Is measurement uncertainty from sampling related to analyte concentration?

Article (Accepted Version)

Is measurement uncertainty from sampling related to analyte concentration?

Stephen L R Ellison[a], Michael H Ramsey[b], Paul Lawrance[a], Brian Stuart[a], Jesus Minguez[a], Michael J Walker[a]

[a] LGC limited, Queens Road, Teddington Middlesex TW11 0LY

[b] School of Life Sciences, University of Sussex, Falmer, Brighton, UK, BN1 9QG

Abstract

The contribution of sampling to the combined uncertainty of measurement is assessed using a combination of literature review and experimental determination of sampling variability in a range of foodstuffs in order to determine whether there is a consistent relationship between analyte level and proportion of variation attributable to sampling. Experimental determinations used the duplicate method, an economical method of assessing the relative contributions of sampling and analytical variability to the overall variance of results. The experimental work covered sampling of retail foodstuffs. 101 estimates of between-target, between-sampling,and within-sample variance were obtained. It is shown for the first time that sampling variance across the food sector appears to follow a Horwitz-like relationship sufficient to provide estimated between-sample standard deviation to within approximately an order of magnitude. The results from different methods of data processing for sampling uncertainty experiments are also compared. It is shown that for the data sets obtained experimentally in this study, log-transformation is of minor importance while the use of robust statistical methods can have greater but less predictable effects on estimated sampling variance.

Introduction

The importance of measurement uncertainty for improving the reliability of compliance decisions in the food sector has been recognized internationally by Codex Alimentarius[1]. This recognition currently has only considered the uncertainty arising in the analytical part of the measurement process, but is now becoming clear that the sampling process is also part of the measurement process and often a larger source of uncertainty[2]. For analytical uncertainty there is a wealth of knowledge of values for a wide range of analytes in food and feed generated over at least ten years; for precision of test methods, the knowledge base goes back considerably further and has resulted in the identification of several useful relationships between analyte level and reproducibility standard deviation[3,4] which are useful in setting criteria for analytical test method performance[3,5] or for proficiency testing.[6,7] For sampling uncertainty such a volume of information has not so far been available and it is not known whether any consistent relationships exist between, for example, analyte level and the proportion of variability due to sampling. This project therefore aimed to collate and extend the available data for sampling uncertainty across a wide range of food and feed materials, at various stages of their production, and for various analytes, with a view to identifying any consistent relationship that might assist in setting regulations or guidelines for future sampling strategies.

The majority of the data presented here were assembled via a review of the available literature. This included data from previous research on cost-effective methods of characterising sampling uncertainty studies,[8-11] together with a range of additional sources (see below). Much of the published data, however, focused on bulk sampling of a comparatively small range of commodities. To ensure that data on typical analytes in retail foods was included, a range of additional retail foods, selected to provide complementary data to that found in previous studies, were sampled and analysed.

Historically, sampling variation[1] has been characterised by large studies involving hundreds of observations. This was not feasible for the present work. Instead, advantage was taken of more cost-effective methods. International guidance published by Eurachem on approaches to the estimation of measurement uncertainty arising from the sampling process provides a simple and comparatively economical approach to estimating sampling variation.[12] This guidance was informed by prior research on simple methods for characterising sampling variation[13-16] which has demonstrated that it is capable of characterising sampling variation with as few as eight sampling targets.[16]

In addition, the project sought to compare different methods of processing the data obtained from these simple studies. Existing guidance provides a range of options, including classical and robust ANOVA,

[1] We use ‘variation’ here for the general concept; we will use ‘variance’, ‘standard deviation’ and ‘standard uncertainty’ for particular statistical parameters characterizing the variation

with sampling uncertainty expressed as a relative standard deviation. It has, however, been suggested that robust ANOVA is not always appropriate[17] and, in addition, that log-transformation might be appropriate prior to analysis where the variation is known to be proportional to analyte level and the range of results is wide.[18] In this work the experimental data were processed by a number of different methods in order to assess the differences between robust and classical variance analysis in the context of sampling uncertainty evaluation and to assess whether log-transformation had any material effect on results.

In this paper, we first describe the initial literature review. We then discuss the results of experimental determinations of different retail products, with particular attention to the various calculation methods used. Finally, we review the accumulated data with a view to identifying any general trends.

Experimental

Sampling was carried out by staff with experience in food survey sampling following available guidelines for trading standards enforcement. Sampling plans followed the duplicate method, using eight distinct sampling targets for each foodstuff studied with one exception, for which 10 materials were acquired. The use of eight sampling targets has been shown to provide a sufficiently reliable estimate of the sampling variance for comparison purposes.[16] Sampling was from multiple retail outlets; where products were packaged and on shelving, each item was taken as far as possible at random. All the materials were analysed using standard methods of analysis for foodstuffs in an accredited laboratory using appropriate internal quality control methods.[19] Table 1 lists the foodstuffs and analytes studied.

Statistical analysis used two principal methods for comparison: classical analysis of variance (ANOVA) and Robust Analysis of Variance. Variance component estimates were also calculated using restricted maximum likelihood estimation (REML) as this is a recommended alternative to classical ANOVA, giving closely similar results for balanced data and generalising more easily in cases of imbalance.[20] To assess the effect of log-transformation, classical ANOVA was also applied to log-transformed data. Classical ANOVA and REML estimation were carried out using the R statistical programming environment,[21] version 2.9.2 or later; robust analysis of variance used the ROBAN package.[22,23]

Literature survey of sampling uncertainty in food analysis

Literature review covered a range of food and analytical journals and other publications extending to 2010. Twenty separate sources were identified, including references cited above[8,9,12,15] and a further sixteen sources[24-39]. Table 2 (see Electronic Supplementary Information) summarises the range of foodstuffs, sampling methods and sampling variance estimation methods found. Seventeen products had been studies in wholesale, farm or factory environments; thirteen in retail. The majority of

products, including all the products studied in retail environments, had been studied by the duplicate method using 8-10 sampling targets. Wholesale products, largely studied by a group in the United States,[32-39] had employed purpose-designed sampling schemes. One example[12] had been modelled following a general scheme due to Gy[40]. In all, 21 products (some in more than one packaging or storage state) and 59 distinct analyte/product combinations were identified. Analytes included trace metals (primarily contaminants), proximates, pesticides and mycotoxins. Mycotoxins are well known to be important food contaminants, present at low levels and prone to high sampling variance, so many of the larger studies - though not the majority of analyte/product combinations - had focused on mycotoxins in granular products such as grain, nuts, or green coffee.

