soil pit; or do they represent a “true” duplicate, i.e. collected at a second site in the vicinity of the first site? For the Kola Project all field duplicates were taken at a second site about 100 m removed from the first site. This is a “worst case” scenario that will provide a very realistic picture of at-site variability.

An ANOVA of the field duplicate pairs decomposes the variability into two components: (1) the variability between the sites where duplicates were collected; and (2) the variability at the field duplicate site. In addition, a formal F-test can be undertaken to determine if the variability

Table 18.5 ANOVA table for the 49 field duplicate pairs for Cu in Kola Project C-horizon soils

Source	DF	MSS	F	<i>p</i> -value	Variance	Percentage of variation
Between sites	48	0.16536	16.5	<0.0001	0.077673	88.6
Within sites	49	0.01001			0.010009	11.4

at the sites is significantly different from the variability between the sites. This test uses what is called a “random effects model” which is somewhat different from the “fixed effects” one-way ANOVA model discussed in Section 9.7.1. When only two sources of variability are being considered the two methods, “fixed effects” and “random effects”, are computationally identical except for the calculation of the proportions of the variability related to the two sources. If the variability between the sites is not significantly greater than the variability within the sites, preparing maps of the data may be misleading. Using the logarithmically transformed Cu data for C-horizon soils, the ANOVA table presented in Table 18.5 may be generated. A logarithmic transformation was used in order to meet the requirement of homogeneity of variance and to approach a normal distribution (see discussion in Section 16.1.1). Additionally an inspection of the data shows that the Cu data span more than one order of magnitude, a useful guide to whether a transformation is necessary.

The ANOVA (Table 18.5) indicates that some 89 per cent of the total variability is due to variations between the regional sites where field duplicates were collected, and only 11 per cent was due to variability at the field duplicates sites. The F-test and the associated *p*-value confirm that this partition is highly significant and that maps can be prepared with confidence.

The same approach can be applied to analytical duplicate pairs, see Table 18.6.

The between sites variability now comprises the variability between the sites chosen for analytical replication across the survey area, and the only remaining variability is due to analysis, i.e. weighing out and analysing a second aliquot of the prepared sample. As might be expected, virtually all the variability now lies between the regional sampling sites and only 0.4 per cent is due to analysis. The F-test and *p*-value demonstrate that this partition is highly significant.

As a formal staggered unbalanced design (Garrett, 1983b) was not employed in the Kola Project, an approximation has to be used in order to partition the variability between the three sources, regional, at sites, and analytical. This can be done because variances are additive. The first task is to determine if the regional variability estimates, between sites, as determined from the data used in both ANOVAs, are sufficiently similar to proceed. The ratio of the “between” variances, both measures of the variability across the survey area, from the two ANOVAs is calculated; it is 1.45, which corresponds to a *p*-value of 0.1. This is above the five per cent significance level, and it cannot be accepted that the two variances are equal.

Table 18.6 ANOVA table for the 52 analytical duplicate pairs for Cu in Kola Project C-horizon soils

Source	DF	MSS	F	<i>p</i> -value	Variance	Percentage of variation
Between sites	51	0.22600	454	<0.0001	0.11274	99.6
Between analyses	52	0.00051			0.00050838	0.4

However, some estimate of the relative variability between the sources is still desirable. Therefore, the field duplicate “between sites” variance was selected, as sites selected for duplicate sampling are likely more evenly distributed across the project area than those selected to monitor analytical variability. As variances are additive, the analytical, “between analyses”, variance determined from the analytical duplicate pairs can be subtracted from the “within sites” variance to estimate that part of the “within sites” variance that is due to the “at sample sites” variance alone. These three variances may now be expressed in percentage form. Thus, for Cu the variability is partitioned: 88.6 per cent to regional, 10.8 per cent to variability at sampling sites, and 0.6 per cent to analytical variability. This is a satisfactory outcome for a regional survey.

The above approximation will sometimes lead to negative estimates for the “at sampling sites” variances; these are impossible. However, they are the outcome of not employing a proper unbalanced sampling design for estimating the variance components. By convention, negative estimates are set to zero, and the computation completed.

