Use of an ‘excess-variance’ approach for the estimation of a key comparison reference value, associated standard uncertainty and degrees of equivalence for CCQM key comparison data

Maurice Cox, Peter Harris
National Physical Laboratory, UK

Steve Ellison
LGC Ltd, UK

Executive summary

This document illustrates the use of an ‘excess-variance’ approach for the statistical analysis of CCQM key comparison data that makes allowance for unexplained laboratory effects.

In ‘excess-variance’ approaches, the variances (squared standard uncertainties) provided by participating laboratories are augmented by an additional variance, common to all laboratories, which is regarded as characterizing unexplained laboratory effects. The reported variances are combined with an estimate of the excess variance to provide weights that are used to estimate the key comparison reference value (KCRV).

The uncertainty associated with this estimate takes account of both the laboratory uncertainties and the estimated excess variance. For mutually consistent data the approach reduces to the classical weighted mean.

Degrees of equivalence are determined accordingly.

The present document demonstrates the general approach by reference to a particular implementation, the DerSimonian-Laird method, applied to some illustrative examples of CCQM key comparisons.

Supporting software can be made available to carry out the calculations should CCQM agree with the content of this document.

1 Introduction

This document is concerned with a CCQM key comparison in which each laboratory independently provides a measured value and an associated standard uncertainty for one or more measurands. Technical review of the data by the relevant working group identifies those laboratory data that are considered appropriate for use in estimating the KCRV. For many studies, this review results in a set of values and uncertainties that include no serious outlying values, but that may show a dispersion of data that cannot be fully explained by the reported uncertainties. While a variety of methods can be applied to estimate a KCRV and its associated standard uncertainty in these circumstances, one class of estimators that appears promising is a class referred to in this document as 'excess-variance' approaches.

Since the situation described—excess dispersion with no evidence of serious outlying values—is one of the general scenarios identified in draft guidance document CCQM-10-03 [2], the present document can be considered a specialized case within CCQM-10-03. Document CCQM-10-03 remains applicable as a background document to be used for those sets of key comparison data for which it is judged inappropriate to apply the approach here.

In 'excess-variance' approaches, the variances (squared standard uncertainties) provided by participating laboratories are augmented by an additional variance, assumed to be common to all laboratories, which is regarded as characterizing unexplained laboratory effects. The reported variances are combined with the estimated excess variance to provide weights used to estimate the KCRV. The standard uncertainty associated with the KCRV takes account of both the laboratory variances and the estimated excess variance. For mutually consistent data the approach reduces to the classical weighted mean.

The present document uses a particular estimator in this class, the DerSimonian-Laird estimator, to illustrate the use of such an estimator. The approach draws on principles given in document CCQM-09-03 [1]. Due account is taken of the GUM [7].

Though the approach illustrated here is straightforward, it is not intended to replace critical evaluation of the data and appropriate use of measurement and statistical expertise, nor should it be regarded as prescriptive. However, the general approach is likely to be applicable to data for many CCQM key comparisons. Such an approach would be compatible with the remit of the CCQM KCRV WG in attempting to harmonize as much as possible the calculation of the KCRV and its associated uncertainty

Section 2 gives the suggested approach and Section 3 provides supporting technical information. The strengths and weaknesses of the approach based on the DerSimonian-Laird estimator compared to the use of some other estimators are considered in Section 4. Section 5 makes concluding remarks. Annexes contain the rationale for the choice of estimator and illustrative examples.

2 Model approach

The suggested approach is based on the use of weighted means with an appropriate choice of weights. An interlaboratory variance is estimated, which is used to augment the above laboratory variances and hence adjust the weights. This additional variance term is chosen so that the combination of reported standard uncertainties and estimated additional variance is sufficient to account for the observed dispersion of values.

The appropriate weighted mean is taken as the KCRV, the standard uncertainty associated with the KCRV is evaluated, and DoEs accordingly determined. The approach is described below as a sequence of steps in which the measured values and associated

standard uncertainties provided by the participating laboratories are denoted by x_i and u_i, respectively. Cited equations and formulae are given in Table 1.

