

Chemometrics and Intelligent Laboratory Systems 53 (2000) 37-55

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

A new model for the inference of population characteristics from experimental data using uncertainties. Application to interlaboratory studies

Wim P. Cofino^{a,*}, Ivo H.M. van Stokkum^b, David E. Wells^c, Freek Ariese^d, Jan-Willem M. Wegener^d, Reneé A.L. Peerboom^d

^a Ministry of Transport, Public Works and Water Management, Directorate General for Public Works and Water Management, Institute for Inland Water Management and Waste Water Treatment RIZA, P.O. Box 17, 8200 AA Lelystad, Netherlands

^b Department of Physics Applied Computer Science, Division of Physics and Astronomy, Faculty of Sciences, Vrije Universiteit,

De Boelelaan 1081, 1081 HV Amsterdam, Netherlands

^c Fisheries Research Station, Marine Laboratory, Victoria Road, Aberdeen, UK

^d Institute for Environmental Studies, Vrije Universiteit, De Boelelaan 1115, 1081 HV Amsterdam, Netherlands

Received 20 December 1999; accepted 3 August 2000

Abstract

A new model to make inferences about population characteristics from experimental datasets is presented. It derives concepts and procedures from quantum chemistry. The model uses the observed values and the uncertainty estimates thereof. It provides the different modes of the distribution and for each mode the expectation value, the standard deviation and a percentage indicating the fraction of observations encompassed. An implementation of the model that does not require uncertainty estimates is provided too. In this paper, the model is elaborated and applied to the evaluation of interlaboratory studies. It has, however, a much wider generic application. It is demonstrated that the model can cope with asymmetric, strongly tailing and multimodal distributions and that it is superior to existing techniques (e.g. ISO 5725, robust statistics). © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Population characteristics; Multimodality; Uncertainty; Robust statistics; Interlaboratory studies

1. Introduction

Statistical inferences of population characteristics from experimentally obtained datasets, e.g. measures for location and dispersion, need to be made in many fields of science. The calculation of such characteristics is frequently complicated by the nature of the distribution underlying the data. A range of procedures has been developed to deal with this problem, such as outlier tests based on assumptions about the distribution functions [1] and nonparametric algorithms like robust statistics [2] and bootstrapping [3]. Better procedures are, however, still desirable.

A new model is presented in this paper which derives concepts and practices from quantum chem-

^{*} Corresponding author.

^{0169-7439/00/\$ -} see front matter © 2000 Elsevier Science B.V. All rights reserved. PII: S0169-7439(00)00093-9

istry, in particular wavefunctions and matrix algebra. No physical laws can be used to derive wavefunctions underlying the experimental data encountered in practice. The square of a wavefunction is a probability density function. Using an estimate for the probability density function of the measurement process concerned, a 'wavefunction' can be derived which underlies the observation and which comprises the observed value and its uncertainty. The wavefunctions of all individual observations constitute a basis set in which wavefunctions describing the properties of the total dataset can be constructed. The procedure outlined in this paper constructs the wavefunctions for the dataset as a whole as linear combinations of the wavefunctions of the observations. For each linear combination, which corresponds to a mode of the distribution of observations, the expectation value, the standard deviation and a percentage indicating the fraction of the observations encompassed is calculated. The first mode provides the population characteristics sought.

Although this model can be applied to a variety of situations, this paper will only be providing information for the evaluation of data from interlaboratory studies. Interlaboratory studies play an important role in analytical chemistry to assess the performance parameters of a method, to assess the performance of laboratories and to certify materials [4]. The design, conduct and evaluation of interlaboratory studies have received considerable interest in the past years. In close collaboration, IUPAC, AOAC and ISO have issued guidelines for method-performance studies [5] and for laboratory-performance studies [6]. The European Union has provided a detailed document describing how reference materials should be prepared and certification studies should be conducted in the framework of the Standards. Measurements and Testing Programme [7]. The Analytical Methods Committee of the UK-Royal Society of Chemistry (denoted as AMC in this paper) reported on the organisation and statistical assessment of proficiency tests [8]. ISO recently has published guides on the organisation of interlaboratory studies [9,10]. A common element throughout these standards and guidelines is the need to provide a statistical evaluation of interlaboratory studies.

Youden and Steiner [11] have provided a cornerstone with respect to the development of statistical procedures related to the evaluation of method-performance studies. The work of these authors has evolved in time in to AOAC guidelines [12], and the standard ISO 5725 [13,14]. Basically, a series of outlier tests are applied to the data after which an ANOVA based procedure is adopted to obtain the expectation value and the reproducibility and repeatability of the method under study.

The procedure is based on the assumption that all of the within-laboratory variances are equal and normally distributed, and that a normal distribution underlies the between-laboratory variance. These assumptions seem acceptable for method-performance studies, in which participating laboratories experienced in the method tested, follow the same, single method protocol. In practice, however, these assumptions are frequently violated. The direct application of ISO 5725 for laboratory-performance studies is highly questionable. In such studies, laboratories are free to use their own methodology. In addition to difficulties related to the distribution of the laboratory means, there is a priori no reason to assume that the within-laboratory variances will be the same. Some researchers have used ISO 5725 to evaluate laboratory-performance studies after a transformation of the data with the aim to reduce the problems regarding deviations from normality and non-equal withinlaboratory variances. (e.g. Ref. [15]). A more fundamental criticism of ISO 5725 concerns the use of outlier tests [1].

In the past 10 years, robust statistics have been put forward as a suitable technique to analyse the data from interlaboratory studies [1,16-21]. The AMC and Lischer methods have been tested by different groups [22-24]. In studies on marine samples, it has been observed that many data distributions tend to have positive outliers owing to, for instance, calculation or transcription errors, errors with units or contamination [22]. Such data normally form up to 5% and, in some cases, as much as 10% of the total amount of data. The robust statistics of AMC and Lischer [26] appear to have difficulties in coping with these asymmetrical distributions, which is manifested as inflated values for the robust means and robust standard deviations. In practice, this problem can be circumvented by inspecting the data and removing 'obvious outliers' prior to the statistical evaluation [22], but this brings an element of subjectiveness which the

robust statistics intended to avoid. Other robust methods are also reported in the literature, each with its own advantages and disadvantages (e.g. Refs. [20,21,25,26]).

Lischer [26] has listed the criteria for procedures to evaluate interlaboratory studies [28]. These criteria include a good efficiency for an ideal model (i.e. the normal distribution), a small impact of minor deviations from the ideal model on the outcome of the calculations and a high resilience towards large deviations. The model presented in this paper meets the criteria stated by Lischer [26]. The method is, however, more robust for asymmetrical distributions and includes numerical and graphical means to assess data distributions. A particular feature of this procedure is that explicit use is made of the uncertainty of the individual laboratory data to establish the expectation value and its uncertainty.

The mathematical approach is provided and explained in Section 2. Thereafter, the procedure is illustrated with simulated datasets and with real datasets derived from international laboratory-performance studies.

2. Design and implementation of the model

In the model proposed in this study, so called measurement functions are defined. The measurement functions are written in the Dirac notation [27] as $|\varphi_i(x)\rangle$ (where x is the measurand), which is abbreviated to $|\varphi_i\rangle$, for laboratory i and as $|\Psi\rangle$ for the interlaboratory study. The laboratory measurement function $|\varphi_i\rangle$ is set equal to the square root of a probability density function which encompasses the expectation value μ_i and uncertainty of the measurement of laboratory i. Likewise, Ψ^2 represents a probability density function which comprises the consensus value and uncertainty of the interlaboratory study.