Table 18.7 presents the results of the above partitioning of the variability into variance components (expressed as percentages) for the major element oxides determined by XRF (left) and selected trace elements (right), and the p -value is for the F-test with the ratio of the two “between” variances determined from the field duplicates and analytical duplicates.

In most cases the p -values in Table 18.7 are greater than 0.05, indicating an inequality of the “between” variances for the field and analytical duplicates, raising some doubt over the

Table 18.7 Distributed percentage variabilities for the major element oxides (determined by XRF) and selected trace elements for Kola Project C-horizon soils. The p -value is for the F-test to determine if the variances at the “between” level are equal for the field and analytical duplicates. With the exception of pH all variables were log-transformed prior to the calculation

Element	Regional	Site	Analytical	p -value	Element	Regional	Site	Analytical	p -value
Al ₂ O ₃	88.65	1.05	10.30	0.88	Ag	63.14	17.29	19.57	0.08
CaO	86.17	0	13.83	1	As	90.50	3.52	5.98	0.03
Fe ₂ O ₃	80.11	0	19.89	0.05	Be	95.30	3.51	1.19	0.45
K ₂ O	80.32	0	19.68	0.8	Bi	79.65	7.56	12.80	0.11
MgO	86.86	2.40	10.75	0.54	Cd	80.25	17.47	2.28	0.48
MnO	66.73	0	33.27	0.15	Co	92.76	5.47	1.77	0.15
Na ₂ O	87.62	0	12.38	0.11	Cr	94.68	4.15	1.18	0.42
P ₂ O ₅	91.37	4.20	4.43	0.94	Cr_INAA	90.44	8.12	1.44	0.82
SiO ₂	82.56	0	17.44	0.88	Cu	88.59	10.83	0.58	0.10
TiO ₂	82.53	0	17.47	0.04	Hg	49.99	10.20	39.81	0.10
					Mo	72.32	13.88	13.80	0.25
					Ni	92.02	7.08	0.90	0.25
					Pb	80.01	11.50	8.48	0.87
					pH	28.19	11.23	60.57	0.01
					S	72.73	23.96	3.31	0.07
					Sb	65.32	25.09	9.59	0.09
					Se	67.40	19.11	13.48	0.06
					Sr	95.91	1.52	2.57	0.86
					Th	82.05	9.92	8.03	0.30
					V	95.34	0.68	3.98	0.40
					Zn	93.26	5.59	1.15	0.26

validity of this *ad hoc* decomposition of the variances. The critical fact is that all the field duplicate ANOVAs, except Hg and pH, indicated that the combined sampling and analytical variability is significantly smaller than the “between sites” regional variability. That is the most important conclusion and supports the validity of the maps. As this critical test was passed, it was considered informative to proceed with the *ad hoc* estimation of variance components so as to be able to place the relative levels of “at site” and “analytical” variability in context with the “overall regional” variability.

The table (Table 18.7, left) indicates that the “at sampling sites” variability for the major elements is very small or zero relative to the analytical variability. This indicates that the soils at the 100 m scale are very homogeneous with respect to major elements. For all oxides, except Mn, in excess of 80 per cent of the variability is at a regional scale. The partitioning for the trace elements (Table 18.7, right) presents a more diverse story. Elements such as Ag, Hg, Sb and the pH all have high local variabilities, i.e. the “between sites” regional variability is less than 67 per cent. This reflects the difficulties in determining these elements at low levels and the problems of measuring pH in soil samples. Bi, Mo, S and Se all have “between sites” variabilities in the range of 67–80 per cent, indicating that maps of these elements are likely reliable. For all of the remaining elements in excess of 80 per cent of their variability is at the regional scale indicating that they may be mapped with confidence.