The steps involved are as follows:

1. For purposes of the calculation (but not for reporting), replace the u_i by the original u_i augmented in quadrature by standard uncertainties relating to any well-quantified inhomogeneity or instability effects.
2. Carefully examine the participants' data, making use of graphical tools and any statistical tests (including, as appropriate, tests for inconsistency and outlying values), as an aid to identifying possible anomalies. The working group may exclude anomalous data from the calculation of the KCRV on technical grounds. (For a more detailed discussion see CCQM-10-03 [2].)
3. For the remaining data (x_i, u_i), $i = 1, \ldots, p$, use formulae (1) to (3) to determine a value λ for the interlaboratory variance and the DerSimonian-Laird (DL) mean x_{DL}. Evaluate the standard uncertainty $u(x_{\mathrm{DL}})$ associated with x_{DL} using formula (4).
4. Take x_{DL} and $u(x_{\mathrm{DL}})$ as the KCRV and its associated standard uncertainty.
5. Form the DoEs for the laboratories, using formula (5) or (6), as appropriate.

Note: The same general formulae apply to some other excess-variance estimators

Table 1. Formulae used in suggested approach

Parameter, etc.	Formula	Ref
Graybill-Deal mean	$x_{\mathrm{GD}} = \frac{1}{W_1}\sum_{i=1}^{p} w_i x_i, \quad w_i = 1/u_i^2, \ i = 1,\ldots,p, \qquad W_1 = \sum_{i=1}^{p} w_i.$	(1)
Interlaboratory variance	$\lambda = \max\left[0, \ \frac{\sum_{i=1}^{p} w_i (x_i - x_{\mathrm{GD}})^2 - p + 1}{W_1 - W_2 / W_1}\right], \quad W_2 = \sum_{i=1}^{p} w_i^2.$	(2)
DerSimonian-Laird mean x_{DL}	$x_{\mathrm{DL}} = \sum_{i=1}^{N} \tilde{w}_i x_i, \qquad \tilde{w}_i = \frac{(u_i^2 + \lambda)^{-1}}{\sum_{j=1}^{p} (u_j^2 + \lambda)^{-1}}.$	(3)
Standard uncertainty $u(x_{\mathrm{DL}})$	$u(x_{\mathrm{DL}}) = \left[\sum_{i=1}^{p} \tilde{w}_i^2 (x_i - x_{\mathrm{DL}})^2 \big/ (1 - \tilde{w}_i)\right]^{1/2}.$	(4)
DoEs (d_i, $2u(d_i)$)	For laboratory data used for KCRV calculation, $d_i = x_i - x_{\mathrm{DL}}, \qquad u^2(d_i) = u_i^2 + \lambda \ - u^2(x_{\mathrm{DL}}).$	(5)
	For data not used for KCRV calculation, $d_i = x_i - x_{\mathrm{DL}}, \qquad u^2(d_i) = u_i^2 + \lambda \ + u^2(x_{\mathrm{DL}}).$	(6)

3 Supporting technical information

The technical information here applies generally to excess-variance estimators and in particular to the step-by-step approach in Section 2.

1. In Step 1, samples measured by the laboratories correspond to different measurands. Heterogeneity or instability causes differences between a common measurand and the measurands corresponding to the samples. The augmented uncertainties take into account the need to relate the measured values to a common measurand.
2. Although ideally interlaboratory effects should have a scientific explanation (heterogeneity or instability of samples, for instance), such an explanation is often not forthcoming. Another possible cause of inconsistency may be under-stated uncertainties u_i for some or all comparison participants.
3. For a mutually consistent data set, the interlaboratory variance λ is taken as zero. In a case of inconsistency, as λ is increased, the influence of those laboratories that provide the smallest uncertainties is reduced. For extremely inconsistent data sets, λ becomes very large compared with the u_i^2 in order to achieve consistency, and the DL mean will approach the arithmetic mean. There is a smooth transition from the classical weighted mean to the arithmetic mean as the data inconsistency increases.
4. The standard uncertainty $u(x_{\mathrm{DL}})$ associated with x_{DL}, determined in Step 4, is compatible with the augmented standard uncertainties $(u_i^2 + \lambda)^{1/2}$ associated with the measured values x_i. These standard uncertainties can be interpreted as *posterior* standard uncertainties associated with the x_i. These posterior standard uncertainties are necessary to overcome inconsistency in the data set (x_i, u_i), $i = 1, \ldots, p$ [14].
5. The DoE for the ith laboratory (Steps 5) consists of a value component d_i and an uncertainty component $U(d_i)$ (at the 95 % level of confidence). The uncertainty component is expressed as $U(d_i) = 2u(d_i)$, where $u(d_i)$ is the standard uncertainty associated with d_i, under a normality assumption.
6. Formulae (5) and (6), used for calculating DoE uncertainties, are compatible with the posterior uncertainties associated with the x_i. In particular, $u(d_i)$ obtained from formula (5) does not reflect only the measurement uncertainty stated by laboratory i. Formula (5) is chosen to be fully consistent with the formal definition of the DoE given in the Technical Annexe to the CIPM MRA — that is, it is the standard uncertainty associated with the difference between the reported measured value and the KCRV after taking account of all terms in the statistical model underpinning the KCRV.

