Workshop
on
Laboratory Proficiency Testing –
Experiences in Data Assessment

organised by the Quality Assurance Panel at the Federal Environmental Agency

7 October 2004
Federal Environmental Agency
Berlin, Germany

-Presentations-

Umwelt
Bundes
Amt
Für Mensch und Umwelt

Aims

The workshop will address the current situation and likely future developments in Proficiency Testing (PT) and Laboratory Performance Studies (LPS) schemes in environmental monitoring. Lectures and discussion are designed to highlight:

- Evaluation of participant performance, including application of measurement uncertainty into PT/LPS
- Application and use of PT/LPS by participants, their customers, accreditation and regulatory bodies
- Selection of appropriate schemes
- Legislative demands for PT/ LPS
- The effect of ISO/IEC 17025 on laboratories' participation in PT/LPS including frequency of participation
- International harmonisation of PT/LPS
- Accreditation of PT/LPS providers
- New technical areas and challenges in PT/LPS
- Use and abuse of PT/ LPS schemes.

Content / Programm

A new model for the evaluation of interlaboratory studies

Wim P Cofino
Wageningen University and Research, Alterra, Centre for Water and Climate, P.O. Box 47, 6700 AA Wageningen ,The Netherlands
Email: wim.cofino@wur.nl

Abstract

Recently, a new model to establish the population characteristics of datasets has been established [1]. This model has been extensively tested in the Quasimeme Laboratory Performance Testing scheme. In the past two years, the FAPAS scheme has investigated the merits of the model. The model has been extended to deal with ' less than' data [2].
The principles of the model will be introduced and illustrated with a few examples. In its most comprehensive implementation, uncertainties are specified for each laboratory individually. For most interlaboratory studies it is difficult to obtain these estimates. The model has an implementation, denoted the ' Normal distribution approximation (NDA)', which does not need estimates for the uncertainties. The NDA approach will be discussed, as well as a recently developed improvement of the NDA approach.
Recently, a paper has appeared in literature about the model [4]. This paper, which relates to the most comprehensive implementation of the model, will be treated.

Kurzfassung

Vor kurzem wurde ein neues Modell zur Ermittlung der Gruppenmerkmale von Datensätzen etabliert [1]. Das Modell wurde weitgehend im Rahmen des Quasimeme Ringversuchs-programms getestet und in den letzten zwei Jahren auf seine Vorzüge durch das FAPAS-Programm untersucht. Um auch die Bearbeitung von „kleiner als " Daten zu ermöglichen, wurde das Modell erweitert [2].
Zweck des vorliegenden Beitrags ist es die grundlegenden Prinzipien des Modells anhand einiger Beispiele vorzustellen und zu veranschaulichen. Innerhalb seiner sehr umfassenden Anwendung werden für jedes Labor die Messunsicherheiten individuell angegeben. Bei Ringversuchsstudien ist es häufig schwierig solche Schätzungen zu erhalten.
Daher wurde innerhalb des Modells ein „ Näherungswert der Normalverteilung (NDA)" eingeführt, der die Notwendigkeit der Unsicherheitsabschätzungen überflüssig macht. Dieser NDA- Ansatz sowie auch seine kürzlich entwickelte Modifizierung werden zu Diskussion gestellt. Ein unlängst in der Literatur erschienener Artikel über das vorgestellte Modell [4], der sich auf seine umfassende Anwendung bezieht, soll außerdem zur Diskussion beitragen.

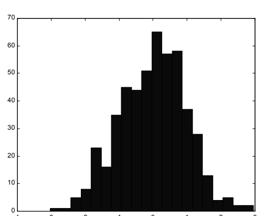
References

1. Cofino, W.P., I. van Stokkum, D.E. Wells, R. A.L. Peerboom, F. Ariese, A new model for the inference of population characteristics from experimental data using uncertainties in data. Application to interlaboratory studies, Chemometrics and Intelligent Laboratory Systems 53 (2000) 37-55
2. Cofino, W.P., I. van Stokkum , J. van Steenwijk, D.E.Wells. A new model for the inference of population characteristics from experimental data using uncertainties. Part II. Application to censored datasets *(accepted for publication in Analytica Chimica Acta)*
3. Eppe, G., W.P. Cofino, E. De Pauw, Performances and limitations of the HRMS method for dioxins, furans and dioxin-like PCBs analysis in animal feedingstuffs : Results of an interlaboratory study, *(accepted for publication in Analytica Chimica Acta)*
4. T. Fearn, " Comments on Cofino Statistics", Accred Qual Assur (2004) 9:441–444

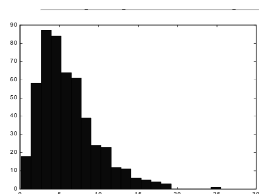
Wim Cofino, Ivo van Stokkum & David Wells

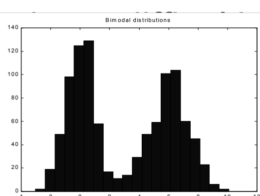
A new model for the inference of population characteristics from experimental data using uncertainties.

Why search for a new model?



- Currently, eithe tlier tests (thus assuming a ce the data) or robust statistic
- The to cop ractice (bir

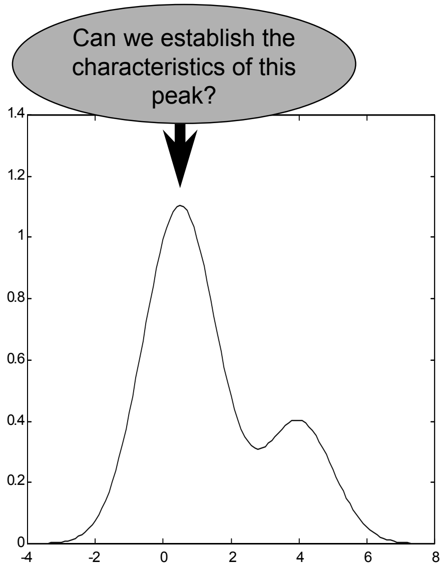




Tailing distributions

Bimodal distributions

The concept of the new model- look for the mode



The overall probability distribution

The model-Quantum chemistry as analogon

- Stationary states of system (e.g. an atom) have wavefunctions Ψ which are solutions to the Schrödinger equation $H\Psi=E\Psi$
- $|\Psi|^2$ is a probability density function

- Our knowledge of measurement processes enables us to to estimate probability density functions of labs/datapoints
- the square root of the estimated pdf for a lab/datapoint can be used as a 'wavefunction' (**Observation measurement function** OMF ϕ)

- Molecular orbitals are constructed from atomic orbitals using matrix algebra:

$$\Psi_j = \sum_i c_{ij} \phi_i$$

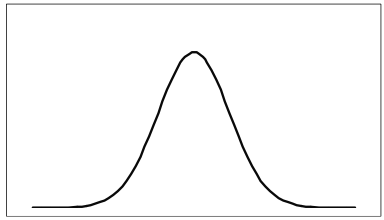
coefficients are determined by finding minimum energy

- A '**Population measurement function'** PMF $\Psi(x)$ is defined as

$$\Psi_j = \sum_i c_{ij} \phi_i$$

Some basics

- The normal distribution: **N(μ,s)**
- The normal distribution is one of the probability density functions (PDFs) used in this model
- PDF's are written as $\phi_i^{\,2}$
- When Normal distributions are used as PDF's, $\phi_i^{\,2}$ stands for a normal distribution characterised by a mean μ_i and standard deviation s_i

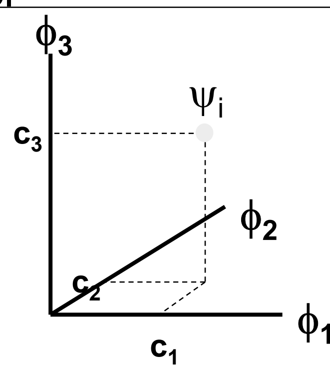


$$\int N(\mu, s)dc = \int \varphi_i^2 dc = 1,$$

$$\int c * N(\mu, s)dc = \int c * \varphi_i^2 dc = \mu$$

The concept of the model

- A 'basisfunction' is associated with each datapoint
- A basisfunction is defined as the square root of a probability density function
- The basisfunctions constitute a space (here $\{\phi_1, \phi_2, \phi_3\}$), in which a 'Population measurement function' (PMF) is constructed
- Knowledge of the PMF enables the calculation of mean and std



The '**Population measurement function'** PMF Ψ(x) is obtained as

$$\Psi_j = \sum_i c_{ij} \phi_i$$

The model - continued (I)

- The expectation value and variance of the population measurement function Ψ_i is calculated as

$$\bar{x}_i = \int x\Psi_i^2 dx / \int \Psi_i^2 dx,$$

$$s_i^2 = \int x^2\Psi_i^2 dx / \int \Psi_i^2 dx - \bar{x}_i^{\,2},$$

$$\Psi_j = \sum_i c_{ij}\phi_i$$

The problem: how to obtain the coefficients?

ALTERRA
WAGENINGEN UR

The model- continued (II)

- The coefficients c_i are obtained by
  - in words

    selecting the largest number of laboratories/data which 'agree'
  - mathematically

    maximising the probability of the (unnormalised) population measurement function, i.e.

$$\int \Psi^2 dx$$

Ψ *being given by*

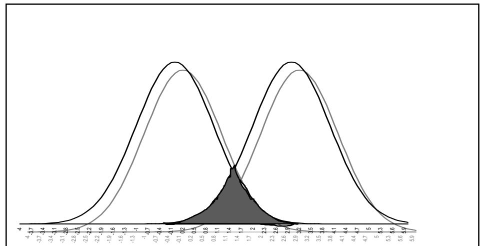
$$\Psi = \sum_i c_i\phi_i$$

using the method of Lagrange multipliers subject to the constraint that the sum of the coefficients equals 1

ALTERRA
WAGENINGEN UR

Overlap as central concept

- The results of two laboratories/datapoints (may) overlap
- The magnitude of overlap is a measure for 'how well the laboratories/data agree'
- The overlap is calculated as follows:



Red: Area of Overlap

$$\int \phi_i \phi_j dx = S_{ij};\quad 0 < S_{ij} < 1;\quad S_{ii} = 1$$

$$\int \phi_i \phi_j dx = S_{ij}$$

$$0 < S_{ij} < 1$$

$$S_{ii} = 1$$

The model - continued (III)

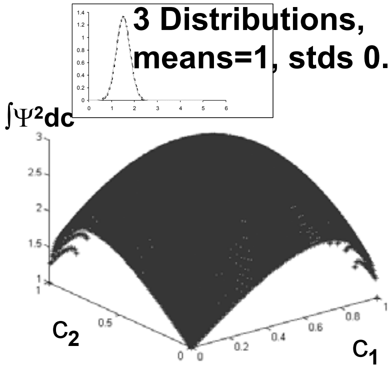
- The mathematical approach amounts to solving the eigenvalue-eigenvector equation,

$$Sc = \lambda c$$

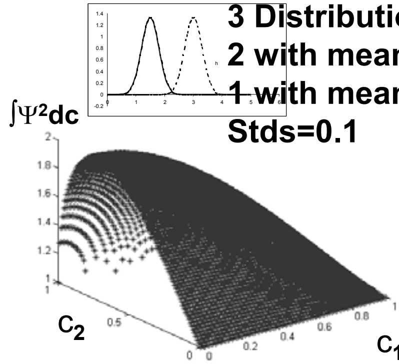
- The eigenvector c_λ with highest eigenvalue λ gives the pursued coefficients
- the eigenvalue λ_i is a measure for the probability of the interlaboratory measurement function Ψ_i, λ_i can be expressed as a percentage