The measurement function has by definition the following properties

$$\langle \varphi_i | \varphi_i \rangle = 1, \quad \langle \varphi_i | \varphi_j \rangle = S_{ij}, 0 \le S_{ij} \le 1,$$
 (1)

$$\langle \varphi_i | x | \varphi_i \rangle = \mu_i. \tag{2}$$

In these formulas, $\langle \varphi_i | \varphi_i \rangle$ implies an integration over the entire measurement range, S_{ij} is the overlap integral quantifying the 'similarity' of results of two laboratories *i* and *j*, and μ_i is the expectation value of laboratory *i* for the sample tested.

The measurement functions of all N laboratories constitute a basis set, $\{ | \varphi_i \rangle, i = 1, 2, ..., N \}$, from which the interlaboratory measurement function is constructed, i.e.

$$|\Psi\rangle = \sum_{i} c_{i} |\varphi_{i}\rangle.$$
(3)

A procedure needs to be developed to estimate the expansion coefficient c_i .

It is assumed that the best estimate of the expectation value of the dataset is provided by the combination of laboratory measurement functions which has the highest probability. This combination comprises the best trade-off between the number of overlapping observations in relation to the intensities of the overlaps. We search for the maximum of $\langle \Psi | \Psi \rangle$ under the constraint $\sum_i c_i^2 = 1$. For this, we use the method of Lagrange multipliers

$$\frac{\partial}{\partial c_{\alpha}} |\Psi\rangle - \lambda \left(\sum_{i} c_{i}^{2} - 1\right) = 0, \qquad \alpha = 1, \dots, N$$
(4)

which is equal to

$$\frac{\partial}{\partial c_{\alpha}} \left(\sum_{i,j} c_i c_j \langle \phi_i | \phi_j \rangle - \lambda \left(\sum_i c_i^2 - 1 \right) \right) = 0,$$

$$\alpha = 1, \dots, N \tag{5}$$

and can be elaborated to the equation

$$\sum_{i} c_i S_{\alpha i} - \lambda c_{\alpha} = 0, \qquad \alpha = 1, \dots, N.$$
(6)

The *N* equations given in Eq. (6) can be written in matrix form as an eigenvalue problem $Sc = \lambda c$, where the symmetric $N \times N$ overlap matrix **S** contains elements S_{ij} and **c** is the $N \times 1$ vector with elements c_i . There are *N* solutions λ with accompanying eigenvectors **c**. The trace of the overlap matrix **S** is equal to *N*, because all the diagonal elements are equal to one. Therefore, the sum of the *N* eigenvalues is also equal to *N*.

40

Let $|\Psi_1\rangle$ be the interlaboratory measurement function corresponding to the first (highest) eigenvalue λ_1 and eigenvector c_{λ_1} . Then the expectation value X_1 and variance s_1^2 can be obtained with the following formulas

$$X_{1} = \frac{\langle \Psi_{1} | x | \Psi_{1} \rangle}{\langle \Psi_{1} | \Psi_{1} \rangle} \tag{7}$$

$$s_1^2 = \frac{\langle \Psi_1 | x^2 | \Psi_1 \rangle}{\langle \Psi_1 | \Psi_1 \rangle} - X_1^2.$$
(8)

When all laboratories work in perfect agreement, i.e. all laboratory measurement functions are identical, all overlap integrals S_{ii} will be equal to one and the eigenvalue λ_1 of $|\Psi_1\rangle$ will be equal to N. Such a situation will normally not be encountered, and λ_1 will have a value between 1 (all off-diagonal overlap integrals are zero) and N. The value of λ_1 is a measure for the degree of overall comparability of the laboratories. More general: eigenvalue λ_i is proportional to the probability of eigenvectors c_{λ_i} . The formula $(\lambda_i/N) \times 100, 0 \le \lambda_i \le 1$, represents the percentage of basis functions (or observations) incorporated in the interlaboratory measurement function $|\Psi_i\rangle$. For instance, if the eigenvalue λ_2 of the second eigenvector c_{λ_1} is close to λ_1 , a bimodal distribution is present. This may be inspected in detail by plotting the laboratory means in a histogram or, in analogy with principal component analysis, by the construction of a biplot (eigenvector c_{λ_1} versus eigenvector $c_{\lambda_{\lambda_{\lambda_{\lambda_{\lambda}}}}}$).

The procedure outlined above renders the total variance s_{total}^2 . Frequently, separate estimates of the within- and between-laboratory variances (s_{within}^2 respectively s_{between}^2) are required. This is accomplished by estimating s_{total} and s_{within} followed by the calculation of s_{between} using the formula $s_{\text{total}}^2 = s_{\text{between}}^2 + s_{\text{within}}^2/n$, *n* being the number of replicates.

The within-laboratory standard deviations can be obtained for each interlaboratory measurement function $|\Psi_i\rangle$ individually. To this end, a new basis set is employed: $\{c_{ik} | \theta_k\rangle$, $k = 1, 2, ..., N\}$. The functions $|\theta_i\rangle$ are obtained from $|\varphi_i\rangle$ by replacing in the formula of the latter the expectation value μ_i of the laboratory *i* by the expectation value X_i of the interlaboratory measurement function $|\Psi_i\rangle$. The uncertainties of the individual laboratories are not changed.

Consequently, the functions $|\theta_i\rangle$ have the same expectation value and differences in the overlap integrals are caused solely by differences in the uncertainties of the laboratories concerned. A 'reproducibility measurement function' $|\Theta\rangle$ is now sought as the linear combination $|\Theta_i\rangle = \sum_{k} \gamma_{ik} c_{ik} |\theta_k\rangle$. The coefficients γ are found using the method of Lagrange multipliers as described previously. The variance calculated with formula (8) using $|\Theta\rangle$ now represents the within-laboratory variance. In the following, the within-laboratory standard deviation obtained according to this approach is denoted as $s_{\text{within,weighted}}$. The procedure can also be carried out without the coefficients being used as weighing factors (all c_{ik} are set equal to 1), the within-laboratory standard deviation obtained in this way is indicated as Swithin.all.

For the implementation of the model, a mathematical formulation needs to be established for the laboratory measurement functions. In this paper, the square root of the normal distribution is used, i.e.

$$|\varphi_{i}(x)\rangle = \sqrt{\frac{1}{\sigma_{i}\sqrt{2\pi}}} e^{\frac{-(x-\mu_{i})^{2}}{2\sigma_{i}^{2}}}.$$
(9)

The use of the normal distribution for the probability function of a laboratory makes the approach conceptually transparent to laboratories. It is also suggested by the use of this distribution as basis for control charts. In most cases, however, the number of observations from which the laboratory mean is estimated is rather small, so that Student's *t*-distribution may be considered more appropriate. In this paper, calculations are carried out using basis functions based on both the normal distribution and Student's t-distribution. The model can also be based on different probability functions. As an example, the symmetric triangular probability distribution is employed.¹ It is emphasized that no assumptions are made regarding the relative magnitude of the within-laboratory variances, nor about the nature of the distribution of Ψ^2 .