18.6 Using maps to assess data quality

One additional quality criterion will be the appearance of the map when the analytical results are mapped (see Chapter 5). Figure 18.4 shows two kriged surface maps for Ca: left map, Ca in the O-horizon; and right map, Ca in the C-horizon. The question to be asked is, do the maps contain any clear regional features or could they as easily represent random variability

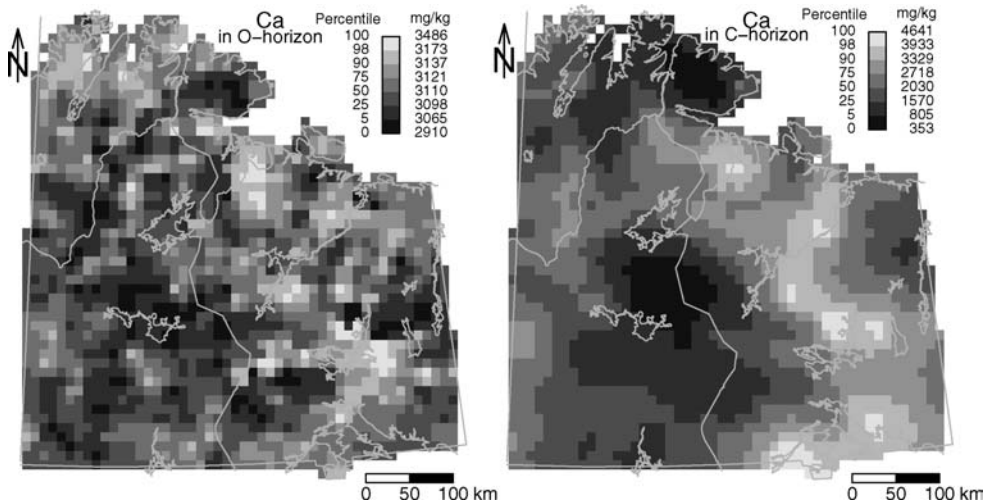


Figure 18.4 Kriged surface maps for Ca in the O-horizon (left) and Ca in the C-horizon (right) of Kola Project soils

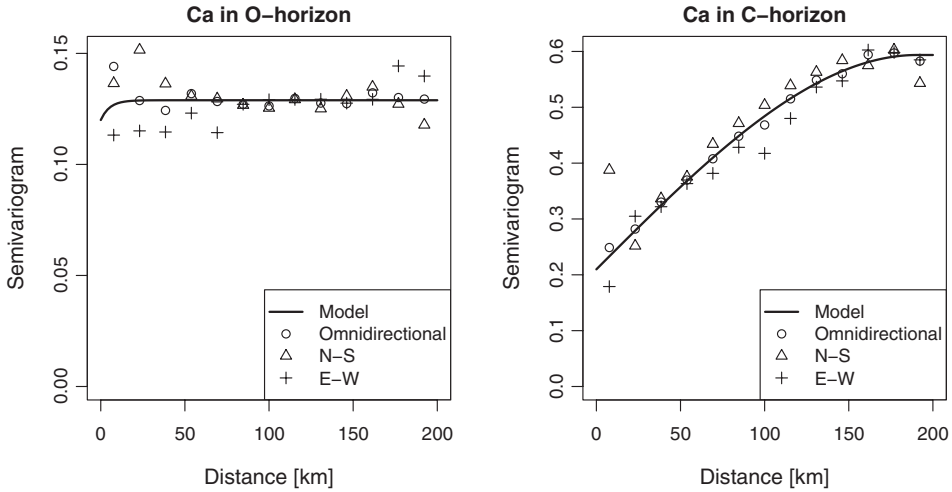


Figure 18.5 Semivariograms for Ca in the O-horizon (left) and C-horizon (right) Kola Project soils

due to analysis, sampling variability, and a lack of any significant regional trends? The map of Ca in the O-horizon is very noisy without any clear spatial structure, while the map of Ca in the C-horizon shows clear spatial structures (Figure 18.4, right). In the first case it could be assumed that the sample density was too low or (and) the sampling and analytical variability was too high relative to the regional variability to obtain a useful map. It is also possible that no clear regional structure for Ca in the O-horizon exists because its regional distribution is dominated by local small-scale effects.

Information about data quality, or better suitability for mapping, can also be directly derived from the semivariogram (see Section 5.7.2), if kriging was used as the interpolation method. Figure 18.5 shows the semivariograms used to construct the above maps. The semivariogram for Ca in the O-horizon (Figure 18.5, left) is a typical example of the nugget effect, i.e. spatial independence of the samples; the nugget effect is almost 100 per cent of the total variance. The semivariogram for the C-horizon data (Figure 18.5, right) shows a clear spatial dependency for a range of about 200 km, here the nugget effect is approximately 30 per cent of the total variance.