Illustration model

3 Distributions, means=1, stds 0.1



3 Distributions
2 with mean=1,
1 with mean=3,
Stds=0.1



$c_3^2=1-c_1^2-c_2^2$

Max = 3 for $c_1=c_2=c_3$ **Max=2 for $c_1=0$; $c_2=c_3$**

The example with three data continued

Three data: 1, 1 , 1
Std: 0.1, 0.1, 0.1

PMF	mean	std	λ	(λ %)
1	1	0.1	3	100
2	1	0	0	0

Three data: 3 , 1 , 1
Std: 0.1, 0.1, 0.1

PMF	mean	std	λ	(λ %)
1	1	0.1	2	66.7
2	3	0.1	1	33.3

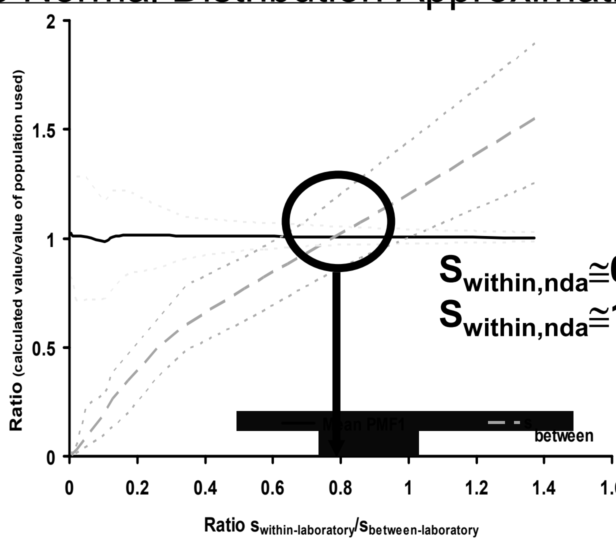
The model requires uncertainties- how to obtain these?

- Estimation of a reasonable value for the parameter/concentration level based on e.g. previous interlaboratory studies, standards, expert judgement, literature,…
- Making use of prescribed performance characteristics
- Making use of performance characteristics of reference

The model has been extended to cope with situations were no estimates of uncertainty are available
- ***The normal distribution approximation***
- ***The bandwidth estimator (to be published)***

ALTERRA WAGENINGEN UR

The Normal Distribution Approximation

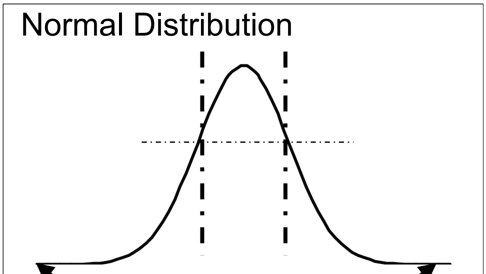


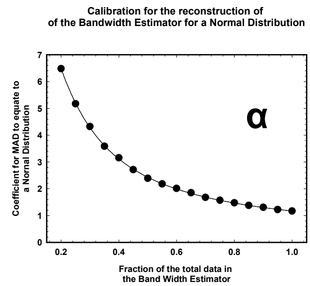
$S_{within,nda} \cong 0.78 * S_{between}$

$S_{within,nda} \cong 1.16 * MAD$

ALTERRA WAGENINGEN UR

The Bandwidth Estimator (I)





•NDA: good estimate mean
and std of population

• all data are used

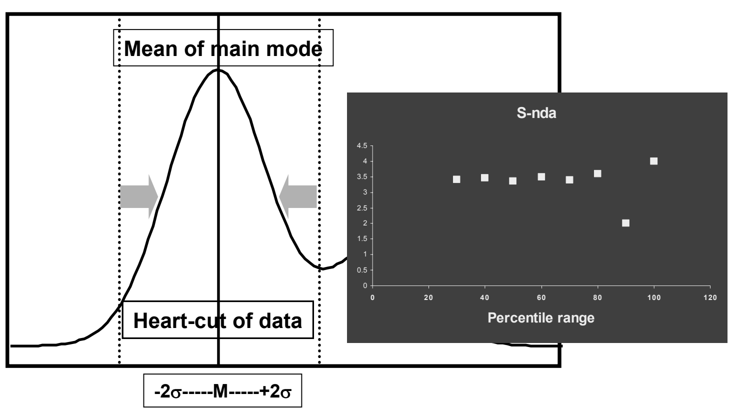
•$S_{within,nda}$=1.16*MAD

Alternative NDA approach with equally good estimation mean and std population:

Use range of data

$S_{within,nda}$=α*MAD

The Bandwidth Estimator (II)



The model summarised

- Model needs estimate of uncertainties of laboratories and an estimate of the type of probability density function of the laboratories/data
- Implementations exist which do not need uncertainty estimates : normal distribution approximation and the bandwidth estimator
- The model can use with different estimates for the probability density function of labs/data (*examples: normal, Student-t, symmetric triangular, rectangular*)
- No assumptions are made regarding the distribution of the population of labs/data
- In addition to the variance, a new measure for the degree of intercomparability is obtained (λ, a measure for the probability of the population measurement function)
- The model provides powerful graphical tools to obtain insight in the data structures

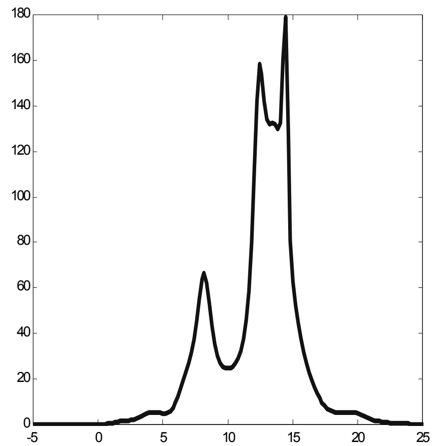
ALTERRA WAGENINGEN UR

Examples

ALTERRA WAGENINGEN UR

Pb in marine sediment

- Laboratories analysed with both ‘total’ (e.g. HF) and ‘partial’ (e.g. aqua regia) methods
- 6 replicates
- Bimodal distribution



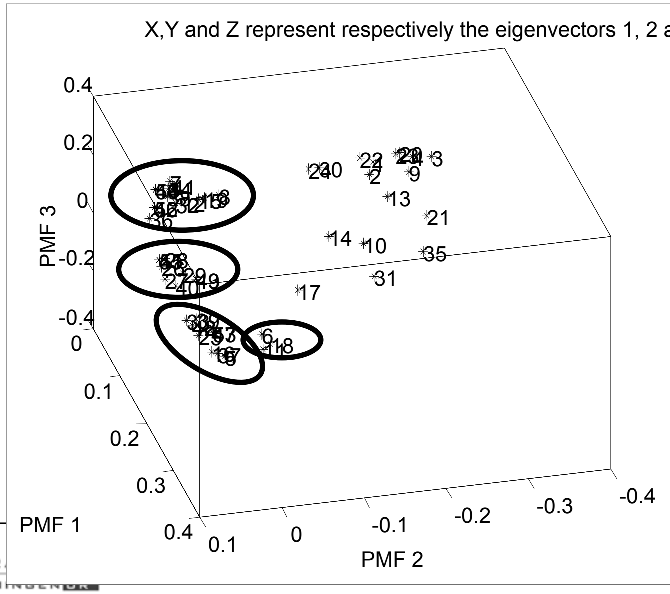
ALTERRA WAGENINGEN UR

Pb in marine sediment: results of statistics (mg/kg)

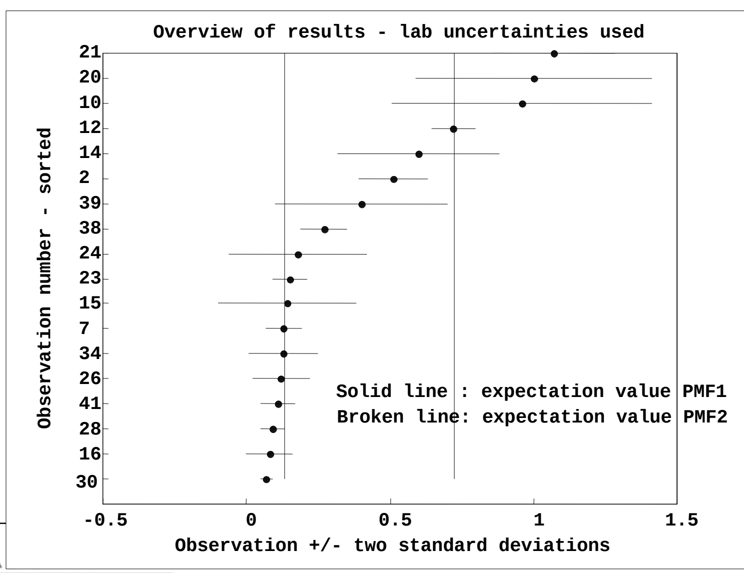
	All data N=53			‘Total’ methods N=29			Partial methods N=24		
	x	**s**	**%**	**x**	**s**	**%**	**x**	**s**	**%**
Robust statistics , UK-AMC									
	11.8	4.2		14.2	2.0		8.7	3.5	
This model									
PMF1	13.4	1.6	35	[redacted]					
PMF2	8.7	2	17	[redacted]					