¹ With upper bound a_+ and lower bound a_- and $a = (a_+ - a_-)/2$, this distribution is $p(t) = (t - a_-)/a^2$ for $a_- \le t \le (a_+ + a_-)/2$, $p(t) = (a_+ - a_-)/a^2$ for $(a_+ + a_-)/2 \le t \le a_+$, p(t) = 0 otherwise [34].

The procedure requires a knowledge or estimate of the uncertainty of the results of each laboratory. In interlaboratory studies where the analyses are carried out in replicate, the standard deviations of the individual laboratories can be calculated and used in an approximative sense for this purpose. When the standard deviations or other measures of uncertainty are not known, an estimate may be made giving each laboratory the same standard deviation. In analogy with the international harmonised proficiency testing protocol [6], these estimates may include a value chosen by perception, for instance derived from literature, other interlaboratory studies or experience, by prescription, based on the performance characteristics required for a specific task, by reference to validated methodology and/or by reference to a generalised model. In this paper, examples are given for most possibilities. In addition, in the following an implementation of the model is provided which reproduces for a normal distribution the mean and standard deviation. This implementation, denoted as 'the normal distribution approximation', can be used when no estimates of the uncertainties are available.

The model has been programmed in MATLAB. When normal distributions are used as basis functions, the overlap integrals can be expressed analytically. The formulas are given in the annex. When Student's t-distribution is employed, analytical expressions for the overlap integrals can only be obtained for specific cases, e.g. when the standard deviation and the number of degrees of freedom n of the two distributions are equal and n = 3.7.11...Therefore, numerical integration has been used for all calculations invoking probability density distributions other than the normal distribution. To this end, the Matlab toolbox for Composite Gauss integration provided by Wilson [28] has been used. The programs described in this paper are available as MAT-LAB toolbox upon request.

3. An illustration of the model

3.1. A small number of simulated observations

As a first illustration of the procedure, a situation with only two laboratories is discussed. The mea-

surement functions of laboratories 1 and 2 are $|\varphi_1\rangle$ and $|\varphi_2\rangle$, the overlap between them S_{12} . The solution to this problem is given by $|\Psi_1\rangle = (|\varphi_1\rangle + |\varphi_1\rangle)$ $\langle \varphi_2 \rangle / \sqrt{2}$ and $| \Psi_2 \rangle = (| \varphi_1 \rangle - | \varphi_2 \rangle) / \sqrt{2}$, the eigenvalues being, respectively, $1 + S_{12}$ and $1 - S_{12}$. The expectation value (Eq. (7)) will be equal to the conventionally employed average of the two laboratory means if the uncertainties in both are equal, otherwise, the expectation value will be slightly more oriented to the laboratory mean with the lowest uncertainty (see formulas in Appendix A). The uncertainty of the expectation value, expressed as the standard deviation (Eq. (8)), is related to the uncertainties of the individual laboratory data and will increase as the latter become greater. When the result of a third laboratory is added, two situations may occur:

- 1. The overlap integral S_{13} and/or S_{23} is distinctly larger than zero, the laboratory is involved in the mixing process and obtains a non-zero expansion coefficient in the interlaboratory measurement function $|\Psi_1\rangle$.
- 2. The overlap integrals S_{13} and S_{23} differ both negligibly from zero. The 3×3 overlap matrix obtains effectively a block-diagonal form with the original 2×2 matrix, the solutions given above, and a 1×1 matrix with eigenvalue 1.

In the second situation, the result of the third laboratory would be classified as 'outlier' using normal statistics, but it has no effect on the expectation value obtained with this model. This argument may be extended to state that observations have no effect on the expectation value $\langle \Psi_1 | x | \Psi_1 \rangle$ if they do not overlap with either $| \varphi_1 \rangle$ or $| \varphi_2 \rangle$. Only the percentage observations accounted for decreases as more non-overlapping observations occur, i.e. from (1 + S)/2 for the originally two laboratories to (1 + S)/N when N - 2 non-overlapping laboratories are involved.

3.2. Normally distributed simulated datasets

It is illustrative to consider the behavior of the model for normally distributed datasets. To this end, calculations have been carried out on data extracted from a population with $\mu = 1$ and $\sigma_{\text{within}} = 1$, and on

datasets which have in addition a normally distributed 'between-sample' (or between-laboratory) component $\sigma_{\text{between}} = 5 \times \sigma_{\text{within}}$. These datasets simulated 50 observations (laboratory means), each based upon seven replicates. One hundred calculations have been carried out for each of the two types of datasets, the results are summarised in Table 1. The correlation coefficients given in this table relate for the parameter proper of the outcome of this model and ANOVA.

Table 1 demonstrates that the model works well with normally distributed data when the data are drawn from one and the same population. The expectation values and the standard deviations compare well with those obtained with ANOVA. The interlaboratory measurement function Ψ_1 accounts for the major part (92.3 + 1.2%) of the observations. The model renders in comparison with the imposed population characteristics and the outcome of ANOVA a distinctly lower stotal and low percentage observations accounted for when a between-sample variance is applied. This observation can be explained as follows. The model calculates the characteristics of the interlaboratory study on the basis of the interlaboratory measurement function Ψ_1 . In this function, the laboratories, which collectively account for the highest area of overlap, are emphasized. The overlap is determined by the difference in laboratory-means in relation to the magnitude of the laboratory-uncertainties. In the second calculation, s_{between} is much larger than s_{within} . Laboratories in the centre of the normal distribution underlying s_{between} have more neighbours at short range than laboratories in the tails. The procedure emphasizes the former laboratories ----driven by the uncertainties of the observations a narrow selection is made which is reflected in the lower value for s_{between} calculated with the model.

The effect of the ratio $s_{\rm within-laboratory}/s_{\rm between-laboratory}$ on the results has been studied into more detail. Twenty datasets of 20 laboratory-means have been generated from a normal distribution with a fixed $s_{\rm between-laboratory}$. For each set, a series of calculations with different $s_{\rm within-laboratory}/s_{\rm between-laboratory}$, each laboratory being given the same value. The results of the computations are plotted in Fig. 1. The expectation value corresponds well with the mean of the population in all cases. As the within-laboratory

standard deviation decreases, the scatter in expectation values increases and the calculated betweenlaboratory standard deviation s_{between} and the percentage of observations accounted for decrease. The decline in the overall standard deviation calculated with Ψ_1 (IMF1 in Fig. 1) with decreasing $s_{\text{within-laboratory}}/s_{\text{between-laboratory}}$ ratio clearly demonstrates the selection mechanism described above. When the within-laboratory standard deviation is very small compared to the between-laboratory standard deviation, then the highest overlap is established by an interlaboratory measurement function giving the highest weight to the two laboratories with the smallest difference in concentration. For this reason, the spread in expectation values is largest at low $s_{\text{within-laboratory}}/s_{\text{between-laboratory}}$ ratios.

The calculations illustrate that the dependence of the s_{between} with the ratio $s_{\text{within-laboratory}}/s_{\text{between-laboratory}}$ is a consequence of the consideration of the uncertainty in the calculations.

