QUALITY OF DATA IN ENVIRONMENTAL ANALYSIS

KEES J.M. KRAMER

Mermayde, P.O. Box 109, 1860 AC Bergen, Netherlands;
kees.kramer@mermayde.nl

Abstract. Environmental science is often the interpretation in an environmental context of information obtained by analysis of representative samples. Clever interpretations and combination of results thus may lead to new discoveries. However, without information from quality analyses, such discoveries will not be made or, even worse, one may arrive at erroneous conclusions. The environmental scientist shall be able to demonstrate the quality of his/her results. This shall be demonstrated within the laboratory (precision) as well as between laboratories (accuracy) in order to avoid bias and non consistency between data sets of different analysts/laboratories. A number of issues that may influence the quality of data, and thus of information, are discussed, including examples of possible pitfalls in sampling, sample treatment and analysis. It is motivated that environmental laboratories shall implement quality assurance (QA) and quality control (QC) principles to demonstrate that they produce quality analytical data, which form the basis for quality science.

Key words: sampling, analyses, quality of results, quality control, quality assurance

INTRODUCTION

In nearly all fields of environmental science, including the marine sciences, scientists have to rely on measurements, often on samples obtained from the rather inhomogeneous natural environment. With the results they intend to explain their findings, correlate, evaluate, and wonder about outlying data points. They usually assume that the data, the results of the measurements are 'true' (accurate) and unbiased.

Already in the late 1970's John Taylor of (then) the US National Bureau of Standards lectured on the quality assurance of chemical measurements at environmental conferences. His message was clear: check on the quality of your results, because with false data you get bad science (Taylor, 1987). Scientists don't like to hear that their measurements may be less true, less accurate than they thought. The feeling among the scientists at such an environmental conference was usually one of 'we are the experts in our fields, and thus our data are OK'.

But were they, or even: are they? Today the environmental scientist usually has training as environmental chemist, marine biologist or sedimentologist. In this capacity they may have to carry out measurements, perform analyses. It is not surprising that their major interest will not necessarily be with the measurements but the interpretation of the results. That is considered their main job, that is what their publications are about. Usually a marine chemist is trained in chemical oceanography, not in the specialism analytical chemistry.

For (environmental) research it is essential that the analytical result obtained, the data, is comparable within the laboratory: day-to-day variation shall be small. In addition, the variation between different laboratories shall be small enough (results are comparable) so that one can use each others data in a scientific interpretation, ensuring quality of information. Methods have been developed that support quality of results. They include 'quality assurance' (QA), the organisation of the work that makes quality measurements possible, and 'quality control' (QC), the necessary checks to verify that indeed the analyses are under quality control. The ISO 17025 standard outlines the general requirements for the competence of testing and calibration laboratories (ISO, 2005). This standard is not to be used by specialised analytical laboratories only. As environmental scientific laboratories are testing (analysing) too, they shall by principle comply with this standard as well. For accreditation of their analyses it is the vital guide.

In this short note the message will be similar to the one of John Taylor: be aware of pitfalls in the analytical process, check on your measurements and analyses. Be sure that they are OK or, even better, demonstrate they are OK. Demon-

strate to yourself and to the outside world, so that proven true information can be used to feed your models, your hypotheses, your monitoring trend analyses, etc. Several pitfalls and errors that commonly occur in environmental analysis are discussed, not to tell that your analyses are wrong, but to help you realise that one shall be vigilant that quality data are produced in support of quality science.

POTENTIAL PITFALLS IN ANALYSIS

'Analysis' is more than only the use of an analytical instrument. It is a combination of connected activities:

- selection of the sampling location,
- sampling operation,
- sample treatment,
- transport and storage,
- sample preparation,
- instrumental analysis,
- calculation,
- statistical evaluation, etc.

Each activity offers possibilities for mistakes to occur. The result of any test on the quality of the environment is no better than the result of all efforts that lead to the final result. When one link in the chain of activities is not under control, in the end data are of poor quality, leading to false information. Each activity shall be evaluated for possible pitfalls, and each activity shall in principle be carried out under a quality assurance regime (Quevauviller, 1995).