4 Advantages and disadvantages of excess-variance estimators

Detailed rationale for choosing the DerSimonian-Laird estimator for this illustration is given in Annex A. Other excess-variance estimators have, however, been proposed for the same general problem, in particular the Mandel-Paule estimate and variations on maximum likelihood estimation. Some of the practical advantages and disadvantages are listed below.

1. The DerSimonian-Laird approach has the advantage that it can be implemented in a single simple calculation and does not require iterative solution. The principal disadvantage is that it is a poorer approximation than those provided by more sophisticated iterative methods. It does not take account of finite degrees of freedom in the individual reported uncertainties, and because it is equivalent to the Graybill-Deal estimator (the classical weighted mean) [5] when the dispersion is fully accounted for

by the reported uncertainties, may substantially understate the uncertainty associated with the KCRV when degrees of freedom are not very large.

2. The Mandel-Paule approach has the advantage of modest complexity and consistency under normality assumptions, resulting in a better estimate of the excess variance than the DerSimonian-Laird approach when all degrees of freedom are large and normality can be assumed. The disadvantages include a need for iterative solution, and, like the DerSimonian-Laird method, lack of treatment of degrees of freedom in the individual uncertainties and the same estimate as the Graybill-Deal mean for apparently consistent data.
3. Maximum likelihood estimation is capable of taking account of finite degrees of freedom in reported uncertainties and restricted maximum likelihood estimation provides minimally biased variance estimates. The principal disadvantage is comparative complexity in implementation, resulting in few currently available software implementations.

In practice, however, all three often produce very similar estimates for a given data set (as is the case for the examples here).

An apparent disadvantage of all excess-variance approaches is the drastic effect at times on degrees of equivalence. Because the same estimate of excess variance is used for all laboratories, laboratories with smaller reported uncertainties will generally be allocated DoE uncertainties that are much larger than their reported uncertainty. This statement is a consequence of a) the need to incorporate an additional variance to provide sensible estimates of KCRV and its associated uncertainty when the reported uncertainties do not account for the observed dispersion, and b) a strict interpretation of the definition of the degree of equivalence given in the MRA. One of the examples here shows DoEs formed with and without this interlaboratory variance.

5 Concluding remarks

An approach for the statistical analysis of CCQM key comparison data that makes allowance for an unexplained interlaboratory effect is suggested for providing a KCRV, its associated standard uncertainty and DoEs. It applies when the measured values provided by the participants in the key comparison are mutually independent.

Such an approach offers a reasonable compromise when the data taken as a whole cannot be explained by the standard uncertainties provided by the comparison participants. The specific approach suggested is based on the DerSimonian-Laird mean [9][10].

Implementation details of the approach are provided and supporting software can be made available. Illustrative examples of CCQM key comparisons are given.

The approach responds to the remit of the CCQM KCRV WG for greater harmonization in key comparison data analysis. It is suggested that CCQM WGs test this approach alongside their ongoing customized data analysis. Some members of CCQM prefer to treat each key comparison on an individual basis rather than use some prescriptive approach. With adequate access to professional statistical advice, that attitude is commendable.

The introduction of a variance relating to interlaboratory effects assures consistency of the key comparison data with the KCRV, but with current interpretation of the CIPM MRA has the consequence that all laboratories have DoE uncertainties that guarantee mutual consistency and that may be considerably larger than reported

uncertainties. It is not yet clear whether alternative calculations of $u(d_i)$ are defensible given the applicable assumptions.