The calculated expectation value and s_{between} are in good agreement with the mean and standard deviation of the underlying population when the ratio $s_{\text{within-laboratory}}/s_{\text{between-laboratory}}$ is about 0.78 (see Fig. 1). This observation can be used to derive empirically an implementation of the model for (near) normal distributions which does not need estimates for $s_{\text{within-laboratory}}$. To this end, a normal distribution of observations (laboratory means) is used and the requirement is imposed that the mean and standard deviation of this normal distribution are reproduced well. In analogy with robust estimators of the standard deviation of a population [17,26], the standard deviation is approximated as $s_{trial} = \alpha \times MAD = \alpha$ \times median($|d_{ij}|$), where $d_{ij} = x_{ij} - M$, M being the median of the dataset. The parameter α is varied to minimize the differences (mean - calculated expectation value) and $(s_{\text{population}} - s_{\text{new procedure}})$. The calculations lead to

$$s_{\text{normal distribution approximation}} = 1.168 \times \text{MAD.}$$
 (10)

The use of Eq. (10) and its underlying approximations will be denoted as the normal distribution approximation of the model.

3.3. A case study: bimodal distributions

Among the calculations which have been carried out and described in Section 4, two cases involving

Table 1 Results of 100 calculations on simulated datasets with normal distributions

	ANOVA			This model				
	Mean	s _{total}	swithin	Mean	s _{total}	Swithin, weighed	Swithin, all	Fraction encompassed (%)
$\mu = 1$ and $\sigma_{\text{within}} = 1$, no between laboratory effect	1.00 ± 0.05	1.00 ± 0.04	1.00 ± 0.04	1.00 ± 0.05 $r^2 = 0.98$	0.98 ± 0.04 $r^2 = 0.24$	0.95 ± 0.04 $r^2 = 0.94$	0.94 ± 0.04 $r^2 = 0.94$	92.3±1.2
$\mu = 1$ and $\sigma_{\text{within}} = 1$ and a between-laboratory component $\sigma_{\text{between}} = 5 \times \sigma_{\text{within}}$	1.00 ± 0.1	5.11 ± 0.53	1.00 ± 0.04	1.00 ± 0.3 $r^2 = 0.43$	2.19 ± 0.4 $r^2 = 0.13$	1.00 ± 0.05 $r^2 = 0.71$	0.94 ± 0.04 $r^2 = 0.94$	32.3±3.6



Fig. 1. Dependence of mean and $s_{between}$ calculated with IMF1 on the assumed ratio $s_{within-laboratory}/s_{between-laboratory}$. The dotted lines indicate the maximum and minimum values encountered in the 20 calculations, which have been carried out.

bimodal distributions are encountered which can be used to illustrate and clarify the model. The two datasets concerned are those for CB28 in the BCR reference material CRM 536 (see Fig. 3 and Table 2) and for Pb in a sandy marine sediment (Fig. 4 and Table 4).

Normal distributions are used for the basis functions. In Fig. 2, the sum of all basis functions and of the corresponding probability density functions are depicted (the solid lines) for both CB28 and Pb. For CB28, two poorly resolved maxima are observed, the highest probability density being observed at about 50 μ g/kg, the next highest at about 38 μ g/kg. For Pb, the graphs are more complex and exhibit a number of well-resolved features, the highest probability density being found for a cluster of peaks at approximately 13 mg/kg and the next highest for a cluster centered at about 8 mg/kg. The model described in

this paper implies an orthogonal transformation of the basis functions in such a manner that a linear combination of basis functions with the highest probability (i.e. the highest eigenvalue λ in the equation Sc = λc) is established. This linear combination is denoted as the interlaboratory measurement function Ψ_1 or IMF1. In addition, N-1 in other interlaboratory measurement functions Ψ_i (or IMF *i*) are constructed. In practice, the latter are ranked according to their eigenvalue λ_i and may be used to obtain insight in structures within the dataset. In most cases, it is sufficient to consider Ψ_2 , in particular when λ_2 attains a value close to λ_1 which indicates that a bimodal distribution is present. With Pb the eigenvalues $\lambda_1 = 18.6$ and $\lambda_2 = 9.1$ correspond to fractions encompassed ($\lambda_i / N \times 100$) of about 35% and 17%. With CB28, the fractions encompassed are about 56% and 37%.

Congene	er	CRM 349								CRM 536							
		Certificate	This mo	bdel						Certificate	This mo	del					
			Basisset		Basisset	ø	Basisset:	.2	Normal		Basisse		Basisset Student	n s	Basisset		Normal distribution
			distribu	tion	<i>t</i> -distrib	ution	triangula distributi	on	approximation		distribu	tion	t-distrib	ution	triangul distribut	ar tion	approximation
			IMF1	IMF2	IMFI	IMF2	IMF1	IMF2	IMFI		IMF1	IMF2	IMF1	IMF2	IMF1	IMF2	IMF1
CB28	Mean (µg/kg)	68.0	66.1	73.0	66.0	69.0	66.1	74.5	65.6	44.4	44.2	44.3	43.2	45.4	44.4	44.0	44.4
	Total SD (μg/kg)	11.7	9.8	16.2	9.9	15.0	10.8	16.6	9.1	6.3	6.1	7.0	6.0	6.2	6.1	<i>T.T</i>	7.9
	SD within (all data, μg/kg)	7.6	9.9		5.9		7.6				3.0		2.1		3.5		
	SD between (μg/kg)	11.3	9.2	16.0	9.3	14.9	10.1	16.3			5.9	6.9	5.9	6.2	5.9	7.3	
	SD within (weighed, μg/kg)		7.5	5.5	7.2	4.9	8.6	6.3			3.1	2.4	2.2	2.0	3.7	3.3	
	Fraction encompassed (%)		63.2	17.5	56.4	16.6	64.3	17.5	72.0		56.4	36.6	49.1	37.1	58.6	34.1	83.3
CB52	Mean (μg∕kg)	149.1	164.9	133.2	166.4	137.3	169.8	136.4	150.3	38.4	36.2	43.2	34.9	43.1	36.5	43.3	38.2
	Total SD (μg/kg)	28.5	25.4	29.4	21.9	23.1	10.5	13.9	34.5	7.1	6.2	7.5	5.6	6.9	6.5	7.6	8.9
	SD within (all data, μg/kg)	14.0	12.7		9.4		4.9				3.1		2.2		3.6		
	SD between (μg/kg)	27.6	24.4	29.1	21.4	22.9	10.1	13.8			6.0	7.4	5.5	6.9	6.3	7.6	
	SD within (weighed, μg/kg)		16.3	8.5	11.2	6.8	5.6	5.3			3.7	2.3	2.6	1.9	4.3	2.8	
	Fraction encompassed (%)		42.7	20.7	36.0	18.8	24.9	15.8	75.3		53.2	29.0	43.9	26.4	55.1	28.7	83.3
CB101	Mean (μg∕kg)	371.8	371.6	370.7	372.5	368.6	375.4	363.0	372.1	43.7	44.8	39.3	44.9	43.3	44.8	37.3	44.1
	Total SD (μg/kg)	29.9	31.5	42.0	29.0	35.0	20.6	26.6	35.5	6.2	5.2	9.0	4.8	7.2	5.5	6.7	6.7
	SD within (all data, μg/kg)	25.7	22.2		16.5		8.6				3.1		2.2		3.6		
	SD between (μg/kg)	27.6	29.6	41.2	27.9	34.4	20.2	26.0			4.9	8.9	4.6	7.1	5.2	9.4	
	SD within (weighed, μg/kg)		23.9	18.6	17.8	14.5	6.6	<i>T.T</i>			3.7	2.8	2.6	2.5	4.3	3.7	
	Fraction encompassed (%)		67.9	19.7	57.9	22.5	37.0	25.7	82.7		55.0	18.1	44.3	19.0	56.9	17.2	74.7
CB153	Mean (μg∕kg)	936.3	938.1	908.0	935.2	907.6	939.2	934.7	927.9	50.3	51.0	46.1	51.2	49.6	51.0	43.8	50.8
	Total SD (µg∕kg)	59.3	58.6	83.3	50.8	57.9	63.6	155.8	47.6	5.7	5.0	7.8	4.8	7.2	5.3	6.7	7.0
	SD within (all data, μg/kg)	56.7	49.5		39.0		55.7				2.9		2.1		3.4		
	SD between (µg/kg)	53.4	55.2	82.2	48.5	56.9	59.6	155.0			4.8	T.T	4.7	7.2	5.0	6.6	
	SD within (weighed, μg/kg)		44.3	30.4	34.1	24.4	49.9	36.3			3.6	2.8	2.5	2.3	4.2	3.1	
	Fraction encompassed (%)		64.8	15.1	57.8	17.3	65.6	14.0	69.1		54.7	19.7	44.8	21.6	56.0	18.5	83.7
CB180	Mean (μg∕kg)	281.5	277.2	275.3	277.7	274.9	276.0	277.1	276.1	22.4	20.8	24.4	20.6	23.4	21.0	24.6	21.6
	Total SD (μg∕kg)	33.3	26.8	48.8	22.5	43.4	29.4	51.2	30.8	3.6	2.4	3.6	2.0	3.1	2.7	3.8	3.3
	SD within (all data, μg/kg)	19.5	18.0		13.7		20.4				1.4		1.0		1.6		
	SD between (μg/kg)	32.3	25.6	48.4	21.8	43.2	28.0	50.7			2.3	3.6	1.9	3.1	2.5	3.7	
	SD within (weighed, μg/kg)		17.3	13.7	12.4	9.9	20.0	15.8			1.7	1.2	1.2	0.9	2.0	1.4	
	Fraction encompassed (%)		55.8	18.6	47.1	17.5	58.3	18.3	70.1		51.9	19.5	44.4	17.1	52.7	19.9	75.9