Representativity in sampling

The objective of sampling is "to collect a portion of material from an environmental compartment (either water, sediment or biota), small enough in volume to be transported conveniently and handled in the laboratory, while still accurately representing the part of the environment sampled". The word representativity has a central place in this passage not only in terms of whether the portion of the sample really represents the original environment, but also whether the sampling and following sample handling remains under sufficient control that no changes in the analyte content, which can be either increase (*e.g.* contamination) or loss ((bio)degradation, sorption, evaporation), occur. The lower the analyte concentration, the more critical these effects become. Materials in contact with the sample (samplers, bottles and jars, tubing, etc.) shall be critically evaluated for such analyte enrichment/loss effects, and cleaning methods shall be validated.

Representativity is also important in the definition of the sampling plan, such as the selection of the sampling location and frequency. Is the sampling strategy applied optimal for a best description of the environmental compartment in space and time? The geographic sampling locations are in practice often determined by logistic possibilities (such as easy access), and availability of staff. One should realise that if not properly selected (or maintained in consecutive sampling events), the sampling location may bias the final analytical result, which is in contradiction with the representativity principle (Kramer, 1994). Validation of the sampling locations (geographic position and sampling depth) is therefore an important aspect of the definition of the sampling strategy.

Sampling frequency may vary from continuous measurements and (semi)continuous sampling, to once or few times per year. It will be obvious that for the compartment water with its rapid changes, no representative result can be achieved when the sampling frequency is too low. On the other hand, changes in soil or sediment matrices are less subject to temporal change. The number of samples that can be collected and analysed is often a function of available staff (including analytical capacity) and availability of technical means (including cars and research ships). It is in the end largely limited by financial constraints. Sampling may be restricted to 'typical' seasons of the year (spring or winter situation, dry period or monsoon), or cover smaller time intervals (following the development of a phytoplankton bloom, fortnightly: spring tides, diurnal: tidal regimes), but nature is often not very predictable. In situations where natural events occur infrequently, such as rain storms causing rapid increases in river discharges, 'peaks' are often not detected, not even when weekly samples are collected (Kramer, 1994). This under-representation of events being sampled was also statistically demonstrated by Walling & Web (1984) for sediment loads in a river/estuarine system. We must realise that 'events' may seriously affect the redistribution of *e.g.* water or sediments, and that they occur in weather conditions that do not allow sampling at all ("oceanography stops at 8 Beaufort wind speed").

Sample treatment

Once the sample has been collected it will be usually transported to the laboratory, but in some analyses sample treatment is required directly after sampling in the field. If this is limited to the addition of preservatives the major concern is contamination of the sample (via the added compound, or through the air (diesel fumes, dust)). Other on site treatments may involve *e.g.* sieving of sediments or the filtration of water. The relatively large surface area of small sedimentary particles results in highest concentrations of pollutants in the smallest fractions. Analysis of total sediment is quite common (after sieving over 2 mm to remove non-sedimentary material), but the analysis of the 'fine fraction' may lead to a better estimation of the distribution of pollution. Thus the mesh size of the sieve is important. Although many use 63 µm as standard mesh, others apply their own standard of 2 µm or 20 µm. Results obtained by such different methods can obviously not be compared. In addition, the construction material of the sieve (nylon, brass) may interfere with the analyte content.

In filtration of water samples similar artefacts may be present. In some monitoring programmes 'total water' samples are analysed, meaning that the concentrations are largely subject to the amount of suspended particulate matter present. If filtration is applied it will matter what type of filter

is used (diameter, pore size, material), the cleaning operation and what type of filtration method (suction or pressure). Although 0.45 µm filtered samples are by empirical standard considered as dissolved, filters of 0.4 µm and 0.7 µm are used as well in environmental analysis. Clogging of the filter will reduce the pore size, even to the limit where colloidal material is trapped. By comparing different filters and methods distinct differences have been reported *e.g.* for the filtration of water samples intended for trace element (Horowitz *et al*., 1992) and for organo-chlorine compound analysis (Hermans *et al*., 1992). Again, different methods may thus lead to results that are not comparable.

In parallel to filtration centrifugation is used, sometimes to collect large amounts of particulate matter (flow-through centrifuge). As the physical principles differ (filtration by size, centrifugation by density) different fractions may be collected. For example, particles that float in the water column, such as phytoplankton, may not be trapped by the centrifugation process, but are collected on filters, rendering comparison of the analytical results difficult, if not impossible. This is also true for comparing filtrate and centrifugate (Hermans *et al.*, 1992).