18.7 Variables analysed by two different analytical techniques

Due to the use of multi-element analytical packages, data for several elements may be determined by more than one technique. This provides an additional means for quality control if the analytical techniques are measuring the elements in the same mineralogical fractions. For example, in the Kola data set (C-horizon) several elements were analysed by both X-Ray Fluorescence (XRF) and Instrumental Neutron Activation Analysis (INAA). Both techniques should yield “total” element concentrations and the results should be directly comparable. At the same time the techniques require different sample preparation procedures and are based on completely different physical principles. Thus they provide an ideal opportunity for quality control investigations, covering the whole range of the data. Some elements may have been

analysed by different techniques, giving some combination of “partial” and/or “total” results. In these instances a comparison can usually not be made (other than in a gross sense) for quality control purposes. However, the data can be used to gain a better understanding of modes of transport for trace elements and the forms in which they are sequestered in the sample materials.

Iron (Fe) is an example of an element that was determined by three different techniques: XRF, INAA and Inductively Coupled Plasma Atomic Emission Spectrometry (ICP-AES) following a partial aqua-regia extraction. In the environmental literature an aqua-regia extraction is often referred to as providing “total” element concentrations because it is a strong acid attack.

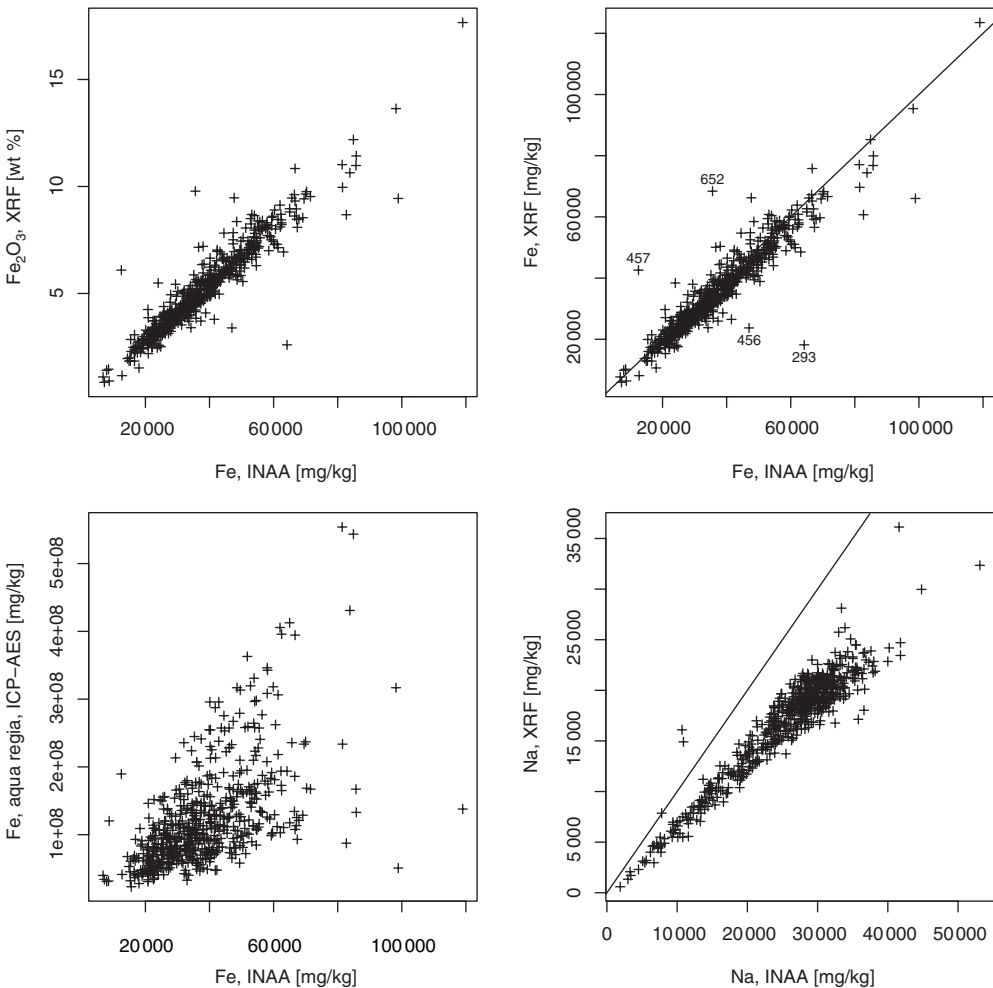


Figure 18.6 Fe in the Kola Project C-horizon soils as determined by three different techniques: XRF expressed as wt% Fe₂O₃ (upper left) and mg/kg Fe (upper right), INAA and in an aqua-regia extraction with ICP-AES finish (lower left); and Na as determined by XRF and INAA in Kola Project C-horizon soils (lower right)

However, elements bound in the lattices of many silicate and oxide minerals are not freed by an aqua-regia extraction and the results obtained are thus often far from “total”.

When plotting the results received from the laboratory (XRF results of major elements are routinely reported in wt% of the oxide – for Fe this is usually Fe_2O_3) a generally close relationship between the XRF and the INAA results is visible (Figure 18.6) – with some samples deviating from this general trend. To obtain a better direct comparison, the XRF results should be re-expressed in mg/kg Fe. The conversion factor from wt% Fe_2O_3 to mg/kg Fe is 6994 (conversion factors for all elements are, for example, given in Reimann and de Caritat, 1998b). Following the conversion XRF and INAA results can be directly compared and a 1:1 line can be drawn in the diagram (Figure 18.6, upper right). Pairs of samples that deviate from the general trend (like samples 456 and 457 or 652 