Table 2 Results of calculations on the analysis of chlorobiphenyls in the fish oil CRM 349 and the fresh water sediment CRM 536



Fig. 2. The model illustrated on observed datasets with a bimodal character: CB28 in CRM 536 and Pb in the sandy marine sediment sample QTM002MS [38].

For CB28 and Pb, the values of the interlaboratory measurement functions IMF1 and IMF2 and the corresponding probability density functions IMF1² and IMF2² are also depicted in Fig. 2 as dashed and dotted curves, respectively. In the case of Pb, it is seen that the region with the highest probability density is contained in IMF1, IMF2 encompassing the next highest mode. The expectation values calculated for these modes were, respectively, 13.37 ± 1.57 and 8.65 ± 2.02 . In the case of CB28, IMF1 and IMF2 have mass at both maxima, the expectation values of both (44 ± 6 and 44 ± 7) are in between. The uncertainty in the observations in relation to the concentration difference between the two maxima is for CB28 relatively large in comparison to that for Pb as measured by the ratio (concentration difference maxima)/(uncertainty in observations), which is about 3.9 for CB28 and 5.8 for Pb. When the calculations are repeated for CB28 using half the value of the reported standard deviations as uncertainties, the clusters are resolved with expectation values close to peak maxima seen in Fig. 2.

These findings indicate that our model can deal with bimodal distributions. The eigenvalues λ_1 , λ_2 , ..., act as indicators for possible multimodality. A separation between the modes will be accomplished

when the difference between the modes is large with respect to the uncertainties in the observations. In other cases, biplots provide insight into the data structures.

4. Calculations with datasets arising from interlaboratory studies

In this section, the model described in this paper will be applied to datasets obtained in material-certification studies of the EU and in international interlaboratory studies conducted in the framework of the EU QUASIMEME project [29,30].

4.1. Calculations conducted on the certified reference materials CRM 349 and CRM 536 issued by the EU-BCR

The former EU-BCR $programme^2$ has issued the certified reference material for chlorobiphenyls in fish oil CRM 349 in 1988 [31] and CRM 536 for PCBs in fresh water sediment in 1998 [32,33]. Laboratories with a proven track record were invited to take part in the material-certification studies. The protocol provided stipulated that the analyses were to be carried in fivefold and that appropriate quality assurance and quality control had to be in place and documented. The analytical results were discussed in a meeting with all participating laboratories. The final dataset was defined on the basis of this technical discussion. The certified values and the overall standard deviations were calculated, respectively, as the mean and the standard deviation of the laboratory means using several statistical procedures which test outliers, normality and homogeneity of variances. The confidence intervals were calculated with the t-factors proper [31–33].

Results of calculations are presented for CB28, CB52, CB101, CB153 and CB180 (Table 2). Four different implementations have been used: three

different basis sets, i.e. the normal distribution, Student's *t*-distribution and the symmetric triangular probability distribution (the parameter *a* [34] is set to $3 \times s_{\text{within}}$), and the normal distribution approximation. The latter uses only the laboratory means, the former three also the within-laboratory standard deviations.

For the model, the values calculated for the first interlaboratory measurement function IMF1 should be compared with the certified values. The results for the second interlaboratory measurement function IMF2 are given for an insight into the structure of the data.

For CRM 536, the characteristics calculated with the model in all implementations correspond well with the certified values. The overall standard deviations calculated with the model are somewhat lower. except when the normal distribution approximation is used. This observation can be attributed to the effect of selection because the between-laboratory standard deviation is in most cases in the order of 1.5-2 times higher than the within-laboratory standard deviation. The overall standard deviation calculated with the normal distribution approximation is in the same order or larger than the value stated in the certificate. In this case, the standard deviation calculated with formula (10) is used as within-laboratory standard deviation for all laboratories. The magnitude of this standard deviation is distinctively higher than the observed within-laboratory standard deviations.

The sum of the percentages of observations accounted for obtained with the model for the first two interlaboratory measurement functions is, in all cases, 75% or more. This indicates that these two interlaboratory measurement functions describe to a large extent the structure in the datasets. In most cases, the results suggest that two modes are present. This was discussed previously, in particular the case for CB28, in Fig. 2 and is clearly demonstrated in the biplot depicted in Fig. 3. It is not clear whether this possible bimodality has an analytical significance. Two-dimensional gas chromatography has been involved in the certification-study [33]. The possibility of coeluting congeners as CB25, CB26, CB29 and CB31 can be ruled out, the differences might be caused by differences in recoveries or by evaporation losses during the clean-up [35]. The certified concentration of CB28 and its uncertainty embodies the concentra-

² Now the EU-Standards, Measurement and Testing Programme.



Fig. 3. A biplot for CB28 in CRM536. The eigenvectors belonging to the two interlaboratory measurement functions IMF1 and IMF2 are plotted against each other.

tions of the two possible modes distinguished in this study. In view of the limited number of observations, there is no reason to question the certified value. Further investigations might give rise, however, to a better analytical understanding in the distribution of the observations and, hence, to a more accurate certified value.