Dissolution of sediments for trace metal analysis is another source of problems in comparing analytical results. Many use *aqua regia* (HNO_3 + HCl), others apply HNO_3 only or a cocktail of different acids: (HNO_3 + HCl + $HClO_4$), or (HNO_3 + $HClO_4$ + HF), in different ratios. Especially in silicate rich matrices true dissolution is only reached when hydrofluoric acid is included.

For the analysis of organic micro pollutants (pesticides, PCBs, PAHs, etc.) many different extraction procedures are used, involving one or more solvents in different ratios, using different methods. Also here the 'total' content is the analytical results sought after, and partial methods shall be avoided. It is common practice to test for the recovery in these analyses. Surprisingly, different methods are applied in how to treat the recovery results. They range from no correction at all, to recalculation of the analytical result based on the recovery factor found. This can be understood for a recovery of 90%, even 80%, but what is the scientific meaning of a recovery of 70 or <60% or, the other extreme, 110% which is also found? Methods shall be validated in each laboratory, and confirmation by interlaboratory comparison studies is essential to demonstrate comparability of results.

Transport and storage

Immediately after sampling the sample shall be preserved in the best possible way depending on both matrix and analyte. Addition of chemicals to stop microbiological activity, addition of acids or solvents to keep the analyte in the dissolved phase are used. Less attention may be paid to the transport conditions. Many environmental samples are sensitive to physical disturbance or degradation. It makes no sense to transport a sensitive sample for half a day at ambient temperature (30°C in summer time) and then store it at safe conditions (-18 °C) in the laboratory freezer. Similarly, a sediment core collected for geochronological analysis of thin sediment slices will become useless if it is transported without precautions in a car on a bumpy road: the originally layered structure will be seriously disturbed, rendering any analysis a waste of time and energy. Transport conditions shall be as good as the well controlled laboratory storage.

Confusing units

Units of expression shall be rightly understood and not be subject to confusion. Although we have adopted the SI system (but not all have it implemented: the inch, gallon and stone still survive), units may lead to confusion. Examples are 'ppm' (parts per million) and 'ppb' (parts per billion), or, worse, ppt (used for parts per thousand and parts per trillion). These 'units' mean nothing when they are not further explained (by weight: w/w, or by volume v/v). The use of ppm or ppb in a mixed nature (w/v) is erroneous, even when used for the matrix water. Preferred nowadays is: mg/kg (for ppm) and µg/kg (for ppb).

For the analysis of nutrients there have been used many different units, their use often depending on the scientific community and/or fashion. The following range can be found in the scientific literature: mg/L, µmol/L , µM, 10-6 mole/L, µg-at/L, nmoles/dm^3 (but for each as what: NO_3^- or NO_3^--N?), µg N/L and meq/L. Trace elements in seawater are reported as nmol/kg and µg/kg. In an analytical sense the unit nmol/kg is correct (in the measuring process volume is subject to ambient changes, *e.g.* due to changes in temperature and air pressure), but did the analyst indeed use the mass of the sample, or converted the volume to the mass assuming that the density of water is 1? Erroneous, if it was a seawater sample (density of seawater of 30 psu at 20°C is close to 1.021).

Regularly seen, even in refereed journals, is the erroneous use of a slash twice in a unit. For example, primary production defined as gram carbon per cubic meter per day is presented as $gC/m^3/d$, where correct would be $gC/m^3.d$ or $gC.m^{-3}.d^{-1}$.

A typical problem with units is the use of the µ (micro). It sometimes still happens that in converting a file from one computer (programme) to the next the 'µ' is not recognised and read as 'm'; this is the more problematic as it does not look strange: *e.g.* both units µg and mg exist, but they differ three orders of magnitude. One should be aware of this problem, notably during proof reading of texts.

Calculation errors

The occurrence of calculation errors is underestimated. The conversion from mg to mmol v.v. may seem too simple to discuss, but reality learns that errors are relatively abundant. Use of spreadsheet programs may result in calculation errors as well. It is simple to copy/paste the contents of a cell to another position, forgetting that the original cell contains not a value but a formula ('paste as value' could solve the problem). Complex calculations shall be checked and verified before-

hand. The worksheet shall then be password protected to ensure that no unintended changes occur in future use.

In interlaboratory comparison studies (proficiency testing schemes, PTs) normally the majority of the data sets converge around the reference value. But regularly there are data sets that stick out, not only random, but by a factor of 2 or 10. This is very often due to errors in the calculations that relate to dilution factors in the analysis. Also 3 orders of magnitude differences occur, but these may be traced back to not checking for the unit of expression (mg/kg is requested, while the laboratory normally reports in µg/kg).