ALTERRA WAGENINGEN UR

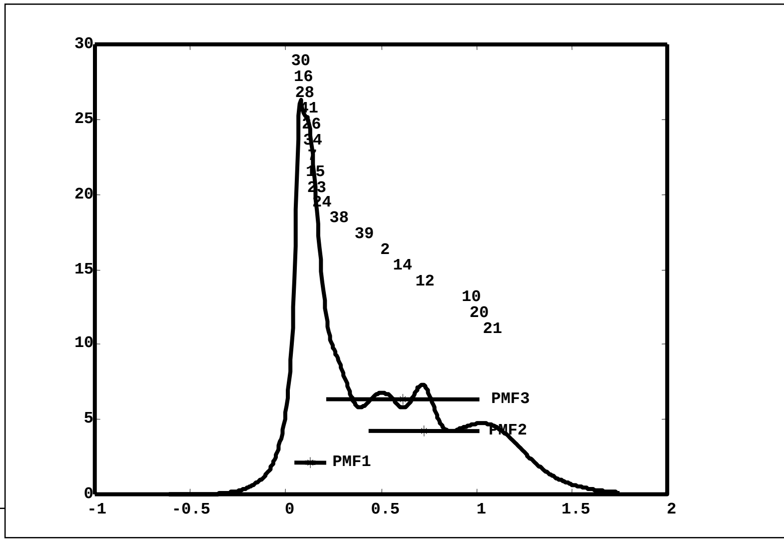
Tri-plot showing structures in data



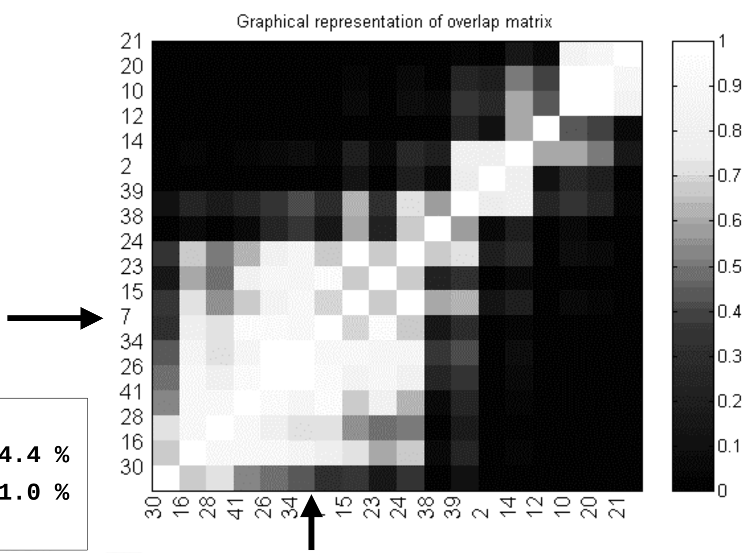
GMO in flour – Overview of data



GMO in flour – Overall measurement function



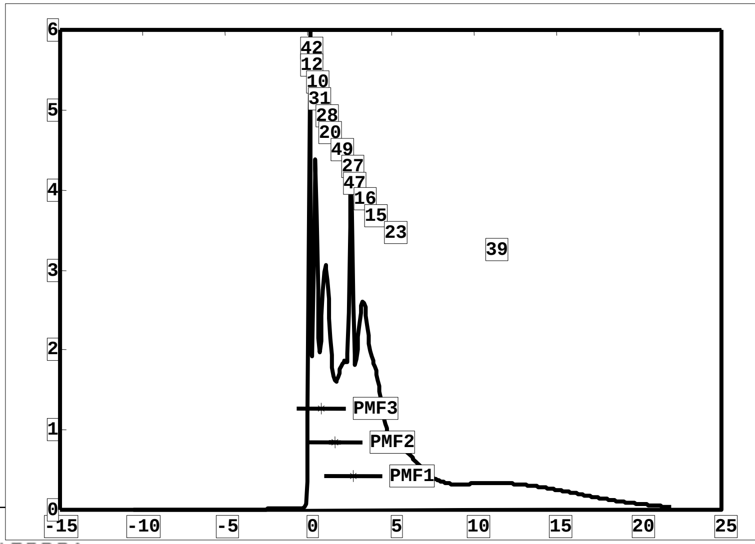
GMO in flour:graphical representation of overlap matrix



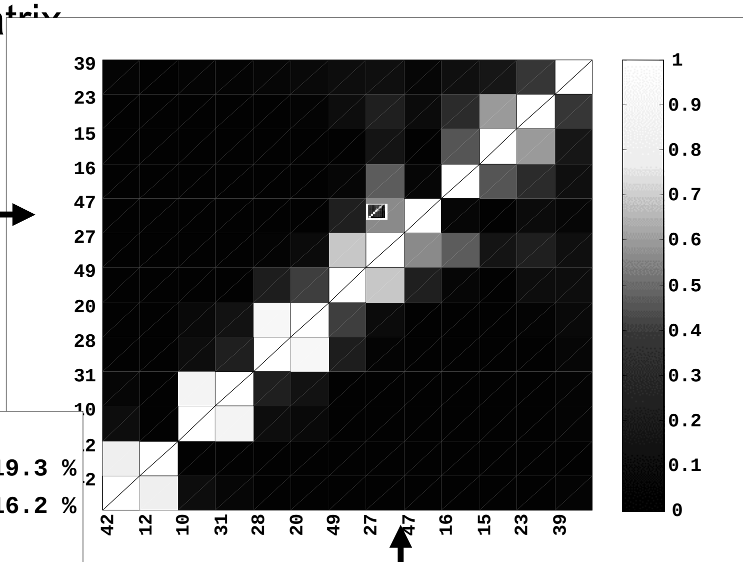
PMF1: 44.4 %

PMF2: 21.0 %

GMO 2 – Overall measurement function



GMO -2, Graphical representation overlap matrix



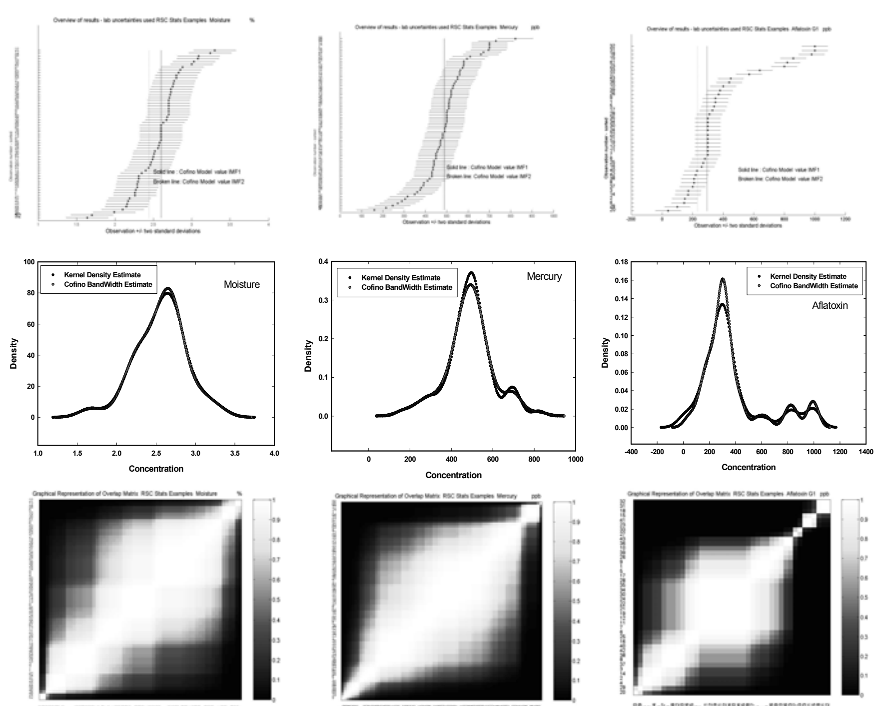
PMF1: 19.3 %

PMF2: 16.2 %

Performance NDA and BWE

Determinand	NObs	AMC[1] Estimates	AMC[1] Method
Moisture	61	2.29 (0.03) 2.67 (0.04) 3.06 (0.5)	Bootstrap-trimmed Bootstrap-trimmed Bootstrap-trimmed
Mercury	77	489.0 (10.6)	RM
Aflatoxin	42	328 (25) 283 (17)	RM RM-trimmed data

ALTERRA
WAGENINGEN UR



Including *less than* values

- Approach
  - Basisfunctions based on normal distributions for data above Limit of Determination (LOD)
  - For ‘less than’values, every value less than the LOQ is equally probable >> ***we can use a rectangular pdf !!!***

Paper accepted

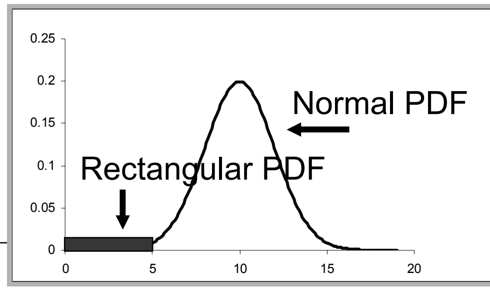
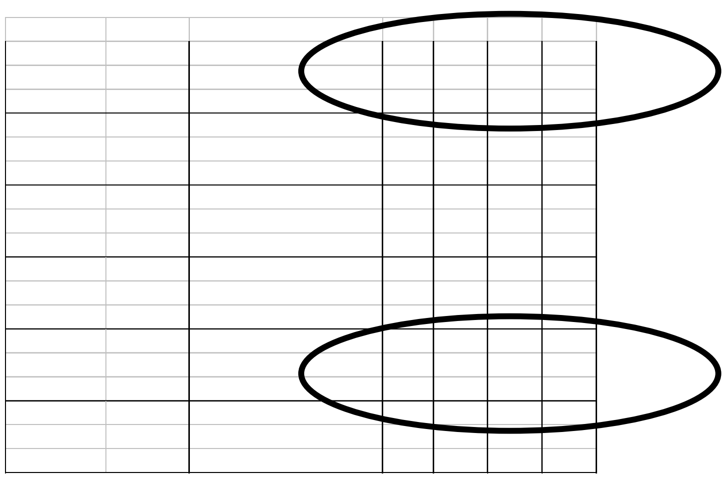
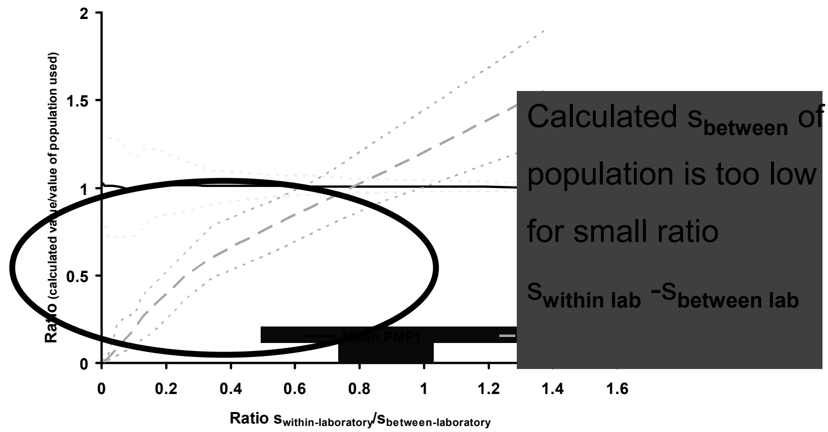


Illustration using Quasimeme datasets



Recent Comment Fearn (AMC-SMC) - background



Comments Fearn - remarks

- Effect already noted in original paper
- Effect may occur when model is used using supplied or estimated (individual) data-uncertainties
- Model has several features to identify occurrence of this effect (graphs, λ,…)
- Occurrence may point to analytical problems (e.g. GMO example in this presentation)
- Effect does **not** apply for NDA or BWE approaches
  - Paper of Fearn suggests differently!
- Not fruitful to create competition between AMC-KDE and this model!