Table 3 (see Electronic Supplementary Information) lists the data on sampling and analytical variation extracted from the literature. Analyte concentrations range from proximates in the high percentage range (for example, fat in butter) to low level contaminants at mass fractions of the order of 10^{-9} (parts per billion). The proportion of measurement variance (defined as the combination of sampling and analytical variances) attributed to sampling variability ranges from zero to approximately 70%. This was encouraging from the point of view of trend analysis, as it is likely that with such large ranges any consistent association between analyte concentration and proportion of variance due to sampling should be readily detectable.

It is important to note that this data set is, by comparison with data sets on analytical variability in interlaboratory studies[3] or proficiency testing schemes,[6] neither large nor representative in the sense of random, representative selection from a larger population of all foods and analytes. The range of product and analyte types is broad considering the size of the data set, but it is clearly possible that product/analyte combinations known to be hard to sample are disproportionately represented in the set. For this reason, and to add to the number of different retail materials studied, a number of additional materials sampled from retail outlets were studied using the duplicate method. These studies are described below before returning to the consideration of trends in sampling variance.

Experimental studies of uncertainty of sampling in retail products

Sixteen product type/analyte combinations were studied to improve coverage of retail products. Each different analyte class was the subject of a separate sampling exercise, resulting in eight separate sampling exercises. Milk and lemonade provided materials with low expected sampling variability. Bread and lamb mince were expected to show moderate sampling variability for major components, and potentially high sampling variability for minor components. Metals (in bread) provided examples of some common nutritional elements rather than contaminants. Histamine in tuna (a spoilage indicator) was expected to show high product to product variation and moderate sample to sample variability. Alkaline phosphatase in milk was chosen to provide an unusual enzyme activity measurand, whilst

proximates (which include ash, protein, moisture, fat and dietary fibre in foods) provided further examples at comparatively high mass fraction.

All of the materials were sampled using the duplicate method, duplicate samples being taken randomly from shelving with the exception of lamb mince, for which duplicate samples were collected on two different days that were equally likely to be selected, given the sampling protocol. Analysis used standard methods and as far as possible was undertaken under repeatability conditions. The results on the raw data are given in Table 4; the corresponding results from the log-transformed data are presented as Table 4a. REML gave essentially identical results to the classical ANOVA calculations for the balanced data sets in this study and the REML results are not reported. To obtain the means and standard deviations in Table 4a, the conversions in equation 1 were used:

$$\bar{x} = \exp\left(\bar{x}_{\ln}\right) \quad \text{and}$$

$$s = \sqrt{\exp\left(2\bar{x}_{\ln} + s_{\ln}^2\right)\exp\left(s_{\ln}^2 - 1\right)} \qquad \text{(eq 1)}$$

where $\bar{x}_{\ln}$ is the mean of the log-transformed data ln(X) and $s_{\ln}$ the appropriate standard deviation calculated from the log-transformed data (this calculation is based on the known variance of a lognormal distribution with known mean and standard deviation of the logarithm; for small relative standard deviation prior to log transform, the standard deviation approaches $\exp(\bar{x})s_{\ln}$).

To compare the effect of log-transformation, Figure 1 shows box plots of the ratio of the respective RSDs calculated using raw data and using log-transformed data. In these plots, if there were no effect the individual box plots would centre at 1.0; values above 1.0 indicate that the standard deviation from log-transformed data is lower than that calculated from raw data. With the exception of three apparent outliers, the plots indicate a slight decrease of approximately 8% in estimated standard deviation when taking logs. Considering that the typical confidence intervals for standard deviations estimated with between 7 and 16 degrees of freedom, as for the balanced nested design used in the 8-target duplicate method, extends above and below the estimate by a factor of at least two, an 8% bias is unlikely to have any important effect. It is, of course, important not to over-generalise; in this data set the range of mean values for each product type is rarely as large as a factor of two itself, leading to a comparatively small change in estimated standard deviation on log-transformation. Were the mean values for different targets within each data set considerably more dispersed, the effect of a proportional standard deviation would be more important and log-transformation more strongly recommended. Considering the relatively small effect in this data set, together with the fact that much of the literature data in Table 3 did not apply any transformation before estimating sampling uncertainties, we chose to retain the standard deviation estimates obtained from the raw data for the purpose of the present review.

Figure 2 compares robust versus classical treatment for this data set by showing the ratio of robust to classical estimate for each of the three standard deviations estimates in each of the 16 product/analyte

combinations. It is clear that while the between-target standard deviation and the analytical standard deviation are relatively stable, with a slight tendency towards lower robust estimates, the sampling variability estimates can change considerably and in either direction on application of robust ANOVA. This range of values has different causes for each individual data set. For example, for magnesium in bread, one pair of analytical replicates shows an unusually high difference, inflating the estimated analytical variance and (because the calculation involves subtraction of ANOVA mean squares) reducing the sampling variance estimate. Robust analysis downweights the outlying replicate pair, resulting in a reduction in estimated analytical variance and a more important increase in estimated sampling variance. For vitamin C in infant formula milk, there were no compelling outlying analytical duplicates, but the data set included one material with generally high vitamin C (ca 15 mg/100 ml compared to the robust mean of 12 mg/100 ml) which additionally included the largest between-sample and between-analysis differences. The effect of robust analysis in this case was to reduce both the between-target and sampling variance estimates, leaving the analytical variance largely unchanged.

It follows that there is no simple generalisation that can be made about the merits or otherwise of robust analysis; the most appropriate approach depends in part on the particular features of the data set and the principal aim of the analysis. For example, analysis aimed at characterising a central majority of variation or likely to suffer from analytical outliers would choose robust analysis; a study aimed at establishing the complete range of variation observed would use classical analysis or modelling that took account of any known distribution properties. For this survey, we chose to retain both estimates for graphical review and to use the robust estimates for any numerical comparisons, again in part because at least some of the prior work summarised here also employed robust estimates.

Trends in analytical and sampling uncertainty

The combined literature and experimental data, comprising 101 estimates of sampling, analytical and between-target standard deviation, were then inspected with a view to identifying any trends, in sampling or analytical uncertainty, with concentration. Sample size was reported quite differently for different products - for example, kg, number of packages, number of individual lettuces etc. – and only one sample size was reported for each product. We were therefore unable to draw meaningful inferences about the effect of sample size on analytical uncertainty.