Statistical methods

Most environmental scientists have no specialist training in statistics. This leads to application of statistical tools as a black box. This does not necessarily offer a problem provided that these methods are fit for purpose and the application is validated in the laboratory. For example, analysis of variance (ANOVA or F-test) is widely used to compare data sets and they are readily available in spreadsheet programmes. But it is easy to misinterpret data entry between columns and rows. In the test results, shall F = MSbetween / MSwithin be larger or smaller than Fcritical for the number of degrees of freedom and selected probability level α. And what shall this α be: 0.05 or 0.01? Validation of the use of the statistical test by applying to a known example (do I get the same result?) is required for validation.

The natural environment is rather inhomogeneous and any environmental scientist will see from time to time outlier results. What is then their strategy: include them or not; and why? Throw away results (automatically if the value exceeds a threshold, like 10%) because they do not fit your model? But maybe your model is not correct, or too limited? What is called an outlier? Is it by your 'expert opinion' *i.e.* you looking at the data set, or is it the result of a statistical outlier test (better)? If the latter is the case (there are many: Grubbs(1, 2), w/s, Tietjen-Moore, Hampel, …), did you select the appropriate one (also meaning: checking on one or more outliers on one side, on both sides)? Beware of using different tests together to the same data set.

Selection of analytical method

In the environmental sample there is only one (total) concentration of a given analyte. Different (validated) methods applied within or between laboratories shall give the same result: confidence intervals of mean values of replicates shall overlap. As long as this is demonstrated (*e.g.* between lab variation is tested by participation in a suitable proficiency testing scheme) the method is suitable. The core of producing quality data is the use of validated methods.

Many use (or are forced to) 'standard methods', issued by *e.g.* national metrological organisations, by ISO, or an industrial sector. They describe in detail all operations (sometimes instruments) that are required for a proper analytical result. Before becoming a 'standard method' they have been thoroughly tested and validated. It is, however, a misconception that by following the standard method the result of the analysis is thus alright, is thus of high quality. Like in cooking, where the mere following of a recipe does not necessarily lead to tasty food, the standard method shall be validated as well by the laboratory. Once proven, validated standard methods may lead to better comparison of data between laboratories. There are, however, a few issues that shall not be forgotten. Firstly, it takes 5-10 years before a method has arrived at the status of being a standard method. It may be that, when finally publicly available, analytical science has further developed and the method tends to be out of date. Similarly, for new techniques no standard methods are available (yet) and it happens that regulations do not allow very useful instruments / analytical techniques to be used without the availability of a standard method. Secondly, a standard method may result in a good overlap of results between laboratories, but one shall be aware of the risk of method dependent, systematic bias.

QA/QC APPROACH

Traditionally the analyst would argue that his results were OK because "We do it for 20 years", or "We use centrally approved standard methods". Since the 1980's also in environmental laboratories one is tuned to implementation of QA and QC (Prichard, 1995, 2000; Wenclawiak *et al.*, 2004). In order to become accredited, laboratories have to implement ISO 17025 and they have to apply QC methods (use Control Charts, regularly participate in PT schemes), and they are visited at intervals by a representative of an accreditation body. In general there is now a tendency that non-accredited laboratories get less work: they can not demonstrate that they deliver quality data. Also in scientific publications there is a tendency that analysts are requested to indicate how the quality of the data has been assured.

Laboratories shall have an adequate system of quality assurance, and shall be able to demonstrate the quality of data. Within the laboratory the day-to-day variation (repeatability) is tested by using reference materials or quality control materials (RMs or QCMs). A QCM is tested in every batch of routine analysis and because it has the same concentration of analyte, it shall give in principle each time the same result. As conditions are never exactly the same (between days variations) a limited deviation is allowed. An elegant method of QC is to plot these results sequentially in so called Control Charts or Shewhart Charts (Prichard, 1995; Cortez, 2002). It consists essentially of a statistical method to check on whether the analysis is still under control. After a number of replicates of the control sample have been analysed the mean and standard deviation (σ) can be calculated. So called control lines at +/- 3σ are defined. If measurements have the same sources of error, 99% of the results falls $-3\sigma < x < +3\sigma$, the control zone. Similarly a warning zone is defined at $-2\sigma < x < +2\sigma$ (covering 95% of the results). When a next QCM result is plotted in the control chart and the point falls outside the control zone

there is a statistical reason to suspect the result. Immediate action is then required to sort out the (analytical) reason for this (statistically) abnormal analytical result, and to take appropriate action, if necessary.