Conclusions

- New model has been developed, can be implemented in many ways
  - Different shapes of pdf's of observations
  - Provided or estimated uncertainties
  - Without any uncertainty estimates using NDA or BWE
- Model can cope with very difficult distributions
- Model can deal with less than values
- Strong visualisation tools

Model is upon request available as Matlab toolbox

ALTERRA
WAGENINGEN UR

Statistical evaluation of proficiency test data with the software ProLab

Steffen Uhlig
Quo data GmbH, Siedlerweg 20, 01465 Dresden-Langebrück.
Email: Uhlig@quodata.de

Abstract

A statistical approach for the analysis of proficiency testing schemes using efficient and robust statistical methods in accordance with the German standard DIN 38402 A 45 is presented. The presentation includes also a demonstration of the software program ProLab for the performance of the calculations.
In the first calculation step the standard deviation between the laboratories is estimated by applying the new Q method. This method is based on the pairwise differences between single test results and requires no target value (mean value) as almost all other methods. The robust standard deviation is then used to calculate the mean value applying the robust Hampel estimator. Robust mean and robust standard deviation describe to some extent the overall competence of the laboratories. In order to assess single test results, they may also be used to derive Z scores and corrected Z_U scores for single laboratories. It is demonstrated that Z scores and Z_U scores can additionally be used to perform more comprehensive statistical analysis, e.g. to assess the laboratory competence over several samples, for several analytical methods or for several analytes.

Kurzfassung

Es wird ein statistischer Ansatz zur Analyse von Daten von Laborvergleichsuntersuchungen vorgestellt, der auf der Anwendung effizienter und robuster statistischer Methoden auf der Grundlage von DIN 38402 A 45 basiert. Gegenstand der Präsentation ist außerdem das Softwareprogramm ProLab, mit dem die statistischen Berechnungen durchgeführt werden können.
Im ersten Berechnungsschritt erfolgt die Ermittlung der Standardabweichung zwischen den Laboratorien mittels der neuen Q-Methode. Diese Methode basiert auf den paarweisen Differenzen zwischen den einzelnen Messergebnissen und erfordert keine vorherige Berechnung eines Ziel- oder Mittelwertes, wie dies ansonsten bei fast allen Methoden der Fall ist. Die robuste Standardabweichung wird erst danach genutzt, um den Mittelwert mittels des Schätzverfahrens von Hampel zu berechnen. Die beiden robusten Kennwerten für Mittelwert und Standardabweichung lassen sich nun einerseits nutzen, um – mit gewissen Einschränkungen - die generelle Kompetenz der Laboratorien zu charakterisieren. Im Rahmen von Laborvergleichsuntersuchungen jedoch lassen sie sich auch zur Berechnung von Z- bzw. Z_U–Scores und damit zur individuellen Bewertung der Messungen nutzen. Es wird gezeigt, dass Z- bzw. Z_U–Scores zusätzlich zur Durchführung komplexerer statistischer Analysen herangezogen werden können, z.B. zur Abschätzung der Laborkompetenz über verschiedene Proben, verschiedene Methoden oder verschiedene Analyten.

Statistical evaluation of proficiency test data with the software program ProLab

PD. Dr. habil. Steffen Uhlig

Gesellschaft für Qualitätsmanagement und Statistik mbH

Overview

1. Z scores
2. Decision matrix
3. Minimum requirements on statistical techniques
4. Determination of the standard deviation
5. Determination of the assigned value
6. Corrected z scores
7. The next level
8. ProLab: An overview

z scores

$$\text{z score} = \frac{\text{measurement value - assigned value}}{\text{standard deviation}}$$

?

|z score| < 2 with ≈95% probability for a lab with mean competence

How to use z scores for several samples/analytes?

Decision matrix

Decisions on a lab in an accreditation or certification scheme might be wrong:

Reality: *Decision:*	High competence (in the future)	Low competence (in the future)
Accept the lab	Ok	**Risk of customer**
Reject the lab	**Risk of lab**	Ok

Tradeoff between lab risk and customer risk

Decision matrix

The higher the risk for the lab, the higher the analytical costs

(since competent labs will disappear from the market)

Tradeoff

The higher the risk for the customer, the higher the environmental/health costs

(since wrong measurement results will cause wrong decisions)

Minimization of <u>both</u> risks requires the application of efficient statistical techniques

Minimum requirements on statistical techniques I

- Reproducibility of the method?
  Can the method be reproduced exactly?
- Interpretability of results?
  Can the results be interpreted in terms of an appropriate model?
- Applicability of the underlying statistical model?
  Appropriate description of reality?

Minimum requirements on statistical techniques II

- Statistical techniques validated?
  - Standard error and bias under appropriate model assumptions, in order to allow comparisons between **different** approaches
  - Consistency
  - Efficiency

➡ Statistical methods for the evaluation of interlaboratory tests for method validation such as ISO 5725-2 are not appropriate: the model is not appropriate!

Determination of the standard deviation I

The standard deviation may be:

- a target value, e.g. 15% of the assigned value (according to practical needs)
- derived from an appropriate variance function (like the Horwitz function)

Advantages:

- cannot be affected by varying laboratories competence.
- can be determined according to the needs of the user

Determination of the standard deviation II

Disadvantages:

- matrix effects not taken into account
- heterogenity of samples not taken into account

➡

The decision matrix may be ***biased***, and there might be a high percentage (eg. >50%) of laboratories not fulfilling the tolerance criterion.

Determination of the standard deviation III

The standard deviation may alternatively be calculated as an empirical, robust standard deviation:

- DIN 38402 A 45 recommends the Q method: highly efficient and highly robust (much better than MAD!).
- If there are several samples, DIN 38402 A 45 allows to calculate an empirical variance function based on the robust standard deviation.
- If the number of laboratories is not too small (and if more or less all laboratories have to take part in the proficiency test), the robust standard deviation is only slightly affected by varying lab competences.

Determination of the standard deviation IV

A compromise:

- DIN 38402 A 45 allows to limit the robust standard deviation in order to fit for the purpose: Very low standard deviations and extremely high standard deviations can be avoided.

Determination of the standard deviation V

Example: 4 laboratories only: 5 5.5 7 100

1) The pairwise differences:

0.5	1.5	93
	2	94.5
		95

2) The 25th percentile of these differences

0.5 **1.5** 2 93 94.5 95

3) Q_n = 2.1914 * 1.5 = 3.23

Rousseeuw/Croux (1993), Uhlig (1997a), Uhlig (1997b), Müller/Uhlig (2001)

Determination of the assigned value I

The assigned value may be:

- a reference value
- a (robust) mean

Determination of the assigned value II

Methods for calculating the robust mean:

Several variants of Huber estimators (highly efficient, but more or less robust: some variants may break down when 3 out of 12 laboratories are outliers!): originally proposed by P. Lischer, Switzerland

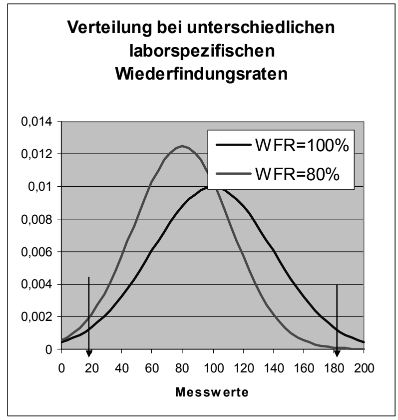
Disadvantage of Huber-type estimators: depending also on extreme outliers!

DIN 38402 A 45: Hampel estimator in connection with the Q method: highly efficient and highly robust. Very similar to Huber-type estimators, but not depending on extreme outliers.

z_u scores

Example: Overall mean=100, close to the determination limit.
Standard deviation = 40,
Tolerance limits =20 and 180

Then: Labs with lower recovery are preferred by an evaluation system based on z scores:
With 100% recovery the z score is falling out of the tolerance interval with 5% probability.
With 80% recovery the z score is falling out of the tolerance interval with 3.1% probability.



z_u scores

Alternative: corrected z scores (Uhlig/Henschel 1997)

$$z_U = \begin{cases} k_1 \cdot z & \text{if} \quad z < 0 \\ k_2 \cdot z & \text{if} \quad z \geq 0 \end{cases}$$

with k_1 <1 and k_2 >1. k_1 and k_2 can be determined so that there is no more advantage for labs with deviating recovery.

➡ DIN 38402 A45

The next level

Use PT data in order to

- assess sources of error (e.g. by PCA)
- calculate the lab uncertainty (see Uhlig/Lischer 1998)
- assess different analytical methods

References

Müller, C. and Uhlig, S. (2001): **Estimation of variance components with high breakdown point and high efficiency.** *Biometrika, 2001*

Rousseeuw, P.J. and Croux, Chr. (1993): **Alternatives to the Median Absolute Deviation.** *Journal of the American Statistical Association 88, 1273-1283.*

Uhlig, S. (1997a): **Robust estimation of variance components with high breakdown point in the 1-way random effect model.** *Industrial Statistics. Hrsg. C.P. Kitsos und L. Edler. Physica-Verlag, 65-73.*

Uhlig, S. (1997b): **Robuste Schätzung in Kovarianzmodellen mit Anwendungen im Qualitätsmanagement und in Finanzmärkten.** *Habilitation thesis. Free University Berlin.*

Uhlig, S. and Henschel, P. (1997): **Limits of tolerance and z-scores in ring tests.** *Fresenius J Anal Chem., 761-766.*

Uhlig, S. and Lischer, P. (1998): ***Statistically based performance characteristics in laboratory performance studies.*** *The Analyst , 167-172.*

Assessment of the QUASIMEME Laboratory Performance Studies using Robust Statistics and the Cofino Model

David E. Wells and Judith A. Scurfield
QUASIMEME Project
FRS Marine Laboratory
375 Victoria Road
Aberdeen AB11 9DB

Abstract

At the outset of the QUASIMEME Laboratory Performance Studies in 1992, data were assessed using Robust Statistics in an attempt to overcome the effects of non-normality of many of the data sets. In some cases the level of tailing and skewed data was such that it was also necessary to trim the data (|Z| < 6) to reduce the effects of extreme values.

Since 1999 QUASIMEME has incorporated the Cofino- Model into the data assessment to provide a more comprehensive and informative evaluation. The Cofino Model, in a single evaluation, can deal with difficult data that are multi-modal and contains outliers and tailing values. It can also include Left Censored Values ('*less than*'). The graphical output provides a clear representation of the data using distribution profiles, ranked distributions and data overlap matrix plots. The models have been tested on over 5000 datasets and a series of comprehensive examples demonstrate the robustness of these evaluation tools.