The results for CRM 349 exhibit more or less the same pattern as found for CRM 536. The data for CB52 present a special case. The calculations with the new model using, for instance, the basis functions based on the normal distributions, reveal a bimodal structure with two different expectation values which are, respectively, distinctively higher and lower than the certified value. When the model is used in the normal distribution approximation, one mode is found with an expectation value close to the value stated in the certificate. The difference between the results of these two implementations of the model can be traced back to the difference between the observed withinlaboratory standard deviations and the constructed value calculated with formula (10). The information presented in this study could, at the time of certification, have resulted in additional analytical assessments and, depending on the outcome thereof, in a more accurate certified value.

4.2. Calculations conducted on the results of interlaboratory studies carried out in the context of the EU QUASIMEME project

QUASIMEME was a project sponsored by the European Union with the objective to improve the quality of measurements carried out by about 95 laboratories involved in European marine monitoring programmes [29,30]. Interlaboratory studies were conducted for nutrients in seawater and for trace metals and for CBs and polycyclic aromatic hydrocarbons in sediments and biological tissues. These interlaboratory studies had the objective to assess the analytical state of practice at that time and be representative for the data submitted to European marine monitoring programmes. Consequently, no methodological requirements were posed. The results of expert laboratories were used to evaluate the assigned values. The data were assessed to establish important features describing the overall performance of the participants. The examples described in this section have been selected to illustrate the potential of the model to handle difficult distributions.

4.2.1. Metals in marine sediment

In the first round of the QUASIMEME interlaboratory studies, marine sediments were distributed to

	Assigned value	ISO 5725	Robust	techniques	New method							
			AMC	Lischer	Basisset normal distribut	ion	Basisset: Student' <i>t</i> -distribu	s ition	Basisset symmetr triangula distribut	: ric ur ion	Normal approximation	
					IMF1	IMF2	IMF1	IMF2	IMF1	IMF2	IMF1	
Mean (mg/kg)	0.03	0.060	0.059	0.033	0.023	0.060	0.023	0.078	0.023	0.012	0.024	
Total SD (mg/kg)					0.010	0.100	0.019	0.167	0.005	0.003	0.011	
SD within (all data, mg/kg)		0.069	0.007	0.006	0.009		0.019		0.003			
SD between (mg/kg)		0.148	0.059	0.026	0.010	0.103	0.019	0.166	0.004	0.003		
SD within, weighed (mg/kg)					0.007	0.008	0.006	0.038	0.003	0.002		
Fraction encompassed (%)					50.1	10.9	44.8	11.1	29.7	17.1	62.1	
Mean (mg/kg)	0.03	0.030	0.030	0.023	0.022	0.039	0.023	0.014	0.023	0.019	0.022	
Total SD (mg/kg)					0.007	0.022	0.005	0.005	0.003	0.007	0.003	
SD within (all data, mg/kg)		0.014	0.006	0.006	0.006		0.004		0.002			
SD between (mg/kg)		0.021	0.026	0.008	0.007	0.022	0.005	0.005	0.003	0.007		
SD within, weighed (mg/kg)					0.006	0.009	0.004	0.003	0.002	0.002		
Fraction encompassed (%)					68.1	12.5	59.5	14.5	48.5	16.4	61.4	

Table 3 Results for Cd in a sandy marine sedime

Cd

all data

N = 22

Cd trimmed

data set N = 13

		Assigned	Robust	nethods	This mo	del						
		value	AMC	Lischer	Basisset normal distribut	:: tion	Basisset: Student's <i>t</i> -distribution		Basisset symmet triangul distribut	: ric ar ion	Normal distribution approximation	
					IMF	IMF2	IMF1	IMF2	IMF1	IMF2	IMF1	
Pb	Mean (mg/kg)		11.84	11.82	13.37	8.65	13.43	8.79	13.32	8.62	12.02	
full data set	$s_{\text{total}} (\text{mg/kg})$				1.57	2.02	1.59	2.17	1.69	2.10	3.69	
N = 53	$s_{\text{between}} (\text{mg/kg})$		4.20	4.16	1.51	1.99	1.55	2.15	1.61	2.07		
	swithin, weighed (mg/kg)				0.99	0.68	0.80	0.71	1.13	0.74		
	$s_{\text{within, all}}$ (mg/kg)		0.76	0.80	0.82		0.69		0.93			
	Fraction encompassed (%)				35.1	17.1	30.5	15.0	36.3	17.6	37.2	
Pb	Mean (mg/kg)	8.7	8.74	8.86	8.66	11.15	8.74	11.28	8.70	11.04	8.71	
partial methods	s _{total} (mg/kg)				1.59	2.02	1.89	2.00	1.63	2.09	2.65	
N = 24	s _{between} (mg/kg)		3.36	3.52	1.57	1.99	1.86	1.97	1.60	2.06		
	s _{within, weighed} (mg/kg)				0.65	0.76	0.74	0.67	0.74	0.83		
	$s_{\text{within, all}}$ (mg/kg)		0.68	0.71	0.67		0.60		0.76			
	Fraction encompassed (%)				34.2	21.8	30.7	20.4	35.8	21.6	66.8	
Pb	Mean (mg/kg)	13.1	14.17	14.09	13.97	14.26	14.03	13.63	13.96	14.30	13.98	
total methods	s _{total} (mg/kg)				1.27	2.65	1.09	1.84	1.40	2.86	1.15	
N = 29	s _{between} (mg/kg)		3.10	2.06	1.20	2.27	1.04	1.83	1.31	2.83		
	s _{within,weighed} (mg/kg)				0.96	0.75	0.73	0.48	1.11	1.31		
	$s_{\rm within, all} ({\rm mg/kg})$		0.96	0.81	0.98		0.77		1.12			
	Fraction encompassed (%)				52.0	13.4	46.1	12.7	53.2	13.5	62.0	

Table 4	
Results for lead in a sandy marine sediment (sample QTM002MS [38	;])
	_

_

assess the analyses of trace metals [36]. The laboratories analysed the samples on six different days in duplicate. During the assessments, it appeared that systematic differences occurred for a number of metals in sediments due to differences in the digestion techniques applied. This problem had to be addressed for each element and sampled separately during the exercise, as the calculation of expectation values for bimodal distributions is not appropriate. The data for Cd and Pb are considered.

The determination of Cd in a sandy marine sediment with digestion methods giving a partial recovery of the total metal content (e.g. aqua regia) proved to constitute a difficult problem for conventional statistics [22]. The concentrations of Cd were very low, resulting in a strongly tailing distribution. Both ISO 5725 and the AMC robust statistics could not cope with this distribution, results in line with the independently established assigned values were only obtained when 'obvious outliers' were removed. As the assigned values were established independently, the dataset could be trimmed prior to the statistical evaluation on the basis of the *z*-scores. It appeared that trimming the dataset by discarding all data with absolute *z*-scores greater than six gave rise to expectation values, which were close to the assigned values. This approach was adopted as a standard procedure in the QUASIMEME programme after round three [30].

In Table 3, the outcome of the present procedure is provided and compared with the calculations reported previously. Trimming the dataset has a great effect on the results obtained with ISO 5725 and both robust procedures. The expectation values obtained with the new model in all implementations change hardly. Upon trimming, the percentage of observations accounted for increases owing to the deletion of the results of nine laboratories, which have absolute *z*-scores greater than six except when the normal distribution approximation is used. The results of the calculations illustrate the high robustness of the new model.