Initial inspection of the analytical repeatability estimates showed a general increase in RSD with decreasing concentration as might be expected from Horwitz' equation;[3,4] this is illustrated in Figure 3a). Figure 3b) shows a broadly similar picture on inspection of the sampling RSD (denoted RSD_{sam} in the figures). Note that, unlike estimates of reproducibility standard deviation, estimates of uncertainty from sampling can be zero; these are omitted from the log-log plot.

The best fit lines, omitting the classical estimates of standard deviation from Table 4, are given by

$$s_{an} = 0.0072c^{0.8176} \quad \text{(eq. 2)}$$

$$s_{sam} = 0.0128c^{0.8099} \quad \text{(eq. 3)}$$

For comparison, Figure 3a) and b) include the relevant Horwitz predictions. For the analytical contribution (Figure 3a), the prediction is taken as 2/3 σ_H/c, where

$$\sigma_H = 0.02c^{0.8495} \quad \text{(eq. 4)}$$

and c is the mass fraction of the analyte in question. The majority of analytical repeatability RSD data obtained in the present study fall near or slightly below the predicted repeatability RSD, though the Horwitz line is somewhat steeper than the fitted line, predicting slightly larger dispersion at higher concentration. Figure 3b) adds the Horwitz predicted SD, σ_H and a further line at Again, there is appreciable scatter about the prediction, but the general picture is of the sampling RSD falling around 1 - 2 times the Horwitz prediction. There is a clear tendency for sampling standard deviation to be slightly lower than RSD_{HOR} at high concentrations (to the right of the plot) and higher at low concentrations; the fitted line is above $2\sigma_H$ for concentrations below 1 ppm

Considerable caution should clearly be used if attempting to estimate sampling uncertainty from equation 3. The residual standard deviation around the prediction of equation 3 was 0.46, corresponding to an approximate factor of $10^{0.46\times2} \approx 10$ possible error for prediction of an individual sampling standard deviation. Estimates from this relationship are therefore order-of-magnitude estimates only. Further, the data set remains comparatively small; although inspection showed that no specific group of analytes or matrices dominates either the present data set or the trends found (for example, the 12 aflatoxins included in the data set follow the same trend as all other analytes), the data set is not guaranteed to be representative, which could lead to additional bias.

An important additional question is whether the proportion of measurement variance attributable to sampling follows a consistent trend with concentration. Figure 4 shows the calculated values for proportion of measurement variance attributable to sampling as a function of concentration; it is immediately evident that although there is a tendency in this data set for very low estimated proportions attributable to sampling to appear only at mass fractions above 10^{-6}, there no evidence of a useful trend. The proportions of measurement variance due to sampling cover almost the whole range from 0 to 1 across essentially the whole mass fraction range from 10^{-6} to 1.0.

Finally, Figure 5 shows the ratio of sampling standard deviation to analytical (repeatability) standard deviation by concentration. Again, there is little evidence of a trend; certainly not a trend useful for prediction. Figure 5 does, however, serve to underline the tendency for sampling to equal or exceed analytical variability: the greater part of the data set falls above the line denoting equal sampling and analytical standard deviation, and the mean ratio of sampling standard deviation to analytical standard

deviation is almost exactly 2.0. A predicted value for the ratio at a particular mass fraction can also be obtained from equations 2 and 3.

Discussion

The possibility of predicting sampling standard deviation in foodstuffs from a simple predictor such as mass fraction is of interest for regulatory agencies, such as the CODEX Alimentarius Commission, because it would offer the possibility of providing more general guidance on the number of sampling increments or individual portions taken when sampling, and would also inform sampling strategies. This is, indeed, the major impetus for much of the published work on sampling of mycotoxins in grains and nuts (see, for example, references 32-39). These extensive studies have generated predictive models for different bulk foodstuffs that have in turn enabled regulators to plan sampling strategies for specific foodstuffs; examples include the detailed sampling plans for aflatoxins in US FDA regulations[41] and the sampling plans adopted in the EU for aflatoxins in nuts[42]. A generally applicable model would simplify planning and potentially simplify regulation.

The evidence from the present review, however, is that while there does appear to be a general trend in sampling standard deviation with analyte concentration, the sampling standard deviation for individual products can differ considerably; perhaps by up to a factor of 10 from the predicted values. While this might be improved via studies of, for example, the effect of processing, it follows that for critical analytes and products, where it is important to have more than an order-of-magnitude estimate, it remains necessary to evaluate sampling uncertainties on a case-by-case basis.

The particular form of the trend found in this study is of a Horwitz-type relationship, that is, a relatively simple power curve relating dispersion to concentration. Broadly similar models have been used to predict sampling dispersion in particular products; for example the sampling variance for aflatoxin in dried figs has been modelled as $(590/n_s)2.219c^{1.433}$, where n_s is the number of figs included in the sample.[43] It is not clear why such a model might apply to sampling variance; it does not appear to follow naturally from sampling theory. However, the form may simply be a matter of choice; with relatively large residuals around the fitted log-log model, a range of possible smooth models could be considered and a linear fit is the simplest of these.

It is also of some interest that the analytical standard deviation is, from equations 3 and 4, close to a factor of two smaller than the sampling standard deviation. Again, the reason for this is unclear. However, it would be good practice to ensure that analytical variance is small compared to sampling variance. A common recommendation is to ensure that a measurement standard deviation is smaller than a sampling standard deviation by a factor of two or more; for example, Ramsey *et. al.* recommended that analytical variance be less than 20% of sampling variance for geochemical analysis, at which point the ratio of sampling to analytical standard deviation is 2. The observed ratio may therefore be unrelated to any natural limitation; it may simply be an outcome of ‘sufficient’

method development, as suggested by Thompson,[45] in which measurement methods are developed to the point of acceptable precision and not beyond.

Conclusions

Experimental data were processed by different combinations of classical or robust analysis and using raw or log-transformed data in order to assess the effect on the resulting estimates of sampling and analytical standard deviation. It was found that for this data set, in which the range of values for a given analyte/material combination was rarely large, log-transformation resulted in only a modest reduction in variance estimates that is inconsequential by comparison to the expected variability of the estimates. For similar applications - that is, where similar products with similar analyte levels are examined to establish an approximate sampling variance - it therefore seems reasonably safe to use raw data without transforms; for data with large ranges, however, appropriate transformation or modelling that takes explicit account of any known distribution properties should be considered.

The effect of robust analysis versus classical analysis was most important for the estimated sampling standard deviation, which was found to vary appreciably and in either direction on applying robust analysis depending on the particular features of the data set. It follows that no clear generalisation can be made based on the present study; rather, the choice of robust over classical estimates depends on the particular intent of a study and on the likely behaviour of the analytical methods used.