and 293) point to the possibility that in the batches containing these samples, there was a sample mix-up, exchange, that should be followed up (i.e. inspect the duplicate and project CRM data). For Na it appears that the INAA results are always higher than the XRF results. The fanning out of the samples towards high concentrations indicates a decreasing precision at higher concentrations. Again samples deviating from the overall trend need to be identified and followed up.

These graphical appreciations of the interrelationships between the determination methods may be formalised through statistical tests. Both the paired t-test and the Wilcoxon test (see Sections 9.5.1 and 9.5.2) may be used to test if the differences between the two different methods of measurement on a suite of individuals are significant.

18.8 Working with censored data – a practical example

Problems related to an inappropriately chosen detection limit for some of the analysed variables have been described in Chapter 2. Here a practical example is chosen to demonstrate the importance of carefully selecting the detection limits for a project’s success or failure.

Gold (Au) results from the Kola Project are a typical example where more than 50 per cent of all data as originally received from the laboratory were “<DL” (Figure 18.7, upper left). When the project was originally planned, Au in the C-horizon soils was only determined by INAA. This technique provides a detection limit of 0.002 mg/kg (2 ppb) that was considered quite acceptable at the time (1995). It had to be expected that a substantial number of the samples would return “<DL” values. When the data were compiled, the CP-plot showed that almost 75 per cent of the data were below the detection limit (Figure 18.7, lower left). A variable with such a substantial number of values “<DL” has to be treated with care during data analysis. In the Kola Project the authors later became interested in studying the regional distribution of Au and the Platinum Group Elements (PGE) palladium (Pd) and platinum (Pt). Thus a special technique to determine these elements with very low detection limits was developed (Niskavaara *et al.*, 2004). Results for the C-horizon soils are shown in Figure 18.7 (upper right). The detection limit achieved was 0.0001 mg/kg (0.1 ppb), a factor of 20 improvement, with only two samples below the detection limit. Thus the complete data distribution can be observed (Figure 18.7, upper right).

When the original data are plotted versus the new Au data set (Figure 18.7, lower right) the major problem caused by the censored data in the original data set can be observed. Even for that part of the data where real values are available the expected 1:1 relationship is not present. This is in part due to poor analytical data quality close to the detection limit (original Au data), and in part caused by high natural sample variability due to the presence or absence of high