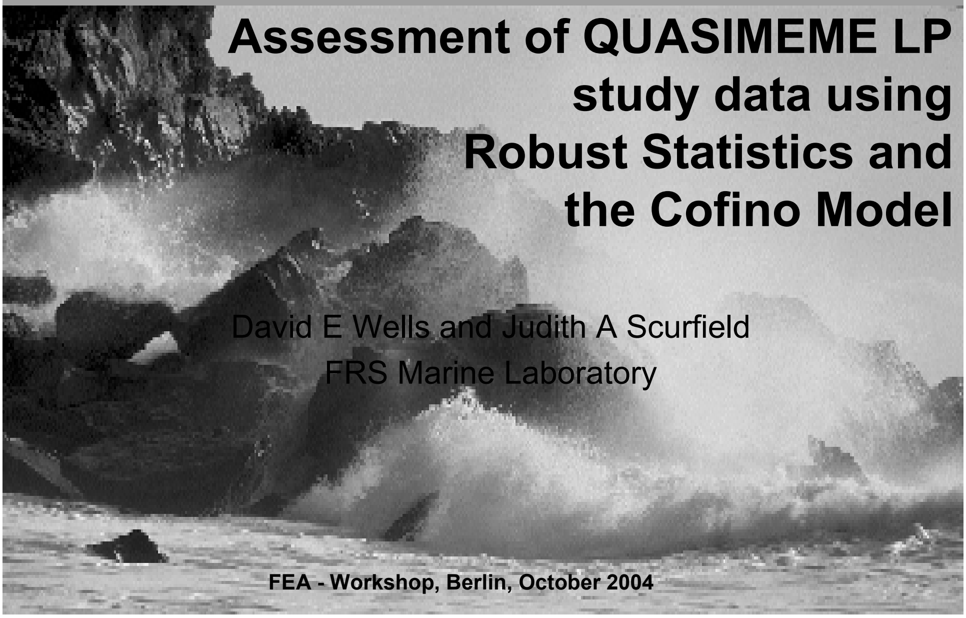
The data include chemical measurements of nutrients, trace metals, organochlorines and PAHs in seawater, biota and sediment which are required for most national and international marine monitoring programmes. A report and CD is available.

Kurzfassung

Am Anfang der Studien zur QUASIMEME-Laborleistung 1992 wurden Daten unter Verwendung robuster Statistiken ausgewertet, in dem Bestreben Effekte der Nicht-Normalverteilung vieler Datensätze zu überwinden. In einigen Fällen war das Niveau der unsymmetrischen Verteilung und der Verzerrung von Daten so, dass es notwendig wurde, die Daten zu eliminieren (|Z|< 6) um eine Reduzierung der Effekte extremer Werte zu erreichen.

Seit 1999 wurde bei QUASIMEME das Cofino-Modell in die Dateneinschätzung miteinbezogen, um eine umfassende und aussagefähige Bewertung zur Verfügung zu stellen.
Mit dem Cofino- Modell, angewandt auf die einzelne Auswertung, können schwierige Daten bearbeitet werden, die multi-modal sind und sowohl Ausreißer als auch die zur Normalverteilung rechtsschiefen Werte enthalten. Ebenso werden die „kleiner als“ Daten berücksichtigt. Das graphische Ergebnis bietet eine klare Darstellung der Daten unter Verwendung von Verteilungsprofilen, abgestuften Verteilungen und Matrix Plots. Das Modell wurde an über 5000 Datensätzen getestet und eine Serie umfassender Beispiele zeigt die Robustheit dieses Auswertungsverfahrens. Die vorliegenden Ergebnisse beinhalten chemische Messungen von Nährstoffen, Spurenmetallen, Chlororganika und PAKs in Meerwasser, Flora und Fauna und in Sedimenten, die für die meisten nationalen und internationalen Überwachungsprogramme der Meeresumwelt gefordert werden. Ein Report und eine CD sind vorhanden.

FISHERIES RESEARCH SERVICES QUASIMEME



Assessment of QUASIMEME LP study data using Robust Statistics and the Cofino Model

David E Wells and Judith A Scurfield
FRS Marine Laboratory

FEA - Workshop, Berlin, October 2004

Principles of Data Assessment

- Based on International Standards
- Equally fair to all participants
- Accurate
- Transparent and traceable methods of assessment
- Confidential results
- QUASIMEME operates to international standards ISO 43 and G 13:2000

Methods of obtaining the Assigned Value

- Use CRM or well characterised material
- Nominal concentration from preparation
- Reference laboratories
- Consensus value of participants data

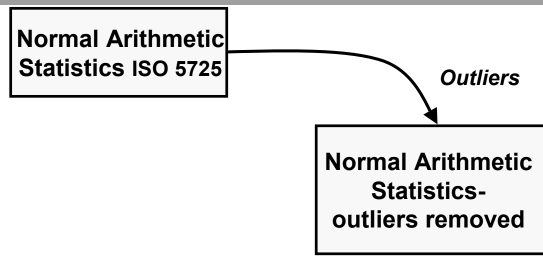
QUASIMEME is a project of Fisheries Research Services

Consensus Values

- Data always available
- Consistent method can be applied to all studies
- In general the bias of a group of reliable laboratories is ca =0 (observation)
- Acceptable approach since the Assigned Value is based on participants data

QUASIMEME is a project of Fisheries Research Services

Development of Data Assessment Statistics 1

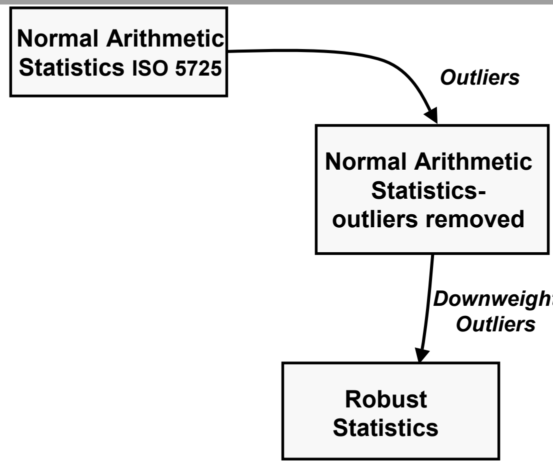


Historical Development of the Assessment

- **ISO 5725 based on a normal distribution is not suitable for most LP studies**
- **Most LP study data has tailing, multi-modal or skewed distributions**
- **The key to sound assessment is a reliable Assigned Value**

QUASIMEME is a project of Fisheries Research Services

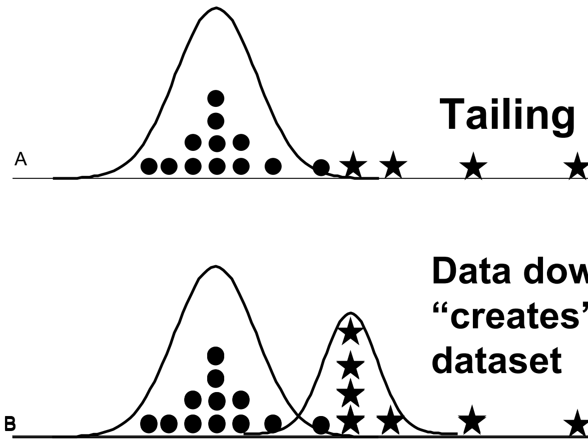
Development of Data Assessment Statistics 2



Historical Development of the Assessment

- QUASIMEME used Robust statistics on trimmed data to obtain the Assigned Value (1992-1999)
- Robust statistics can cope with up to ca 7-10% tailing data
- Robust statistics cannot cope with bi-modal distributions

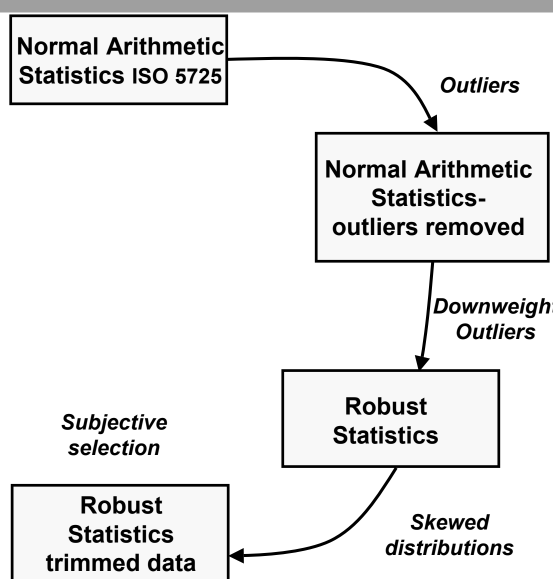
Robust Statistics



Tailing data

Data down-weighted “creates” a bimodal dataset

Development of Data Assessment Statistics 3

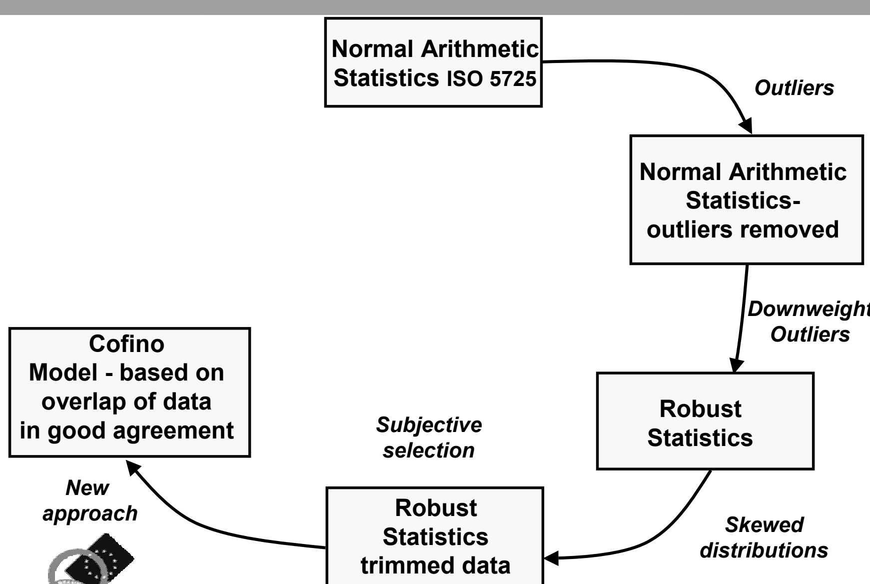


QUASIMEME is a project of Fisheries Research Services

Robust Statistics

- Many examples where Robust Statistics do not provide the best population characteristics
- Outlier tests not appropriate to skewed or bimodal data
- *Trimming* data is subjective
- Require a *non-subjective* method

Development of Data Assessment Statistics 4

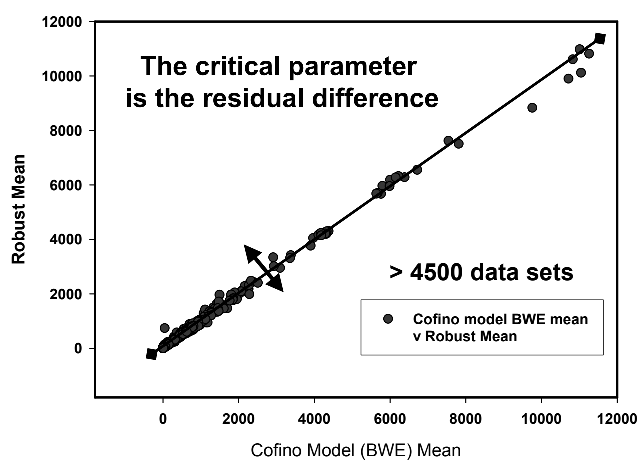


Evaluation of Robust Statistics and Cofino Model using QUASIMEME data (1996 - 2004)