The principal question for this study is whether sampling variation can be predicted usefully from simple dependencies on concentration. The accumulated data set from the literature review and additional experimental studies has been inspected for evidence of trends in uncertainty from sampling. There is a clear indication of an increasing trend in uncertainty from sampling, expressed as a relative standard deviation, as the analyte level falls; fitting a linear log-log relationship led to the relation $s_{\mathrm{sam}} = 0.0128c^{-0.8099}$ (c being the mass fraction of analyte), though with very considerable variation about the fitted line. There was no useful predictive trend in the proportion of variance attributable to sampling variation.

Acknowledgement

This investigation was funded by the UK Food Standards Agency

References

1. Codex (2008) Guidance on measurement uncertainty and uncertainty of sampling guidance on uncertainty from sampling, Codex Committee on Methods of Analysis and Sampling, Twenty-ninth Session, Budapest, Hungary, 10-14 March 2008
2. Ramsey MH (2004) When is sampling part of the measurement process? Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement, 9, 11-12, 727 – 728
3. W Horwitz, L R Kamps, K W Boyer (1980), *J. Assoc. Off. Anal. Chem*., Vol 63, p1344.
4. M. Thompson, P. J. Lowthian, (1997) *J. AOAC International*, Vol 80, 676-679
5. CODEX Alimentarius Commission (2013), Procedural Manual, Twenty-first edition. United Nations Food and Agriculture Organization, Rome, Italy, pp 64-67
6. M. Thompson and P J Lowthian (1995) *Analyst*, vol 120, 271-272
7. M. Thompson (1999) *Analyst*, vol 124, 991
8. Ramsey, M.H., Lyn, J.A. and Wood, R. (2001) Optimised uncertainty at minimum overall cost to achieve fitness-for-purpose in food analysis. Analyst, 126, issue 10, 1777-1783
9. Lyn, J.A., Ramsey, M.H., and Wood, R. (2002) Optimised uncertainty in food analysis: application and comparison between four contrasting 'analyte-commodity' combinations, Analyst, 127, 1252 – 1260
10. Lyn, J.A., Ramsey, M.H., and Wood, R. (2003a) Multi-analyte optimisation of uncertainty in infant food analysis, Analyst,128, 4, 379-388
11. Lyn J.A., Ramsey M.H., Damant A., and Wood R. (2005b), Two stage application of the OU method: a practical assessment, *Analyst*, 130, 1271-1279
12. Ramsey M.H., and Ellison S. L. R.,(eds.) (2007) Eurachem/EUROLAB/ CITAC/Nordtest/ AMC Guide: *Measurement uncertainty arising from sampling: a guide to methods and approaches Eurachem* ISBN 978 0 948926 26 6. (http://www.eurachem.org/)
13. Ramsey, M.H.(1998) Sampling as a source of measurement uncertainty: techniques for quantification and comparison with analytical sources. Journal of Analytical Atomic Spectrometry, 13, 97-104
14. Lyn, J.A., Ramsey, M.H., Damant, A.P. and Wood, R.(2007a) Empirical versus modelling approaches to the estimation of measurement uncertainty caused by primary sampling, (DOI: 10.1039/b709539m.) Analyst, 2007, 132, 1231 – 1237
15. Lyn, J.A., Palestra, I.M., Ramsey, M.H., Damant, A.P. and Wood, R. (2007b) Modifying uncertainty from sampling to achieve fitness for purpose: a case study on nitrate in lettuce *Accred. Qual. Assur.*, 12, 67-74
16. Lyn, J.A., Ramsey, M.H., Coad, S., Damant, A.P.,Wood, R. and Boon K.A. (2007c) The duplicate method of uncertainty estimation: are eight targets enough? *Analyst 132, 1147-1152* (DOI: 10.1039/b702691a)

17. Wilrich, P-T (2010) Comments to EURACHEM/CITAC guide “Measurement uncertainty arising from sampling”, *Accred. QUal. Assur.* Vol 15, pp 529-531

18. Ramsey, M. H., Ellison, S. L. R. (2010) Reply to comments on EURACHEM/CITAC guide “Measurement uncertainty arising from sampling” *Accred. Qual. Assur.* Vol 15 pp 533-535.

19. Thompson, M., Wood, R. (1995) Harmonized guidelines for internal quality control in analytical chemistry laboratories. Pure Appl. Chem., Vol. 67, No. 4, pp. 649-666

20. Searle, S. R., Casella, G., McCulloch, C. E. (2006) "Variance components", John Wiley and Sons, Inc, Hoboken, New jersey, USA ISBN 978-0-470-00959-8

21. R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

22. ROBAN - Robust Analysis of Variance. Water Resource Systems Research Laboratory; Imperial College, London; University of Sussex, Centre for Environmental Research. Available from www.rsc.org/amc.

23. Analytical Methods Committee (2009) The Duplicate Method for the estimation of measurement uncertainty arising from sampling. *AMC technical briefs*, No 40. ISSN 1757-5958. Available at http://www.rsc.org/amc.

24. Lyn, J.A., Ramsey, M.H., Fussel, R.J. and Wood, R. (2003b) Measurement uncertainty from physical sample preparation: estimation including systematic error *Analyst*, 2003, 128 (11), 1391 – 1398

25. Lyn J.A., Ramsey M.H., Damant A., and Wood R. (2005a) Optimising Uncertainty in Physical Sample Preparation, Analyst, 130, 1507-1512

26. Thompson M. (2002) To assess the contribution of sampling procedures to the uncertainty of food analysis. Project report FD97/71 (FSA project FS 2918)

27. Thompson M., Willetts, P., Anderson S., Brereton P. and Wood R. (2002) Collaborative trials of the sampling of two foodstuffs, wheat and green coffee. Analyst, 127, 689-691.