Even when the day-to-day variation is under control, there may be a bias: replicates are precise, not necessarily accurate (it may deviate from the accepted true value). Certified reference materials (CRMs) then serve to test on the accuracy. CRMs consist of samples that are certified for the concentration of one or more analytes. If a laboratory, by analysing a CRM, finds the same value as stated in the certification report (within the uncertainty range), accuracy has been demonstrated. Where QCMs are tested in principle in each batch, CRMs are used much less frequent. They are usually purchased from CRM producers (see *e.g.* www.VIRM.net for a list).

Another method to test for comparability of results is to participate in proficiency testing schemes. They are organised for many types of analysis, often on an international scale, with up to several hundreds of participants. All get the 'same' sample and submit their analytical result (XLab) to the PTs organiser. A common approach in the evaluation is the use of a so called Z-score. During the evaluation a reference value (XRef) is defined (there are several approaches, see *e.g.* Cortez, 2002) with a standard deviation (sRef). The Z-score is defined for each PT participant as: Z = (XLab - XRef) / sRef, thus normalising the expression of the evaluation as being independent of concentration range. If the evaluation for a given laboratory Z > |3| this is considered as the analysis being out of control, action is required; when |2| < Z < |3| this might be considered as a warning signal.

Maintaining Control Charts, use of CRMs and regular, successful participation in PT schemes are essential ingredients for being an accredited laboratory, producing quality data.

CONCLUSION

Quality data are the basis for quality information, for sound scientific results. Quality is in part the selection and use of methods that are fit-for-purpose, in part it is the quality driven organisation of analyses, with built-in tests to demonstrate that quality data are produced. It is easy but rather foolish to hide your quality position behind 'clever' wording:

- "We follow accreditation procedures" may indicate that you are not accredited (yet);
- "We are accredited", but are you for all types of analysis?
- "We have participated in PT schemes" may hide that your Z-score > |3|, or that you do not do so any more;
- "We are a National Reference Laboratory" gives you certain responsibilities, but makes you not immune to mistakes.

In quality of data be honest; with yourself and with your clients (which may be science itself). Good (environmental) science is served by the statement: "Better no data than poor data".

REFERENCES

Cortez, L., 2002 - Use of LRM in Quality Control: Control Charts and Interlaboratory Testing. K.J.M. Kramer (Ed.). Mermayde Publ., Bergen, Netherlands (CD-ROM)

Hermans, J.H. Smedes, F. Hofstraat, J.W. Cofino, W.P., 1992 - A method for estimation of chlorinated biphenyls in surface waters - influence of sampling method on analytical results. Environ. Sci. Technol. 26: 2028-2035

Horowitz, A.J. Elrick, K.A. Colberg, M.R., 1992 - The effect of membrane filtration artefacts on dissolved trace element concentrations. Wat. Res. 26: 753-763

ISO/IEC 17025:2005 (E), 2005 - General requirements for the competence of testing and calibration laboratories. International Organisation for Standardisation, Geneva, Switzerland

Kramer, K.J.M., 1994 - Inorganic contaminants in the water column: sampling and sampling strategy. Intern. J. Environ. Anal. Chem. 57: 179-188

Prichard, E. (Ed.), 1995 - Quality in the analytical chemistry laboratory. Wiley, Chichester, pp. 307

Prichard, E., 2000 - Quality in the analytical chemistry laboratory. EC project Quash report, EUR 19088, Luxembourg, pp. 111. Note: This report has been translated in all west and central European languages. Downloadable in .pdf format from: www.VIRM.net.

Quevauviller, Ph. (Ed), 1995 - Quality assurance in environmental monitoring. Sampling and sample pre-treatment. VCH Publ., Weinheim, pp. 306

Walling, D.E., B.W. Webb, 1985 - Estimating the discharge of contaminants to coastal waters by rivers: some cautionary comments. Mar. Poll. Bull. 16: 488-492

Wenclawiak, B.W., Koch B.M., Hadjicostas E. (Eds.), 2004 -Quality Assurance in analytical chemistry. Training and teaching (with CD-ROM). Springer, Berlin, pp. 280