- **Over 4,500 data sets**
- **Nutrients, Trace Metals, Organochlorine & Organophosphorous Pesticides, CBs, PAHs**
- **Seawater, fish & shellfish, sediment**
- **No. of Observations (Nobs) 5 - 70**

QUASIMEME is a project of Fisheries Research Services

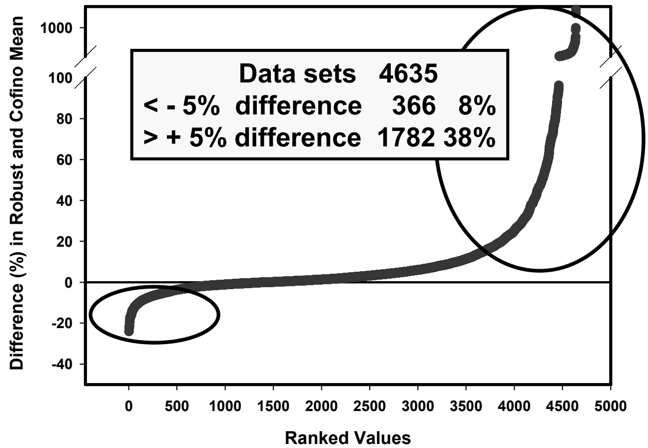
Comparison of Robust & Cofino Means QUASIMEME LPS data 1996-2004



QUASIMEME is a project of Fisheries Research Services

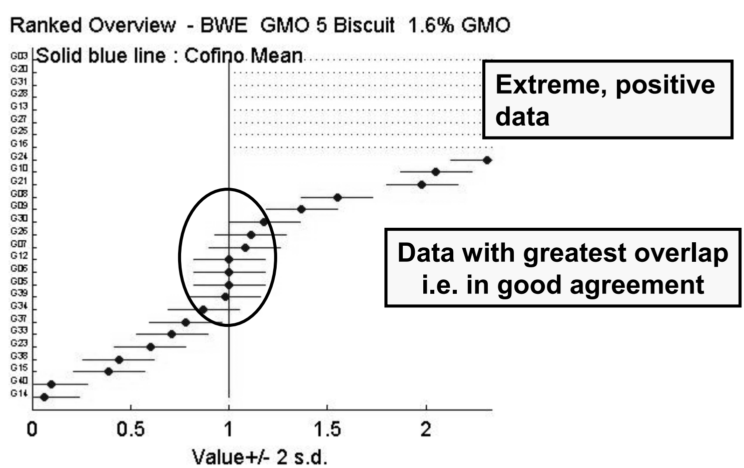
Comparison of Robust & Cofino Means QUASIMEME LPS data 1996-2004

Small differences in the mean (Assigned Value) can significantly affect the assessment of a laboratory



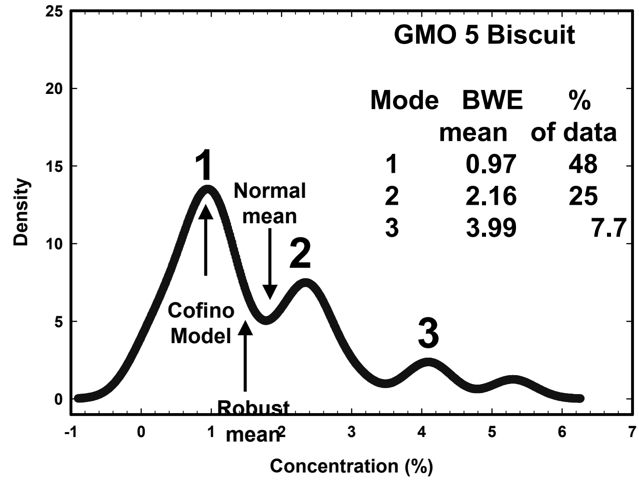
QUASIMEME is a project of Fisheries Research Services

Comparison of Robust Statistics and Cofino Model - GMO in biscuit

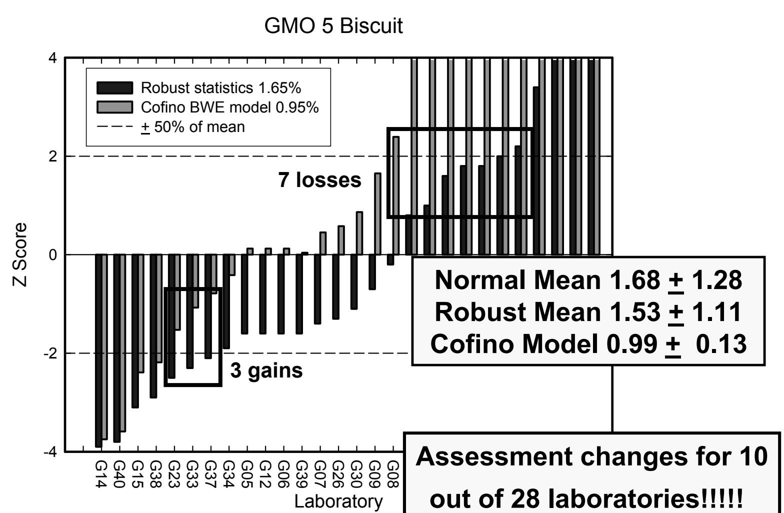


QUASIMEME is a project of Fisheries Research Services

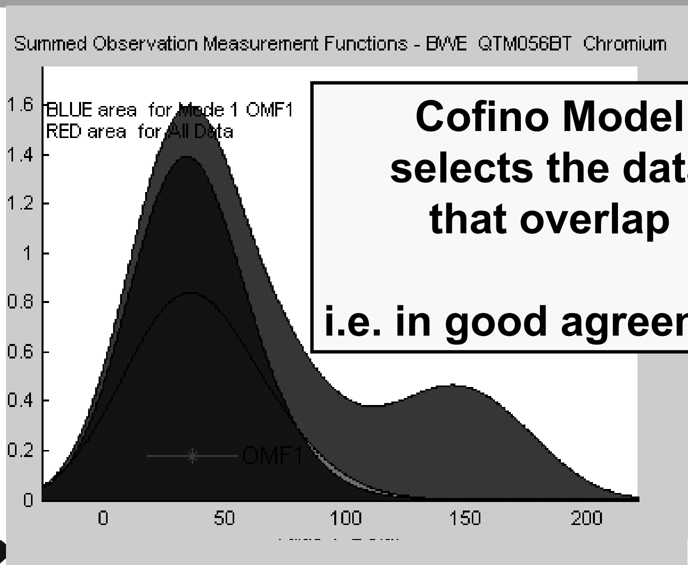
Population Density Function
GMO in biscuit



Assessment by Robust Statistics or Cofino Model? - GMO in biscuit



Selection of data in *good agreement*

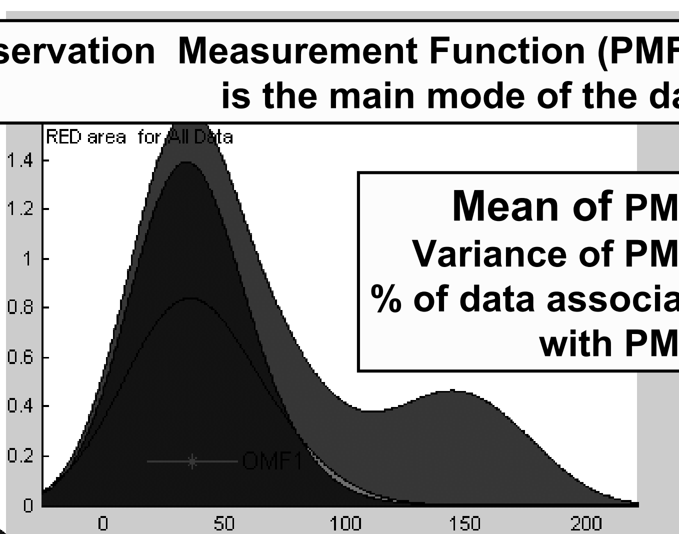


Cofino Model selects the data that overlap

i.e. in good agreement

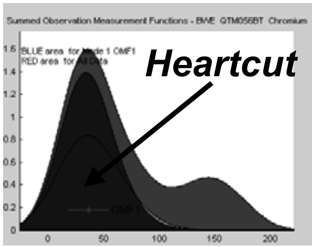
Cofino Model. What does it actually provide?

Observation Measurement Function (PMF_1) is the main mode of the data



Mean of PMF_1
Variance of PMF_1
% of data associate with PMF_1

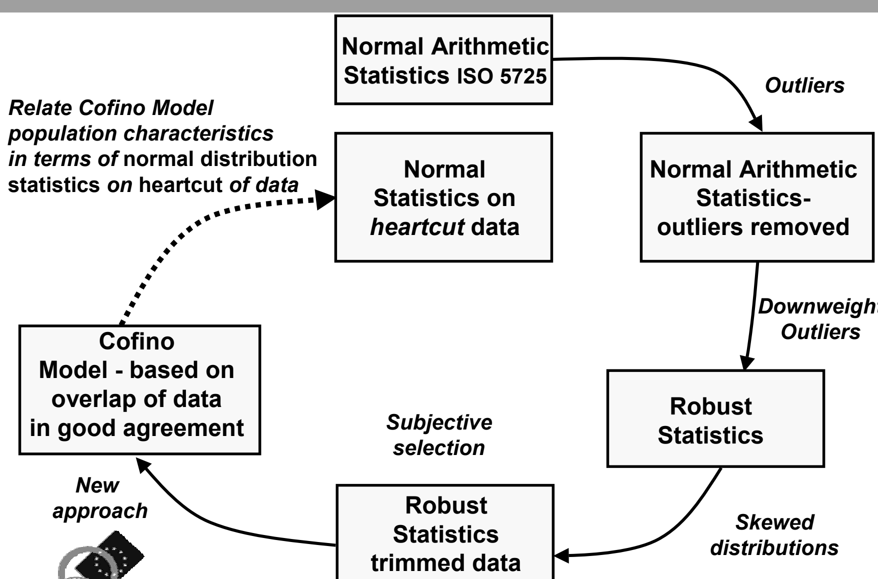
Cofino Model. What does it actually provide?



- **Mean of $PMF_1 \sim \underline{=}$ Mean of *heartcut* of data**
- **Variance of $PMF_1 \sim \underline{=}$ Variance of *heartcut* of data**
- **% of data associate with $PMF_1 \sim \underline{=}$ Data in the *heartcut***

Cofino model characteristic relate to all data
What are the *normal* statistics
of the equivalent *heartcut* of data?