28. Lyn J. A. (2003) PhD Thesis, “Optimising uncertainty from sampling and analysis of foods and environmental samples”, University of Sussex. ISNI: 0000 0001 3614 3415. EThOD ID: uk.bl.ethos.270732. URL: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.270732

29. Food Standards Agency Report 01055

30. Palestra (2008) PhD Thesis, “Estimation and achievement of optimal measurement uncertainty form the sampling of food and feed”, University of Sussex. ISNI: 0000 0001 3464 7193. EThOS ID: uk.bl.ethos.488580 . URL: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.488580

31. Food Standards Agency Report E01070, “Assessment of the practicality and usefulness of sampling proficiency tests in the food sector”

32. Vargas, E.A., Whitaker, T.B., Dos Santos, E.A., Slate, A.B., Lima, F.B., and Franca, R.C.A., 2006. Food Additives and Contaminants, 23 (1), 62-72

33. Johansson, AS; Whitaker, TB; Hagler, WM; Giesbrecht, FG; Young, JH; Bowman, DT (2000) J. AOAC Int., 83 (5): 1264-1269

34. Whitaker, T.B., (2006). Food Additives and Contaminants, 23 (1), 50-61.

35. Whitaker, T.B., Anders S.J. (2005) J. Food Protection, 68 (6), 1306-1313.

36. Whitaker, T.B., 2003. Food Control, 14, 233-237

37. Whitaker, TB; Dowell, FE, Hagler WM; Giesbrecht FG, Wu J; (1994). J. AOAC Int., 77(1), 107-116

38. Whitaker, TB; Hagler, WM; Giesbrecht, FG; Johansson, AS, (2000) J. AOAC Int., 83 (5): 1285-1292

39. Ozay, G ; Seyhan, F ; Yilmaz, A; Whitaker, TB; Slate, AB; Giesbrecht, F. (2006) J. AOAC International, 89 (4): 1004-1011.

40. Gy P M (1998) Sampling for Analytical Purposes John Wiley & Sons Ltd, Chichester

41. Food and Drug Administration: Investigations operations manual (2013) Sample Schedule 6: Aflatoxin sample sizes. http://www.fda.gov/ICECI/Inspections/IOM/default.htm

42. Commission Regulation (EC) No 401/2006 of 23 February 2006 laying down the methods of sampling and analysis for the official control of the levels of mycotoxins in foodstuffs. Official Journal of the European Union L 70/12, 9.3

43. CODEX STAN 193-1995 (Amd 2015) General Standard For Contaminants And Toxins In Food And Feed, pp27-33. Food and Agriculture Organisation (2015).

44. Ramsey, M.H., Thompson M. and Hale M. (1992) Journal of Geochemical Exploration, 44, 23-36

45. Thompson, M (2013), Analyt. Meth., 5, 4518-4519

Tables

Table 1: Summary of products and analytes evaluated experimentally in this report

Product	Analyte	No. Samples
Tuna	Histamine	8
Milk	Alkaline Phosphatase (ALP)	8
Lemonade	Sorbic Acid	10
Bread	Metals: Ca, Cu, Fe, Mg, Na, Zn	8
Bread	Fibre	8
Infant formula milk	Vitamin C	8
Bread	Vitamin B_1	8
Mince	Ash, Protein, Moisture, Fat	8

Table 4: Experimental variance components using classical and robust ANOVA

Method	Product	Analyte	Units	No. of Products	$s_{product}$	s_{sample}	$s_{analytical}$	Mean
Anova	Tuna	Histamine	µg/g	8	2.810	2.710	0.550	5.180
Roban	Tuna	Histamine	µg/g	8	1.450	2.310	0.510	4.540
Anova	Milk	ALP[Note 1]	mU/ml	8	16.300	6.200	5.109	49.780
Roban	Milk	ALP[Note 1]	mU/ml	8	16.100	7.260	5.060	48.770
Anova	Lemonade	Sorbic ac	µg/ml	10	34.180	2.120	3.000	145.290
Roban	Lemonade	Sorbic ac	µg/ml	10	38.760	2.000	2.750	145.290
Anova	Bread	Ca	mg/Kg	8	1011.020	780.910	67.060	2746.660
Roban	Bread	Ca	mg/Kg	8	1265.350	113.860	45.830	2728.560
Anova	Bread	Cu	mg/Kg	8	1.080	0.400	0.200	3.120
Roban	Bread	Cu	mg/Kg	8	1.240	0.330	0.190	3.120
Anova	Bread	Fe	mg/Kg	8	9.210	1.000	1.820	32.870
Roban	Bread	Fe	mg/Kg	8	9.390	1.480	1.430	32.640
Anova	Bread	Mg	mg/Kg	8	312.740	9.770	35.130	704.410
Roban	Bread	Mg	mg/Kg	8	354.530	26.550	13.920	704.420
Anova	Bread	Na	mg/Kg	8	1247.150	108.800	226.550	7026.660
Roban	Bread	Na	mg/Kg	8	983.910	199.840	128.330	7216.730
Anova	Bread	Zn	mg/Kg	8	9.980	0.530	0.880	20.150
Roban	Bread	Zn	mg/Kg	8	10.250	0.870	0.360	16.670
Anova	Bread	Fibre	g/100g	8	0.141	0.012	0.012	0.366
Roban	Bread	Fibre	g/100g	8	0.160	0.008	0.012	0.366
Anova	Milk[Note 2]	Vit C	mg/100ml	8	1.340	0.610	0.720	11.960
Roban	Milk[Note 2]	Vit C	mg/100ml	8	1.550	0.300	0.760	11.940
Anova	Bread	Vit B1	µg/g	8	0.880	0.000	0.450	3.510
Roban	Bread	Vit B1	µg/g	8	0.730	0.000	0.550	3.390
Anova	Mince	Fat	g/100g	8	6.575	3.744	0.533	10.949
Roban	Mince	Fat	g/100g	8	4.319	3.658	0.496	9.760
Anova	Mince	Protein	g/100g	8	1.330	0.000	1.581	19.842
Roban	Mince	Protein	g/100g	8	1.098	0.354	1.146	20.041
Anova	Mince	Ash	g/100g	8	0.0765	0.0672	0.0762	0.984
Roban	Mince	Ash	g/100g	8	0.0894	0.0000	0.0495	0.992
Anova	Mince	Moisture	g/100g	8	5.431	2.350	0.606	66.730
Roban	Mince	Moisture	g/100g	8	4.367	2.163	0.607	67.478

Note 1: ALP = alkaline phosphatase.