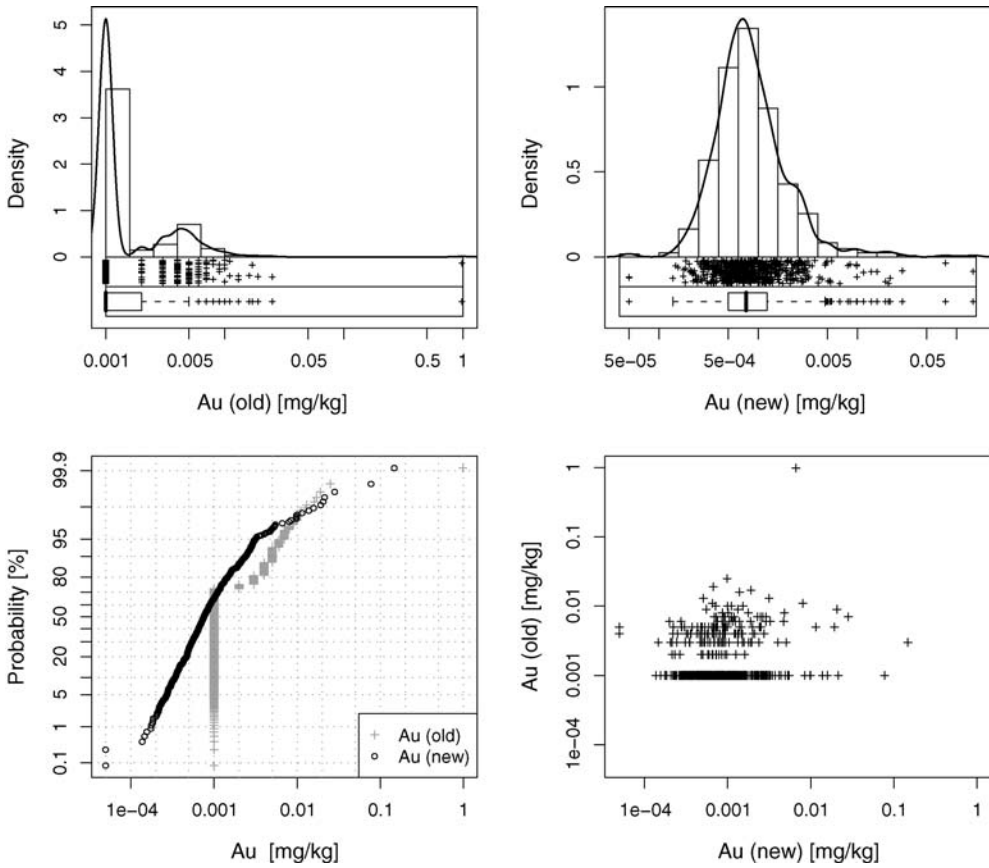


Figure 18.7 Combination plots of histogram, density trace, one-dimensional scatterplot and Tukey boxplot for Au old (upper left) and Au new (upper right) in the Kola Project C-horizon soils. Combined CP-plot for Au old and Au new (lower left), and scatterplot for Au new versus Au old (lower right)

specific gravity micro-nuggets in the aliquots weighed out for analysis (Clifton *et al.*, 1969; Stanley, 1998).

The problems caused by censored data become even more obvious in a scatterplot. When plotting Cu versus Au almost all the structure visible in the diagram will be hidden in that part of the data that is below the detection limit of the original analyses (Figure 18.8, left). The original data do not reflect the relation between Au and Cu that is so apparent in the new Au data (Figure 18.8, right). A clear message from this plot is that variables with a high proportion of censored data should not be used for multivariate data analysis. A typical example would be correlation-based methods, where the results will be distorted and interpretation severely hindered if variables with a high proportion of censored data are included. Such variables can be used to document the data distribution and for mapping (though with care due to possible data quality problems). When highly censored data exist for elements that are characteristically high for some particular process, such as the formation of a mineral deposit type or industrial contamination, their presence in measurable amounts can be used to confirm