Annex A. Rationale for the choice of estimator

When key comparison data are mutually inconsistent, there is a need to force consistency in order to provide a) a meaningful KCRV and value components of the DoEs, and b) credible uncertainties associated with these values. A considerable number of relevant papers in the statistical and metrological literature exist that concentrate on excess-variance estimators for this purpose [3][4][8][9][10][12][13][15]. Also see CCQM-10-03 [2].

Weighted mean statistics (where the location estimate is expressed as a linear combination of the x_i) are mainly used for this purpose. See the review by Rukhin [11]. Such estimators include Mandel and Paule (MP), Vangel and Rukhin (VR), DerSimonian and Laird (DL) and the maximum-likelihood estimate (MLE).

Rukhin [11] gives results of simulations that indicate that MP consistently outperforms MLE and its variants. Moreover, an MLE solution is governed by a non-linear algebraic equation, which generally involves a non-monotonic function [13][15], for which there is a possibility of a non-unique solution. MP is also given by the solution of a non-linear algebraic equation. However, the function involved is monotonic and convex [6], with the result that a unique solution always exists. This solution can straightforwardly be determined by an algorithm with guaranteed convergence [8][15]. Approximations to MP exist [3][4][8][11][12][13], with several, particularly DL [4], being effective. Rukhin's simulations [11] demonstrate that DL performs almost as well as MP.

DL has the advantage that it is a direct method, requiring implementation of only a small number of formulae. There could conceivably be cases, not covered by Rukhin's simulations [11], where DL does not perform so well. As mentioned, software for an implementation of DL can be made available.

Annex C. Illustrative examples

Presented for each of three examples of CCQM key comparisons is (a) a graph of the data (for each laboratory the measured value and ±1 standard uncertainty associated with that value, with a green bar if used in obtaining the KCRV and red otherwise), the KCRV given by the MP estimate (black horizontal line) and ±1 standard uncertainties (blue horizontal lines) associated with that estimate, (b) a table giving the MLE, MP and DL estimates and the associated standard uncertainties, and the excess standard uncertainty in each case, and (c) a graph of the DoEs. The third example show the DoEs computed with and without the estimated interlaboratory variance.

K61 Plasmid DNA in solution (fg μl^{-1})

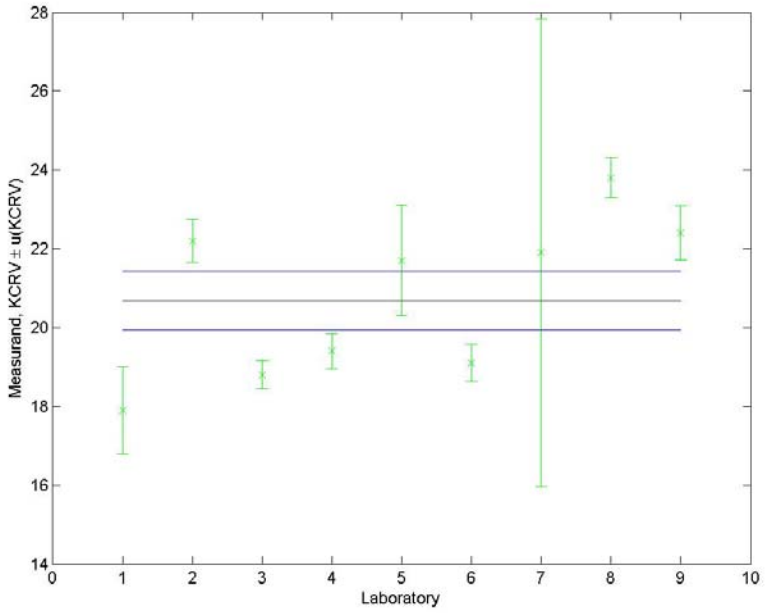

Figure 1. Laboratory data and KCRV (DL) for K61 Plasmid DNA in solution

Table 2. MLE, MP and DL estimates, standard uncertainties and excess standard uncertainty for K61

Estimator	Estimate	Std unc	Excess std unc
MLE	20.68	0.70	1.86
MP	20.69	0.70	1.86
DL	20.68	0.75	2.04