Note 2: Infant formula milk

Table 4a: Experimental variance components using classical and robust ANOVA - via log transform[Note 1]

Method	Product	Analyte	Units	No. of Products	$s_{product}$	s_{sample}	$s_{analytical}$	Mean
Anova	Tuna	Histamine	µg/g	8	2.159	2.150	0.555	4.217
Roban	Tuna	Histamine	µg/g	8	1.716	1.897	0.545	3.992
Anova	Milk	ALP[Note 2]	mU/ml	8	15.586	7.291	4.363	47.024
Roban	Milk	ALP[Note 2]	mU/ml	8	17.848	7.773	4.516	46.910
Anova	Lemonade	Sorbic ac	µg/ml	10	36.527	2.023	2.993	141.451
Roban	Lemonade	Sorbic ac	µg/ml	10	41.962	2.150	2.688	141.451
Anova	Bread	Ca	mg/Kg	8	1288.165	446.613	81.154	2491.722
Roban	Bread	Ca	mg/Kg	8	1560.309	113.310	46.366	2505.872
Anova	Bread	Cu	mg/Kg	8	1.254	0.342	0.228	2.905
Roban	Bread	Cu	mg/Kg	8	1.487	0.305	0.208	2.905
Anova	Bread	Fe	mg/Kg	8	9.802	1.210	1.580	31.661
Roban	Bread	Fe	mg/Kg	8	9.144	1.066	1.552	32.278
Anova	Bread	Mg	mg/Kg	8	360.034	16.764	21.325	640.612
Roban	Bread	Mg	mg/Kg	8	428.408	23.949	12.641	640.612
Anova	Bread	Na	mg/Kg	8	1457.713	101.369	205.456	6907.117
Roban	Bread	Na	mg/Kg	8	926.456	186.987	131.823	7188.053
Anova	Bread	Zn	mg/Kg	8	10.750	0.493	0.539	18.100
Roban	Bread	Zn	mg/Kg	8	12.847	0.638	0.365	18.100
Anova	Bread	Fibre	g/100g	8	0.172	0.007	0.011	0.338
Roban	Bread	Fibre	g/100g	8	0.202	0.008	0.012	0.339
Anova	Milk[Note 1]	Vit C	mg/100ml	8	1.343	0.538	0.653	11.865
Roban	Milk[Note 1]	Vit C	mg/100ml	8	1.565	0.378	0.733	11.865
Anova	Bread	Vit B1	µg/g	8	0.831	0.000	0.501	3.394
Roban	Bread	Vit B1	µg/g	8	0.764	0.000	0.501	3.329
Anova	Mince	Fat	g/100g	8	5.724	4.811	0.586	8.893
Roban	Mince	Fat	g/100g	8	4.522	5.478	0.604	8.758
Anova	Mince	Protein	g/100g	8	1.396	0.000	1.587	19.744
Roban	Mince	Protein	g/100g	8	1.043	0.356	1.213	19.982
Anova	Mince	Ash	g/100g	8	0.080	0.063	0.085	0.975
Roban	Mince	Ash	g/100g	8	0.088	0.000	0.048	0.986
Anova	Mince	Moisture	g/100g	8	5.738	2.381	0.636	66.489

Roban	Mince	Moisture	g/100g	8	4.352	2.151	0.666	67.388

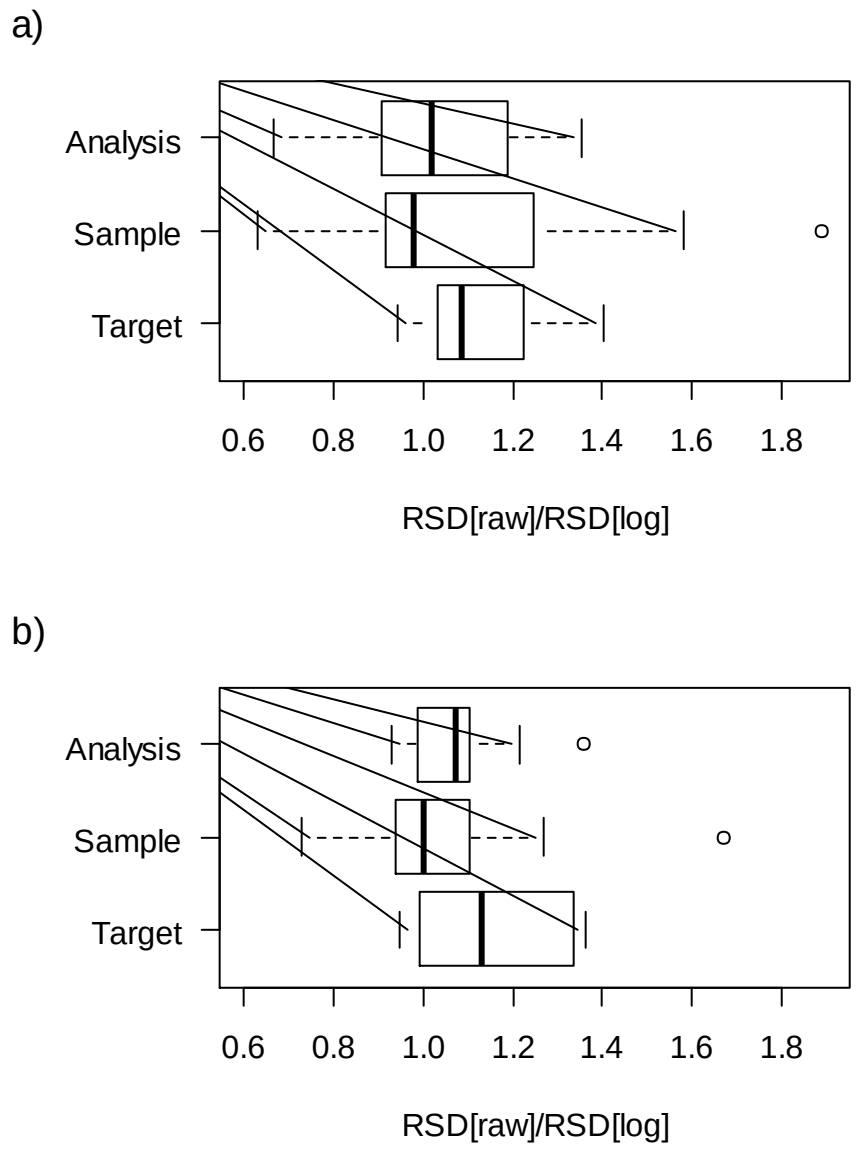
Note 1: The results were obtained by log-transformation of the raw data, treatment by the ANOVA method indicated, and conversion to the concentration scale using equation 1.

Note 2: ALP = alkaline phosphatase.