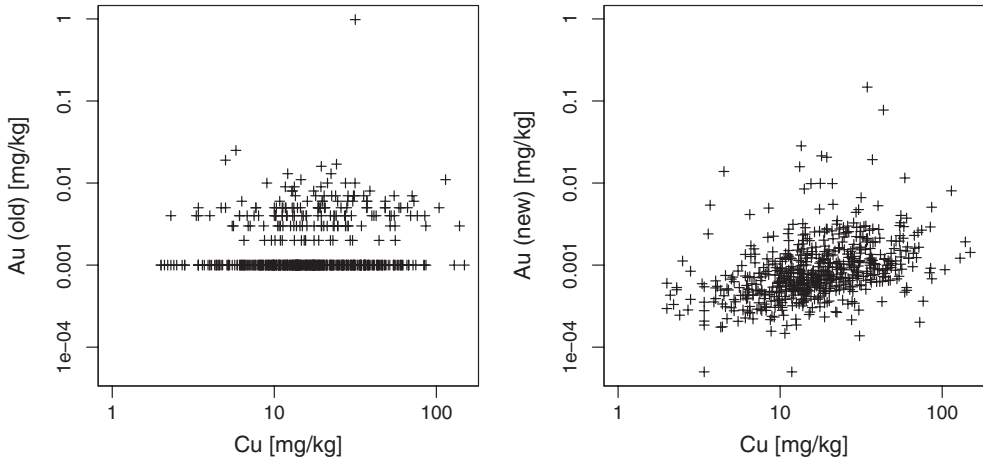


Figure 18.8 Scatterplot for Cu versus Au old (left) and Cu versus Au new (right) in Kola Project C-horizon soils

an interpretation, based on the remaining elements present in reliably measurable amounts. The two maps constructed with the old and the new Au data (Figure 18.9) show the major difference – the regional data structure only becomes visible with the complete data set. In the map constructed with the new Au data (Figure 18.9, right), the granulite belt (see geological map, Figure 1.2) is generally enriched in Au, and a linear Au anomaly marks the Pasvik valley, which follows an important fault zone. An interesting anomaly occurs in the centre of the Finnish survey area, marking a region with many known Au showings. With

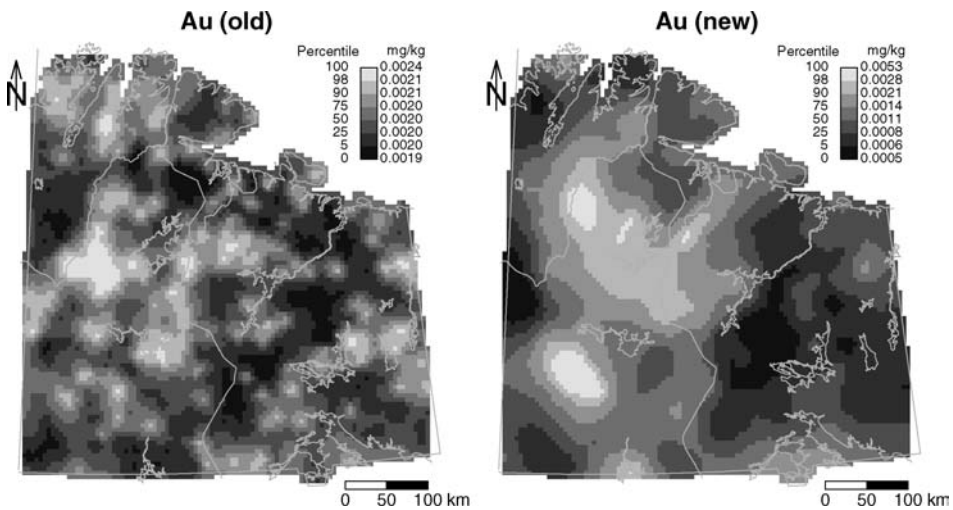


Figure 18.9 Distribution maps of Au determined with the old (left) and new (right) method. For Au determined with the old method about 75 percent of all values were below the detection limit (Figure 18.7)

knowledge of these features it may be possible to detect them in the map for the original Au determinations. However, the original map by itself is so noisy that no clear conclusions about the processes determining the regional distribution of Au can be deduced.

18.9 Summary

Quality control should be an integral part of any project in applied geochemistry and environmental sciences. The design of a QC protocol needs to be an integral part of project planning. International reference materials, at least one project standard, field and analytical duplicates all need to be inserted in any batch of samples to be analysed. A contract with the analytical laboratory should clearly spell out the consequences if quality problems are detected. It is neither sufficient nor prudent to rely on the quality certification of any laboratory. Ultimately the survey activity is the responsibility of the persons organising it, and it is their responsibility to independently check the quality of all aspects of the work from planning, through field and laboratory activities to final report preparation.