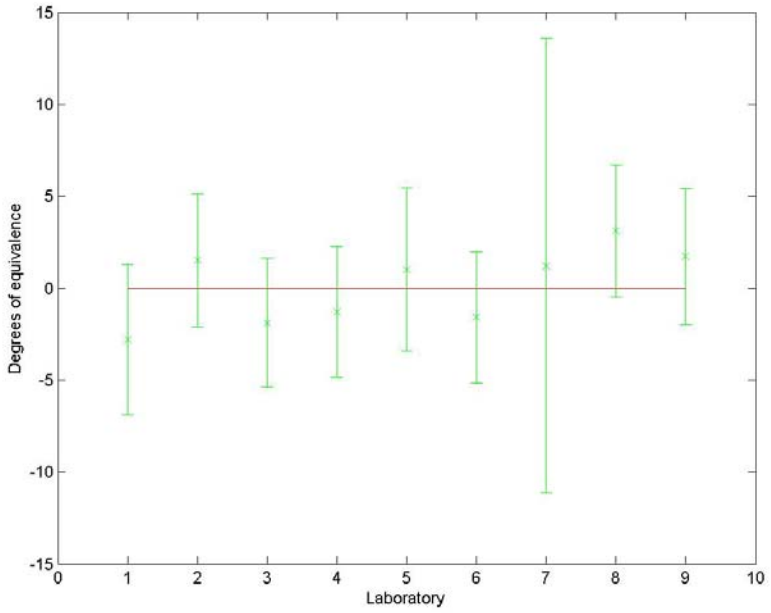

Figure 2. DoEs for K61.

K5 pp'-DDE in fish oil (μg/g)

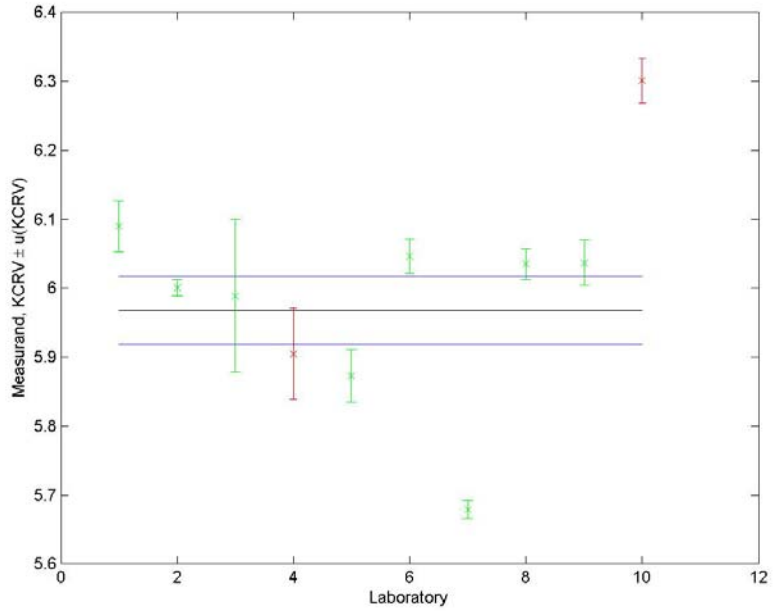

Figure 3. Laboratory data and KCRV (DL) for K5 pp'-DDE in fish oil

Table 3. MLE, MP and DL estimates, standard uncertainties and excess standard uncertainty for K5

Estimator	Estimate	Std unc	Excess std unc
MLE	5.97	0.05	0.13
MP	5.97	0.05	0.13
DL	5.97	0.05	0.18