Note 3: Infant formula milk

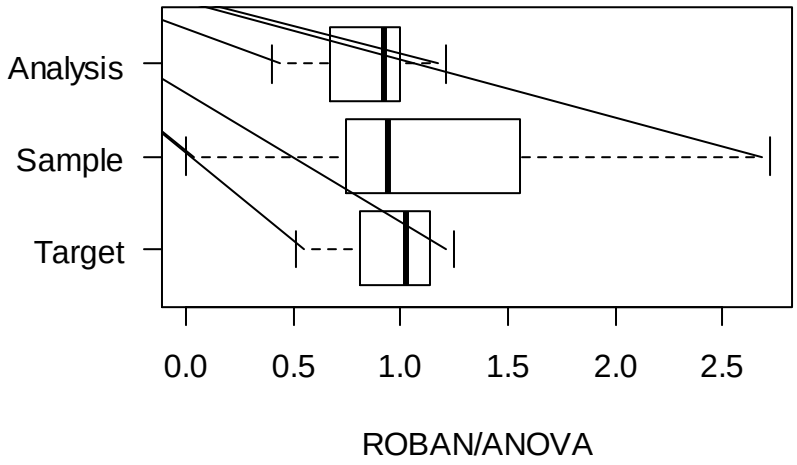
Figures

Figure 1: Effect of log-transformation



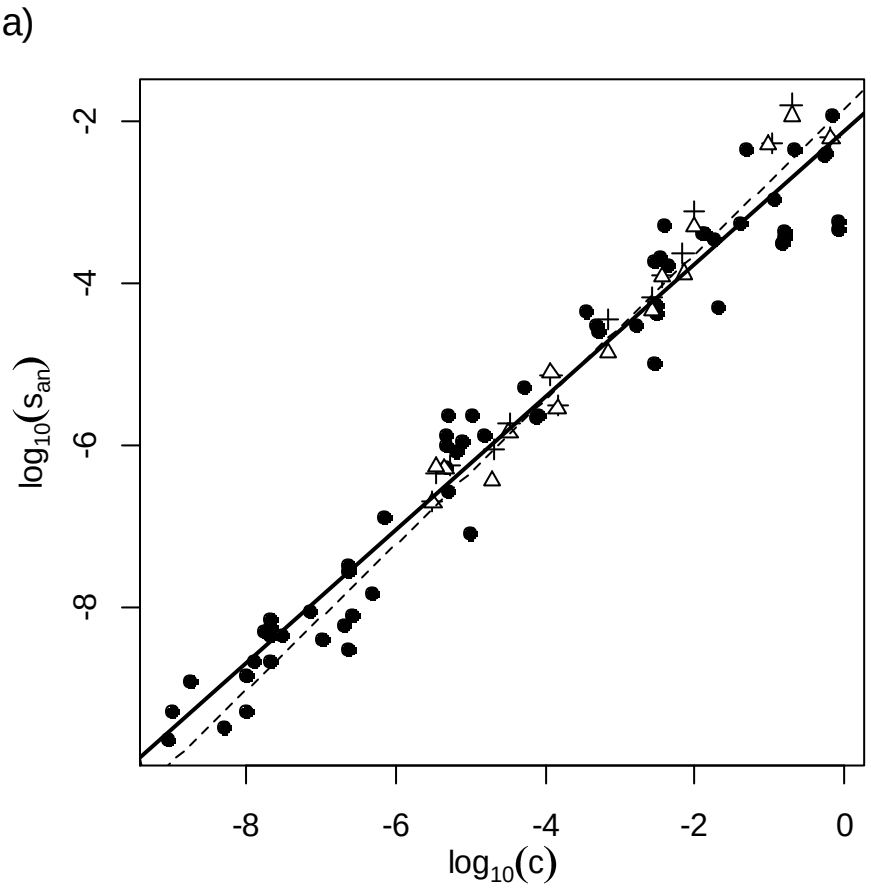
The figure shows the ratio of relative standard deviations for analysis, sampling and target-to-target (product) variation calculated using a) raw data and b) ROBAN. The ratio shown is for the raw data RSD divided by the RSD calculated from the log-transformed data. Values for which both RSDs were zero have been omitted; there were no cases in which log-transformation caused an estimated standard deviation to change from zero to a nonzero value or vice versa.

Figure 2: Robust estimates compared with classical estimates

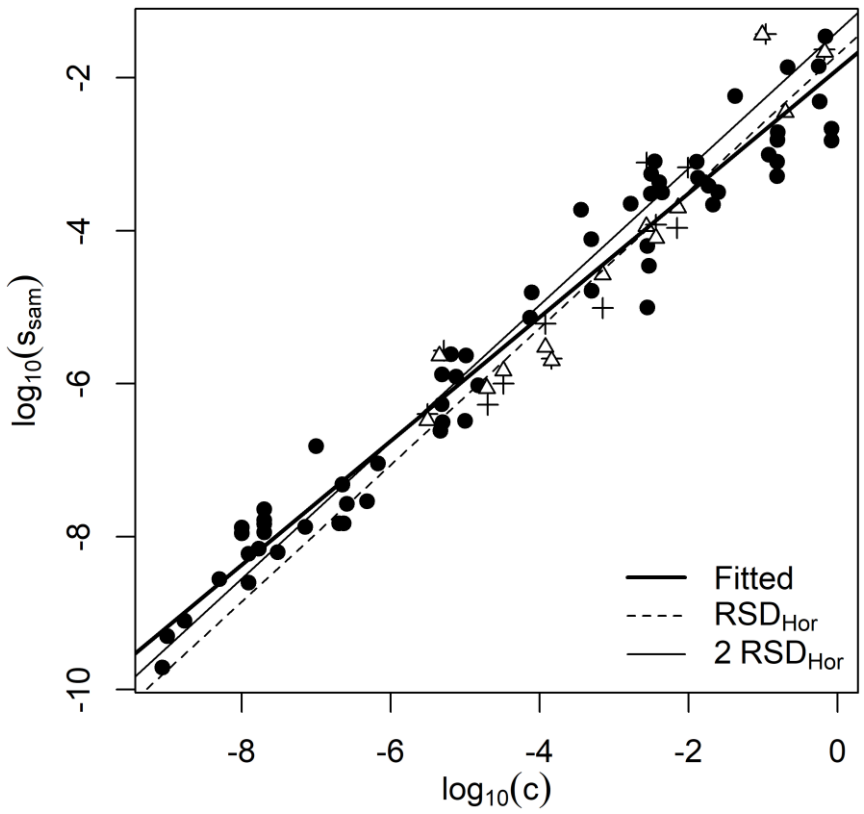


The figure shows the distribution of ratios of robust estimates to classical estimates of the respective standard deviations. One outlier arising from a classical sampling variance estimate of zero has been omitted.