A second element, Pb, was studied in the sandy marine sediment in the first round of the QUA-SIMEME project. A total of 53 laboratories submitted data. Twenty-four laboratories used acid mixtures such as aqua regia to digest the samples, so that Pb was only partially recovered from the material. The



Fig. 4. A biplot depicting the coefficients in the first two interlaboratory measurement functions obtained for the full dataset of Pb in the sandy marine sediment sample QTM002MS [38]. Laboratories using partial and total methods have been indicated separately. Substructures within the sets of partial and total data are visible.

remaining 29 laboratories used a digestion invoking hydrofluoric acid or equivalent, providing a 'total' determination of Pb. The two datasets were evaluated separately, the difference was found to be significant [36].

In Table 4, the results of the two different applications of robust statistics and the results obtained by the present method are given. The expectation values of Pb found with the present procedure agree well for both partial and total methods with those obtained with robust statistics.

For the full dataset, both robust statistics procedures provide an expectation value which is located between the expectation values for the individual total and partial populations. The present procedure renders two interlaboratory measurement functions with low, similar percentages of observations accounted for, indicating the presence of a bimodal distribution. The expectation values of these two interlaboratory measurement functions are quite close to those obtained on the separate total and partial populations. The bimodal nature is also demonstrated by the biplot given in Fig. 4, and by the right-hand part of Fig. 2, which was discussed in the previous chapter. The model also indicates a bimodal distribution for the data observed with partial methods.

The good agreement between the calculations on the full dataset and on the separate populations may be traced back to a within-laboratory precision which is relatively low compared with the difference be-

Table 5 Results of laboratory-performance studies on Cd, Cu, Hg and Pb in a plaice muscle tissue (sample QTM017BT [38])

		Assigned	Number	Robust	statistics	This m	odel		
		value	of laboratories	AMC	Lischer	Basis functio normal distribu	n: ition	Normal distribution approximation	
						IMF1	IMF2	IMF1	
Cd full dataset	Mean (µg/kg)	6.1	37	8.79	7.75	5.42	6.86	5.95	
	s _{total} (μg/kg) Fraction encompassed (%)			5.97	4.88	1.44 46.5	2.54 17.1	3.16 67.3	
Cd trimmed dataset	Mean (µg/kg)	6.1	27	5.78	5.80	5.42	6.76	5.63	
	s _{total} (μg/kg) Fraction encompassed (%)			2.00	1.98	1.43 63.7	2.50 23.3	2.06 79.5	
Cu full dataset	Mean (µg/kg)	0.25	40	0.27	0.26	0.23	0.27	0.24	
Cu trimmed dataset	$s_{\text{total}} (\mu g/kg)$			0.09	0.08	0.05	0.08	0.06	
	Fraction encompassed (%)					59.4	17.3	69.0	
Cu trimmed dataset	Mean (µg/kg)	0.25	34	0.24	0.24	0.23	0.27	0.23	
	s _{total} (μg/kg) Fraction encompassed (%)			0.06	0.05	0.05 69.8	0.08 20.4	0.05 74.4	
Hg full dataset	Mean (µg/kg)	67.4	35	64.12	63.85	64.23	58.66	63.83	
-	$s_{\text{total}} (\mu g/kg)$			14.87	14.59	7.65	15.30	10.94	
	Fraction encompassed (%)					53.1	20.6	66.2	
Hg trimmed dataset	Mean (µg/kg)	67.4	33	64.05	63.85	64.23	58.66	63.84	
	s _{total} (µg/kg)			13.22	13.03	7.65	15.29	10.93	
	Fraction encompassed (%)					56.3	21.9	70.2	
Pb full dataset	Mean (µg/kg)	18.6	29	48.37	38.80	16.87	24.57	23.60	
	$s_{\text{total}} (\mu g/\text{kg})$			51.44	40.09	9.44	14.36	26.60	
	Fraction encompassed (%)					41.6	18.9	68.4	
Pb trimmed dataset	Mean (µg/kg)	18.6	17	18.91	18.91	17.39	22.87	18.14	
	s _{total} (µg/kg)			10.49	10.49	9.00	12.85	10.38	
	Fraction encompassed (%)					67.9	27.9	77.4	

tween the sub-populations corresponding with the total and partial digestion. Such a favourable combination will not always be encountered.

4.2.2. Trace metals in marine biological tissues

In the fourth round of the QUASIMEME project, Cd, Cu, Hg and Pb were measured in a plaice muscle homogenate (sample QTM017BT). Laboratories were requested to report a single result from the analysis of the sample. The concentrations were very low.

The uncertainty of the laboratory data has been estimated as no replicate data are available. Berman and Boyko [37] have organised interlaboratory studies of trace metals in marine biological tissues for the International Council for Exploration of the Sea (ICES). These interlaboratory studies covered a range of concentrations and included estimates of the within-laboratory standard deviations. For the elements included in this study, the medians of withinlaboratory standard deviations for matrices with similar concentration levels were considered. This assessment resulted in the following estimates for the within-laboratory relative standard deviations: Cu ---14%, Cd — 15%, Hg — 8% and Pb — 20%. For each element, the median concentration in the present dataset has been established. Each laboratory is given a s_{within} equal to the appropriate percentage of the median.

The computations are limited to the robust methods of AMC and Lischer [26] and the model using normal distributions as basis functions and the normal distribution approximation. The results are depicted in Table 5.

The number of laboratories included in the trimmed dataset for Cd and Pb are substantially less than in the full dataset. This reflects the difficulty in the analysis of these elements at such low concentrations.

The means calculated with robust statistics for the full datasets are in poor agreement with the assigned value for Cd and Pb, and differ largely from the means obtained for the trimmed datasets for Cd and Pb. Robust statistics give good results for the full and trimmed datasets of Cu and Hg. The calculations indicate that robust statistics cannot handle well the skewed full datasets of Cd and in particular Pb.

The assigned values and the outcome of the model agree for both the full and trimmed datasets with

normal distributions as basis functions. The results obtained for the full and trimmed datasets differ marginally. This demonstrates the ruggedness of the model and that it handles strongly tailing distributions well too. The model applied in the normal distribution approximation appears to have some difficulty to cope with the full dataset of Pb. The full dataset of Pb is highly skewed and the assumptions on which the normal distribution approximations are based are not valid in this case.

5. Conclusions

A new model to infer population characteristics from experimental data is presented and applied to the evaluation of datasets from interlaboratory studies. The calculations using this model offer a considerable number of advantages over the commonly used ISO 5725 and robust statistical procedures. It has been demonstrated that the model copes well with highly skewed and with multimodal distributions. It also provides quantitative measures and graphical means to explore data structures and reveal features which are less or not obvious by inspection or traditional data analysis. The model has been applied in this paper exclusively to the evaluation of interlaboratory studies, but has a much wider generic application.

Acknowledgements

The work described in this paper is the result of years of experience gained among others in the context of material-certification studies and quality assurance programmes funded by the European Union Standards, Measurement and Testing Programme (e.g. contracts MAT1-CT92-0002 and MAT1-CT93-0018). The European Union is acknowledged for this support. I.L. Freriks, Ministry of Transport, Public Works and Water Management, Directorate General for Public Works and Water Management and Waste Water Treatment RIZA, is acknowledged for critically reading the manuscript.