Development of Data Assessment Statistics 5



Comparison of Cofino Model characteristics & Arithmetic Statistics of *heartcut* of data

Obtain characteristics of main mode (PMF_1)

↓

Rank observations by ABS(x_i - mean $_{PMF1}$)

↓

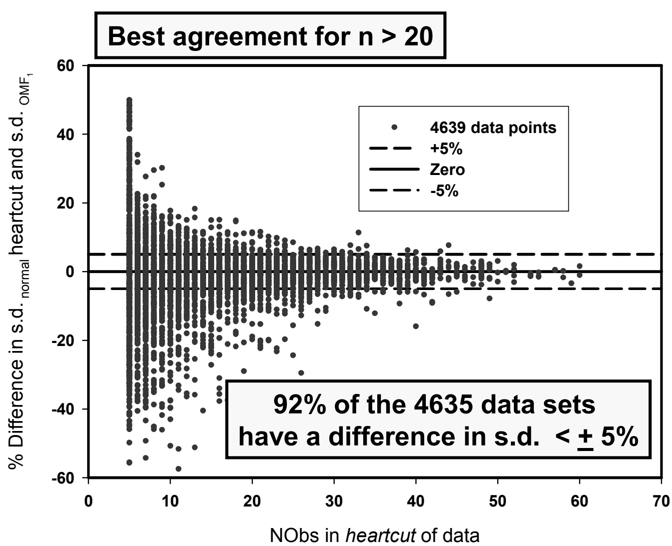
Select *heartcut* NObs where (s.d. $_{Normal}$ of *heartcut* - s.d. $_{PMF1}$) is minimum

↓

Compare!

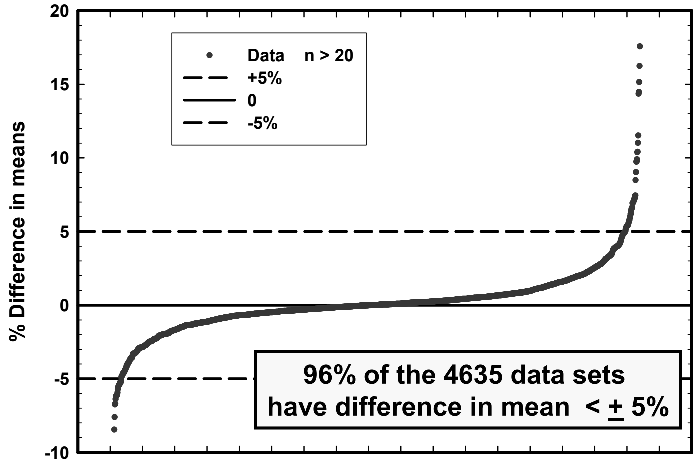
QUASIMEME is a project of Fisheries Research Services

Difference between s.d.$_{normal}$ of *heartcut* and s.d. $_{PMF1}$ in relation to NObs



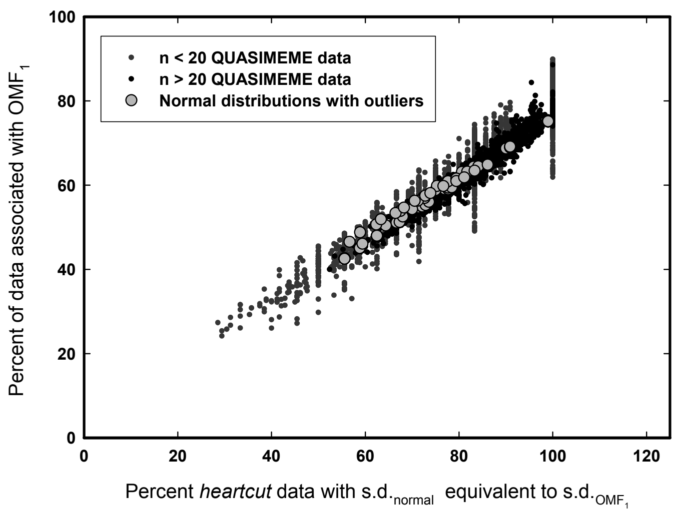
QUASIMEME is a project of Fisheries Research Services

Difference between mean.$_{normal}$ of *heartcut* and mean. $_{PMF1}$ (ranked)



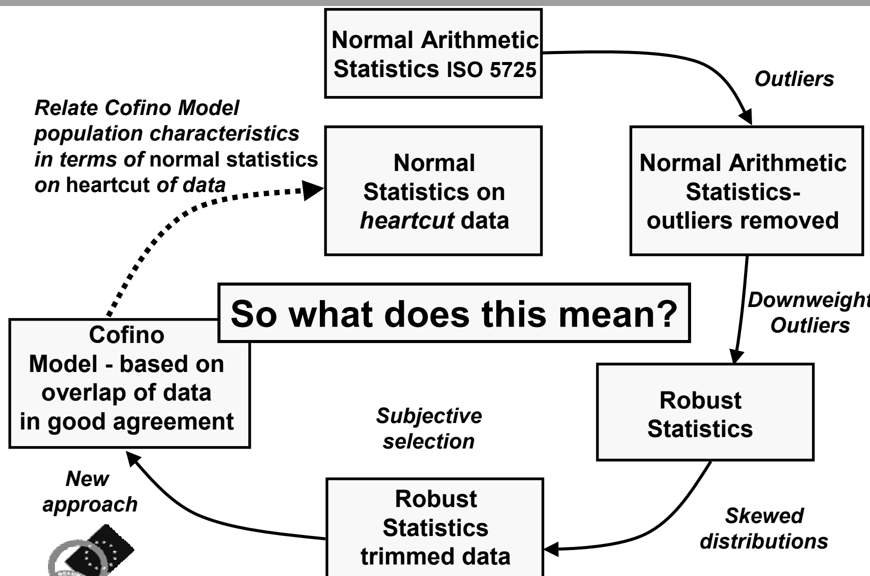
QUASIMEME is a project of Fisheries Research Services

Comparison between % data in the *heartcut* and % data associate PMF1



QUASIMEME is a project of Fisheries Research Services

Development of Data Assessment Statistics 6



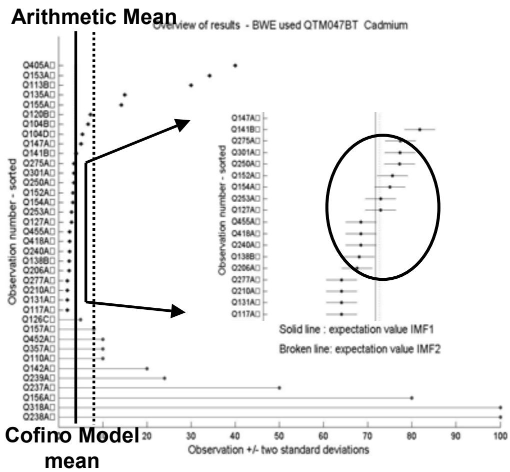
QUASIMEME is a project of Fisheries Research Services

Development of Data Assessment Statistics 6

- **Relates the Cofino Model to established statistics (for comparison)**
- **Cofino Model provides population characteristics equivalent to the Normal Distribution characteristics of the *heartcut* of data**
- **The *heartcut* of data belong to laboratories in good agreement (by definition!!)**
- **Circular argument for confirmation!**

QUASIMEME is a project of Fisheries Research Services

Percentage of data in the Model Mean



Cd in flounder muscle

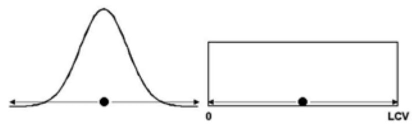
- **Cofino mean = 2.62 $\pm$ 0.56 (n=37)**
- **Arithmetic mean 7.79 $\pm$ 10.5**
- **Only 31% data associated with first mode (i.e. mean)**
- **Model selects data that are in good agreement**

Cofino Model : Features of application

- **Inclusion of ”less than” values - Left Censored Values**
- **Coping with bimodality**
- **Separating population characteristics of different methodologies.**

QUASIMEME is a project of Fisheries Research Services

Cofino model- Left Censored Values (LCVs)

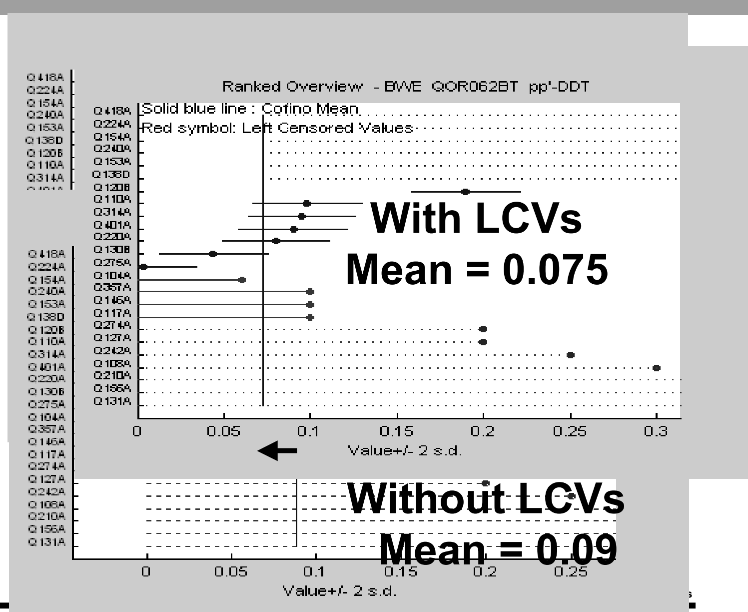


Numerical values Normal distribution

LCVs given a rectangular distribution

- Cofino Model provides a Liner combination of both distributions
- Limit the LCVs to < Cofino mean$_{nda}$
- High LCVs provide no information on the population characteristics

Effect of including LCVs for pp’DDT in biota

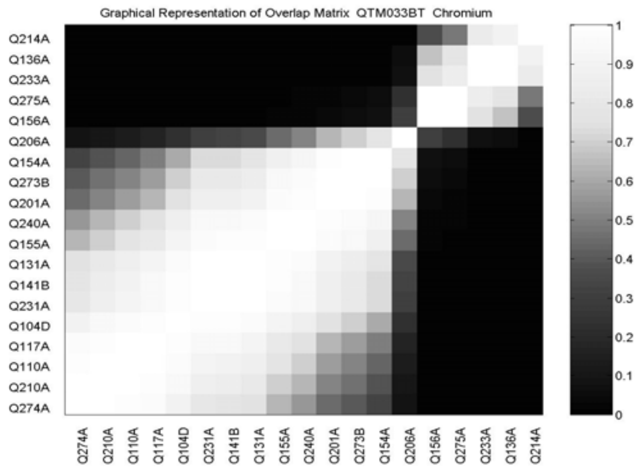


Bimodality - Chromium in biota

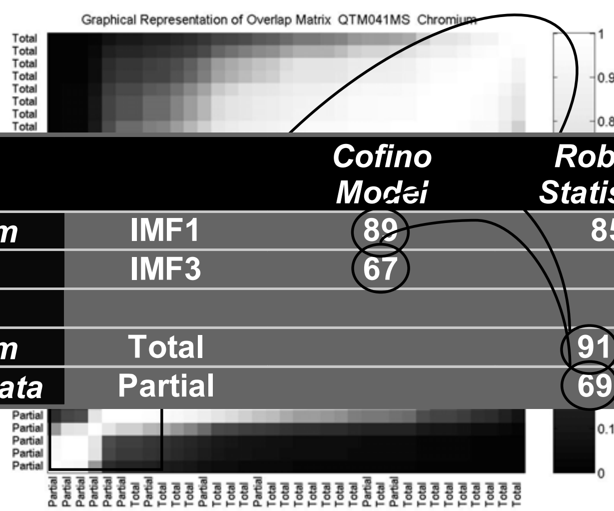
Ranked Overview

Distribution Plot (pdf)