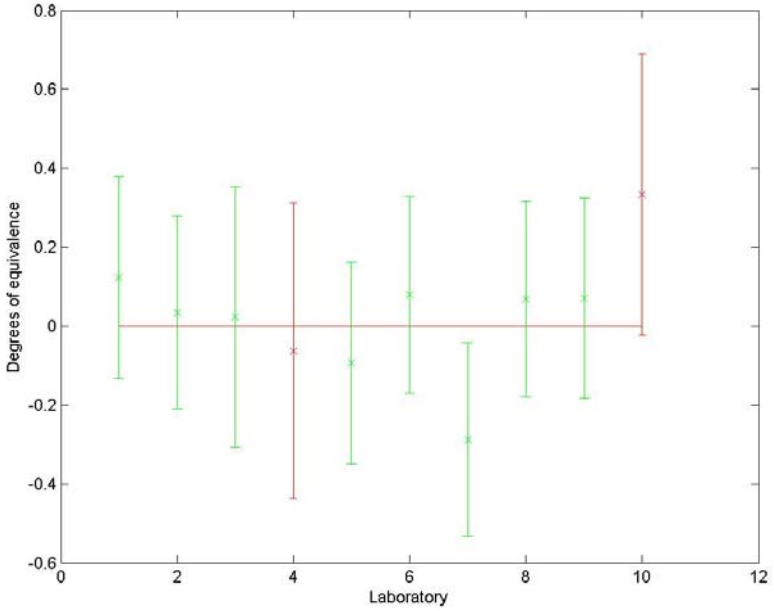

Figure 4. DoEs for K5.

K25 PCB170 in Sediment (ng g^{-1})

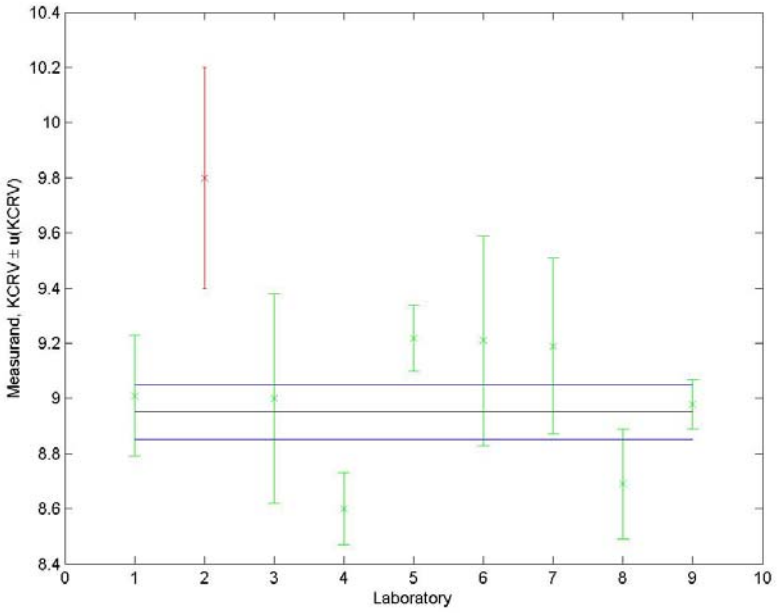

Figure 5. Laboratory data and KCRV (DL) for K25 PCB170 in Sediment

Table 4. MLE, MP and DL estimates, standard uncertainties and excess standard uncertainty for K25 PCB170 in Sediment

Estimator	Estimate	Std unc	Excess std unc
MLE	8.95	0.09	0.17
MP	8.95	0.09	0.15
DL	8.95	0.10	0.18

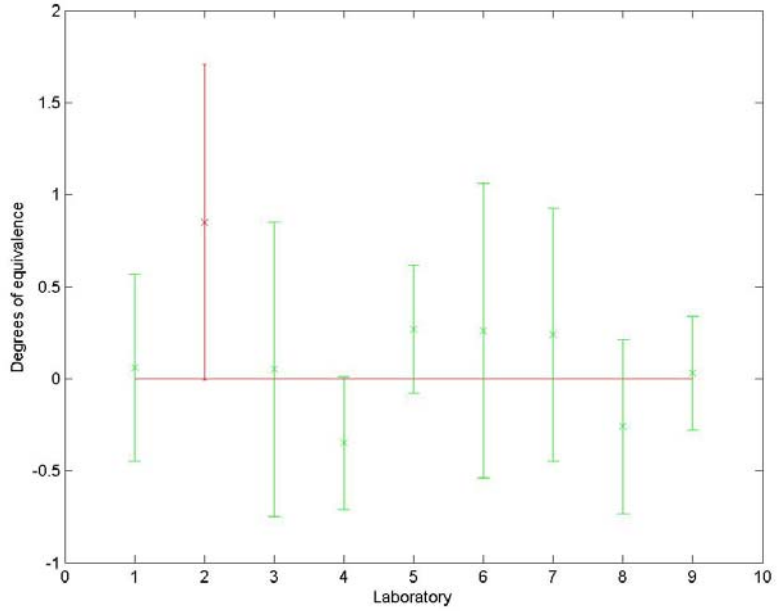

Figure 6. DoEs for K25 PCB170 in Sediment, including interlaboratory variance.