Appendix A. Formulas for the overlap integrals

When the basis functions invoke normal distributions, all integrals required can be calculated analytically

$$S_{ij} = \sqrt{\frac{2}{\sigma_i \sigma_j}} \frac{\exp\left(-\frac{\left(\mu_i - \mu_j\right)^2}{4\left(\sigma_i^2 + \sigma_j^2\right)}\right)}{\sqrt{\frac{1}{\sigma_i^2} + \frac{1}{\sigma_j^2}}}.$$

 S_{ij} is proportional to a normal distribution of the 'variable' $(\mu_i - \mu_j)$ with zero mean and variance $2(\sigma_i^2 + \sigma_j^2)$. For the expectations of x and x^2 , we find

$$\langle \phi_i | x | \phi_j \rangle = S_{ij} \frac{\mu_j \sigma_i^2 + \mu_i \sigma_j^2}{\sigma_i^2 + \sigma_j^2}$$

 $\langle \phi_i | x^2 | \phi_i \rangle$

$$= S_{ij} \frac{\left(\mu_{j} \sigma_{i}^{2} + \mu_{i} \sigma_{j}^{2}\right)^{2} + 2\left(\sigma_{i}^{2} + \sigma_{j}^{2}\right) \sigma_{i}^{2} \sigma_{j}^{2}}{\left(\sigma_{i}^{2} + \sigma_{j}^{2}\right)^{2}}$$

Note that for i = j, these expressions simplify to $S_{ij} = 1$, μ , and $\mu^2 + \sigma^2$, respectively.

References

- [1] J.N. Miller, Analyst 118 (1993) 455-461.
- [2] F.R. Hampel, E.M. Ronch, P.J. Rousseeuw, W.A. Stahl, Robust Statistics, Wiley, New York, 1986.
- [3] C.Z. Mooney, R.D. Duval, Bootstrapping. A Nonparametric approach to statistical inference; SAGE University Paper series on Quantitative Applications in the Social Sciences, Newbury Park, 1993.
- [4] W. Horwitz, Pure Appl. Chem. 66 (1994) 1903-1911.
- [5] W. Horwitz, Pure Appl. Chem. 67 (1995) 331-343.
- [6] M. Thompson, R. Wood, Pure Appl. Chem. 65 (1993) 2123– 2144.
- [7] W.P. Cofino, B. Pedersen, Mar. Pollut. Bull. 32 (1996) 646– 653.
- [8] Analytical Methods Committee, Analyst 117 (1992) 97–104.
- [9] Anonymous, ISO/IEC Guide 43-1, Proficiency testing by in-

terlaboratory comparisons: Part 1. Development and operation of proficiency testing schemes, International Standard Organisation: Geneva, 1996.

- [10] Anonymous, ISO/IEC Guide 43-2, Proficiency testing by interlaboratory comparisons: Part 2. Selection and use of proficiency testing schemes by laboratory accreditation bodies, International Standard Organisation: Geneva, 1996.
- [11] W.J. Youden, E.H. Steiner, Statistical Manual of the AOAC, Association of Official Analytical Chemists, 1975.
- [12] Anonymous, J. Assoc. Off. Anal. Chem., 72 (1999) 694-704.
- [13] Anonymous, Accuracy (trueness and precision) of Measurement Methods and Results: Part 2. Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method, International Standard Organisation, Geneva, 1994.
- [14] M. Feinberg, TrAC, Trends Anal. Chem. 14 (1995) 450-457.
- [15] J. de Boer, J. van der Meer, L. Reutergårdh, J.A. Calder, J. Assoc. Off. Anal. Chem. 77 (1994) 1411–1422.
- [16] V.J. Houba, J. Uittenbogaard, P. Pellen, Commun. Soil Sci. Plant Anal. 27 (1996) 421–431.
- [17] Analytical Methods Committee, Analyst 114 (1989) 1699– 1705.
- [18] Analytical Methods Committee, Analyst 114 (1989) 1693– 1697.
- [19] M. Thompson, B. Mertens, M. Kessler, T. Fearn, Analyst 118 (1993) 235–240.
- [20] M.A. Montfort, Commun. Soil Sci. Plant Anal. 27 (1996) 463–478.
- [21] W. Beyrich, W. Golly, N. Peter, R. Seifert, Die DOD Methode, Kernforschungszentrum Karlsruhe, Karlsruhe, 1990.
- [22] W.P. Cofino, D.E. Wells, Mar. Pollut. Bull. 29 (1994) 149– 158.
- [23] Anonymous, European Laboratory Accreditation Publication ELA-G6: Criteria for Proficiency Testing in Accreditation, European Laboratory Accreditation, 1993.
- [24] M. Thompson, P.J. Lowthian, Analyst 121 (1996) 1597– 1602.
- [25] K. Kafadar, J. Am. Stat. Assoc. 77 (1982) 416-424.
- [26] P. Lischer, Statistik und Ringversuche (Auszug aus dem Schweiz. Lebensmittelbuch), Eidgenoessische Forschungsanstalt fuer Agrikulturchemie und Umwelthygiene: Liebefeld-Bern, 1990.
- [27] C.L. Perrin, Mathematics for Chemists, Wiley, New York, 1970.
- [28] H. Wilson, B. Gardner, MATLAB Numerical Integration Toolbox version 2.0, Mathtools, 1998. www.mathtools.com.
- [29] D.E. Wells, A. Aminot, J. de Boer, W. Cofino, D. Kirkwoord, B. Pedersen, Mar. Pollut. Bull. 35 (1997) 3–17.
- [30] D.E. Wells, W.P. Cofino, Mar. Pollut. Bull. 35 (1999) 18-27.
- [31] B. Griepink, D.E. Wells, M. Frias Ferreira, The certification of the contents (mass fraction) of chlorobiphenyls (IUPAC Nos 28, 52, 101, 118, 138, 153 and 180) in two fish oils, Commission of European Communities, Directorate-General Telecommunication, Information Industries and Innovation, Report EUR 11520 EN, 1988.
- [32] J.W.M. Wegener, W.P. Cofino, E.A. Maier, G.N. Kramer, TrAC, Trends Anal. Chem. 18 (1999) 14–25.

- [33] J.W.M. Wegener, E.A. Maier, G.N. Kramer, W.P. Cofino, The certification of the contents (mass fractions) of chlorobiphenyls IUPAC no 28, 52, 101, 105, 118, 128, 138, 149, 153, 156, 163, 170 and 180 in freshwater harbour sediment CRM 536, European Commission, Directorate Science, Research and Development Report EUR 17799 EN, 1999.
- [34] Anonymous, Guide to the Expression of Uncertainty in Measurement, International Organization for Standardization, Geneva, 1993.
- [35] J. de Boer, personal communication, 1999.
- [36] B. Pedersen, W.P. Cofino, Mar. Pollut. Bull. 29 (1994) 166– 173.
- [37] S.S. Berman, V.J. Boyko, ICES seventh round intercalibration for trace metals in biological tissue (ICES 7/TM/BT part 1 and part 2), International Council for Exploration of the Sea ICES, Copenhagen, 1999.
- [38] B. Pedersen, W.P. Cofino, I. Davies, Mar. Pollut. Bull. 35 (1999) 42–51.