Figure 3: Analytical and Sampling standard deviation vs concentration

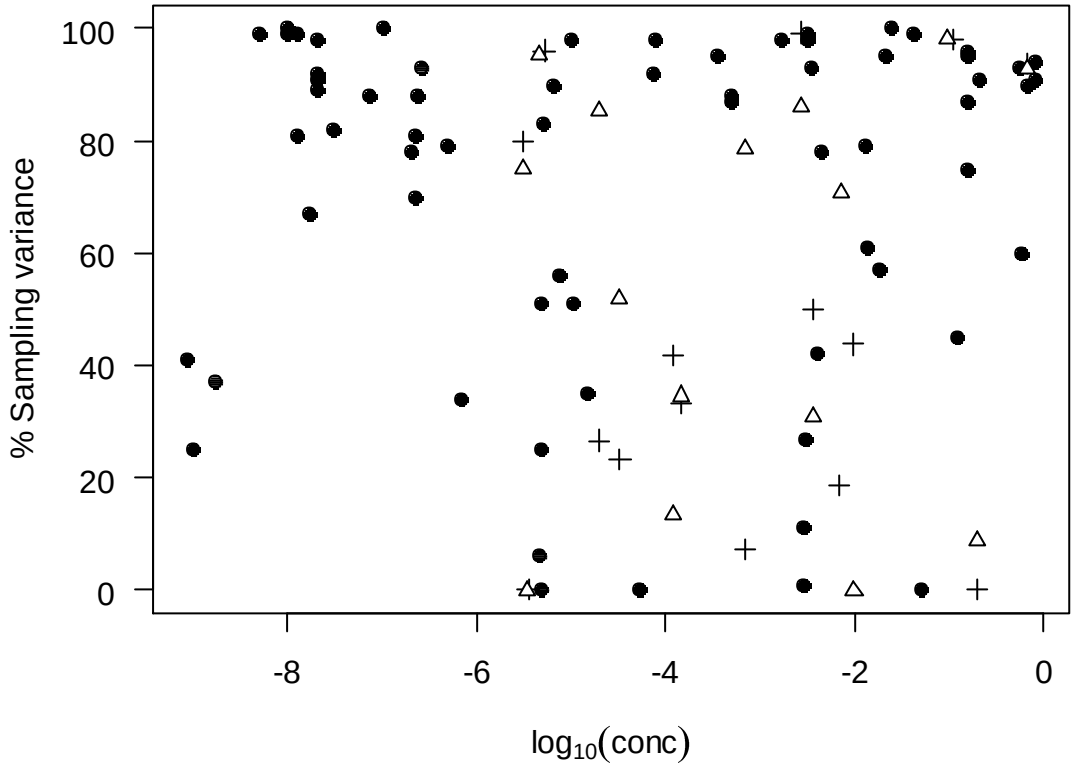


b)



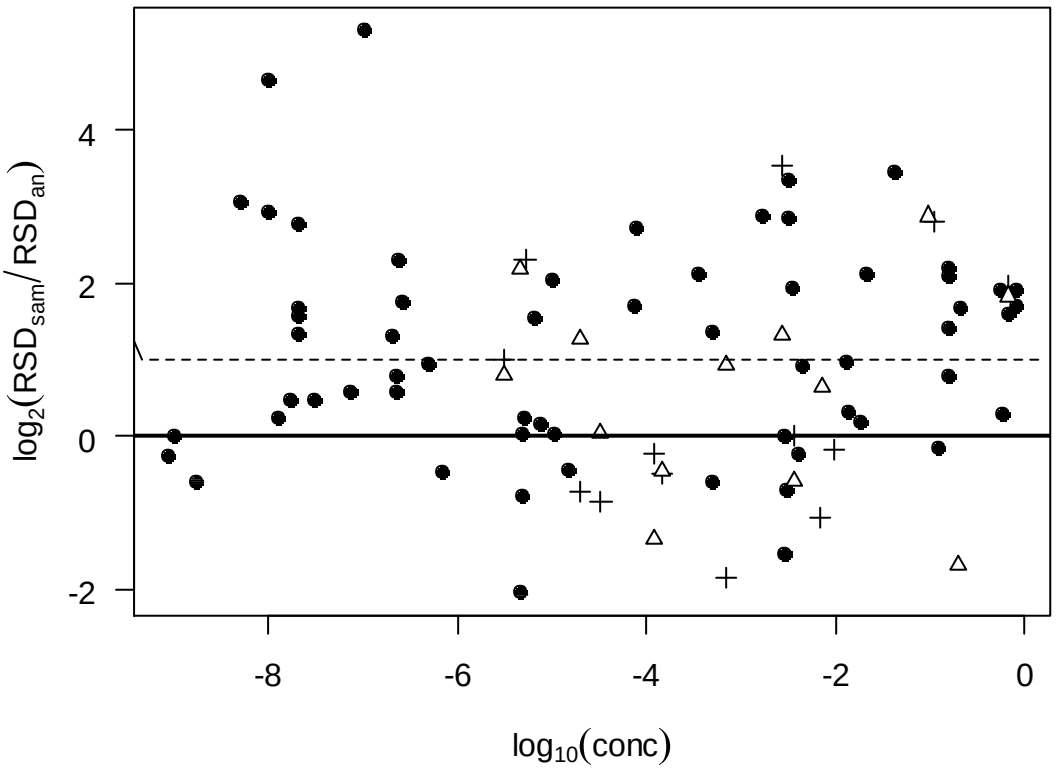
The figure plots a) $\log_{10}(s_{an})$ vs $\log_{10}(c)$; b) $\log_{10}(s_{sam})$ vs $\log_{10}(c)$; where s and c are standard deviation (in units of mass fraction) and mass fraction respectively. Filled circles are values drawn from the literature and summarised in Table 3; triangles are RSDs calculated from robust (ROBAN) estimates from Table 4 and crosses are classical ANOVA estimates from Table 4. The heavy solid line is the linear least squares regression fits to the literature and robust estimates. The dashed line is the Horwitz standard deviation; the light solid line twice the Horwitz standard deviation.

Figure 4: Proportion of variance due to sampling, by concentration



The figure shows the proportion of measurement variance $s^2_{\text{sam}}/\left(s^2_{\text{sam}} + s^2_{\text{an}}\right)$ attributable to sampling variance. For symbols see Figure 3.

Figure 5: sampling/analytical uncertainty ratio by concentration



The figure shows the ratio RSD_{sam}/RSD_{an} plotted on a log scale. For symbols see Figure 3. The solid line is at a log ratio of 0 (representing equal sampling and analysis); the dashed line is at the calculated mean log ratio of 1.00, corresponding to a 2:1 ratio of sampling standard deviation to analytical standard deviation.