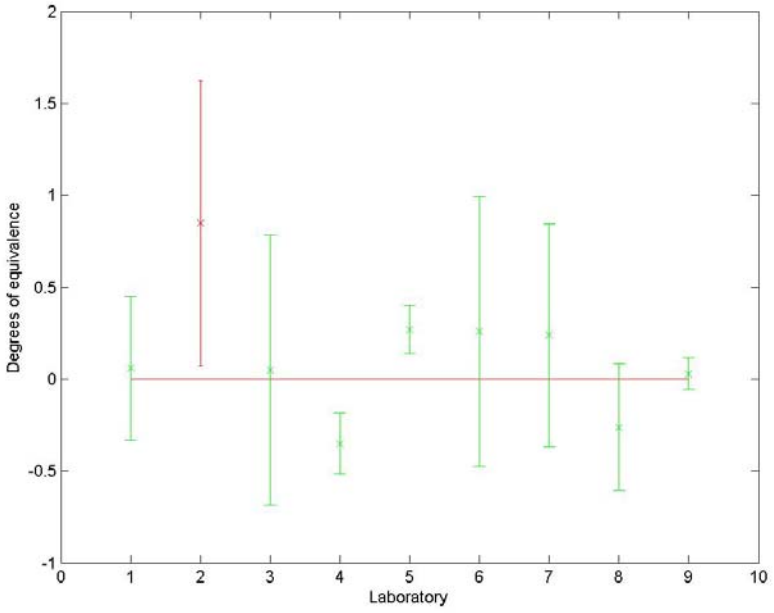

Figure 7. DoEs for K25 PCB170 in Sediment, excluding interlaboratory variance.

References

[1] CCQM-09-03. Data evaluation principles for CCQM key comparisons. CCQM, 2009.

[2] CCQM-10-03. CCQM guidance note: estimation of a consensus KCRV and associated degrees of equivalence. CCQM, 2010.

[3] R. DerSimonian and R. Kacker, Random-effects model for meta-analysis of clinical trials: An update, *Contemporary Clinical Trials*, 2007, **28**, 105–114.

[4] R. DerSimonian and N. Laird, Meta-analysis in clinical trials, *Controlled Clinical Trials*, 1986, **7**, 177–188.

[5] F. A. Graybill and R. B. Deal, Combining unbiased estimates. *Biometrics*, 1959, **15**, 543–550.

[6] H. K. Iyer, C. M. J. Wang and T. Mathew, Models and confidence intervals for true values in interlaboratory trials, *J. Amer. Statist. Assoc.*, 2004, **99**, 1060–1071.

[7] JCGM 100:2008. Guide to the expression of uncertainty in measurement (GUM).

[8] R. N. Kacker, Combining information from interlaboratory evaluations using a random effects model, *Metrologia*, 2004, **41**, 132–136.

[9] J. Mandel and R. Paule, Interlaboratory evaluation of a material with unequal number of replicates, *Anal. Chemistry*, 1970, **42**, 1194–1197.

[10] R. Paule and J. Mandel, Consensus values and weighting Factors, *J. Research Natl. Bureau Standards*, 1982, **87**, 377–385.

[11] A. L. Rukhin, Weighted means statistics in interlaboratory studies, *Metrologia*, 2009, **46**, 323–331.

[12] A. L. Rukhin and N. Sedransk, Statistics in metrology: international key comparisons and interlaboratory studies, *J. Data Sciences*, 2007, **5**, 393–412.

[13] L. Rukhin and M. G. Vangel, Estimation of a common mean and weighted means statistics, *J. Amer. Statist. Assoc.*, 1998, **93**, 303–308.

[14] K. Weise and W. Wöger, Removing model and data non-conformity in measurement evaluation, *Meas. Sci. Technol*., 2000, **11**, 1649–1658.

[15] R. Willink, Statistical determination of a comparison reference value using hidden errors, *Metrologia*, 2002, **39**, 343–354.