Kilt Plot of Overlap Matrix



Separation of methods



		Cofino Model	Robust Statistics
Chromium	IMF1	89	85
	IMF3	67	
Chromium	Total		91.5
Separate data	Partial		69.1

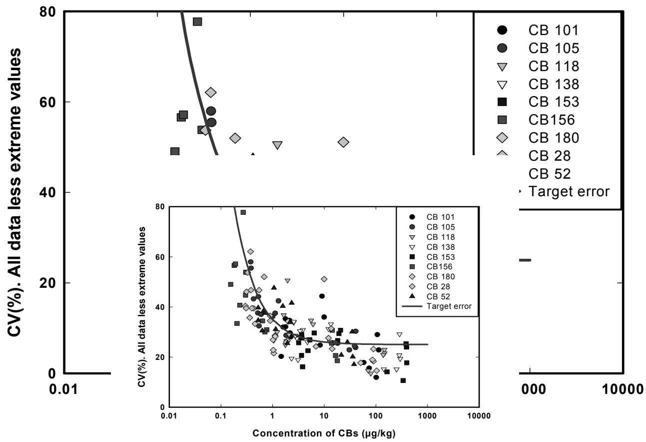
Cofino model- Summary

- Provide a reliable estimate of the Assigned Value
- Not affected by tailing, skewed or bi-modal data
- Can integrate numerical data and left censored values
- Can integrate and separate methodologies
- Provide summary of the population statistics and graphical representations to describe the data

Assessment of Laboratories - The Z Score

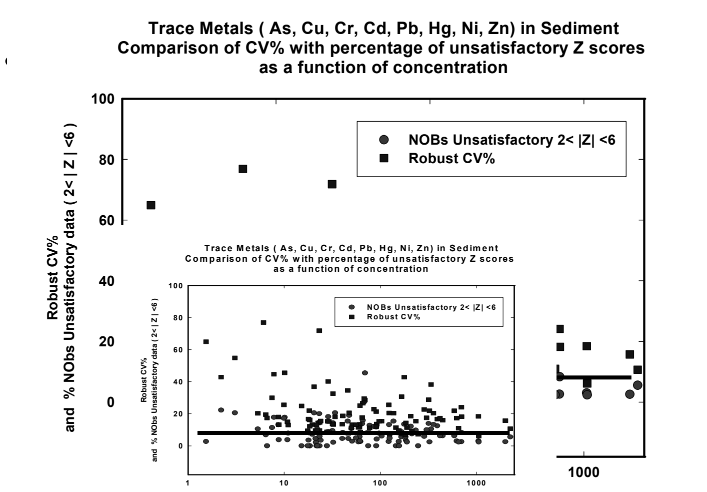
- Normalise data to a prescribed quality objective
- $Z = (x_i - X)/\sigma$ where x_i is the lab. Value, X is the AV, and σ is the allowable error
- The allowable error has a proportional and constant element
- $\sigma = (E_{prop} * AV) / 100 + 0.5 * E_{const}$
- Allows for poorer precision at lower concentrations
- Provides a constant assessment independent of concentration

Effect of Concentration on performance



QUASIMEME is a project of Fisheries Research Services

Performance with concentration
Trace Metals in Sediment



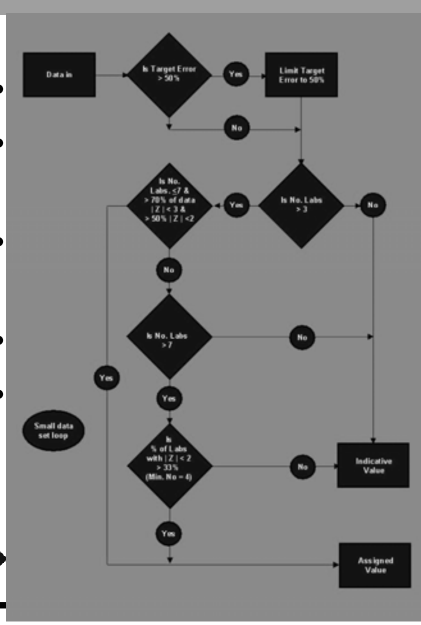
QUASIMEME is a project of Fisheries Research Services

Z Score Assessment Criteria based on ISO 43

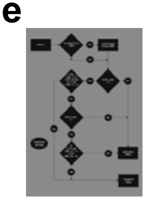
- |Z| < 2 Satisfactory performance
- 2 < |Z| < 3 Questionable performance
- |Z| > 3 Unsatisfactory performance
- Additional QUASIMEME categories
- |Z| > 6 Extreme: gross errors
- For Left Censored Values (LCVs)
- LCV/2 < |Z| $\underline{=}$ 3 Consistent with AV
- LCV/2 > |Z| $\underline{=}$ 3 Inconsistent with AV

QUASIMEME is a project of Fisheries Research Services

Guidelines: Assigned or Indicative Value



- ment are published
- sed on < n=4
- 7 70% data |Z| <3
  nd 50% data |Z| <2
- Z| <2 (n_{min} = 4)
- se criteria the



UASIMEME is a project of Fisheries Research Services

QUASIMEME Report of the Studies

- Summary description of the study
- Summary Statistics Tables
- Z score plots
- Cofino plots
  - Population Measurement Functions
  - Ranked distribution
  - Kilt Plot of the Overlap Matrix
- Confidential Laboratory Assessment Certificate

Encouragement from Garfield……

Know the quality of your data before others tell you!!

QUASIMEME is a project of Fisheries Research Services

Thank you……………..

- **To all participating laboratories**
  - **for all your hard work**
- **Wim Cofino for all the fun of extending the model into QUASIMEME**
- **For the QUASIMEME Team Judith Scurfield, Kerry Smith, Laura Pattison, Brid O'Shea**

QUASIMEME is a project of Fisheries Research Services

Proficiency Testing Schemes in the Field of Drinkingwater Surveillance
– State of the Art and New Developments

DR. ULRICH BORCHERS
IWW Rheinisch Westfälisches Institut für Wasser – Beratungs- und Entwicklungsgesellschaft mbH
Moritzstraße 26, D-45470 Mülheim an der Ruhr, GERMANY, u.borchers@IWW-online.de

Abstract

This contribution highlights the state of the art of German interlaboratory comparison schemes in the field of the national drinkingwater surveillance. According to the German Drinkingwater Regulation (TrinkwV 2001) all laboratories are obliged – beneath all other requirements in conjunction with accreditation - to participate in external quality assurance systems successfully and on a regular basis.

On a regularly basis and successfully part at external. The requirements for this kind of proficiency testing programmes (PTS) are nationally harmonised by a recently published FEA recommendation [1]. The concrete planning, execution and evaluation of interlaboratory comparisons for proficiency testing of laboratories are standardised by the standard DIN 38402-45. This German standard expands the ISO Guide 43 in the field of PTS for water analysis and is now being transferred to an ISO standard [2].

The results of this interlaboratory trials can help to demonstrate that the binding method performance criteria given by the German Drinkingwater Regulation (accuracy, precision and LOD) are met. "*Independent bodies*" according to § 15(5) TrinkwV 2001 are responsible for the compliance checking of the FEA recommendations [1].

An overview of the harmonised Drinkingwater PTS in Germany will be given and the performance of drinkingwater labs will be demonstrated by the means of some relevant parameters. Finally, a case study for the realisation of the traceability of results of an interlab trial for relevant heavy metals will be presented.

Zusammenfassung

Dieser Beitrag beleuchtet den derzeitigen Stand der Trinkwasserringversuche in Deutschland im Bereich amtlicher Trinkwasserüberwachung. Gemäß der Deutschen Trinkwasser-verordnung (TrinkwV 2001) sind alle Laboratorien dazu verpflichtet – neben all den anderen Anforderungen im Zusammenhang mit der Akkreditierung – regelmäßig und erfolgreich an Programmen zur externen Qualitätssicherung teilzunehmen. Die Anforderungen an diesen Typ eines Qualitätssicherungsprogramms (PTS) sind national durch eine erst vor Kurzem veröffentlichte Empfehlung des UBA [1] harmonisiert worden. Die konkrete Planung, Durchführung und Auswertung von Ringversuchen zur Qualitätsüberprüfung von Laboratorien ist durch die DIN 38402-45 genormt. Diese Deutsche Norm ergänzt den ISO Guide 43 bezüglich der Qualitätsüberprüfung im Bereich der Wasseranalytik und wird derzeit in eine ISO-Norm überführt [2].

[1] Empfehlungen des Umweltbundesamtes: Empfehlung für die Durchführung von Ringversuchen zur Messung chemischer Parameter und Indikatorparameter zur externen Qualitätskontrolle von Trinkwasseruntersuchungsstellen - Empfehlung des Umweltbundesamtes nach Anhörung der Trinkwasserkommission des Bundesministeriums für Gesundheit und Soziale Sicherung beim Umweltbundesamt, Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz 2003; **46**: S. 1094–1095

[2] ISO/WD 20612 Water quality – Interlaboratory comparisons for proficiency testing of laboratories (ISO TC 147/SC 2 **N0695**)

Die Ergebnisse dieser Ringversuche können unter anderem dazu genutzt werden, die Einhaltung der von der Trinkwasserverordnung verbindlich vorgegebenen Verfahrenskenndaten (Richtigkeit, Präzision und Nachweisgrenze) nachzuweisen. Die „*unabhängigen Stellen*“ gemäß § 15 (5) TrinkwV 2001 sind dann dafür zuständig, die Erfüllung der in der UBA-Empfehlung [1] festgelegten Kriterien zu überwachen.

Es wird ein Überblick über die harmonisierten Trinkwasserringversuche in Deutschland gegeben und die Leistungsfähigkeit der Laboratorien anhand von ausgewählten Parametern demonstriert. Schließlich wird eine Möglichkeit für die Umsetzung einer Rückführung von Ringversuchsergebnissen auf nationale Normale aufgezeigt.

Possible ways for the realization of traceability of measurement results in environmental protection

Detlef Schiel
Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig

Abstract

Metrological traceability to internationally recognized reference points (national standards) is a prerequisite for comparability, accuracy and reliability of measurement results. The development of national standards and their provision for the link up with measurement procedures which are used for reference or routine measurements is the task of the national metrological institutions. In order to link up the large number of chemical laboratories in environmental protection efficient traceability structures are necessary.

In this lecture possible traceability structures will be presented and discussed.

Kurzfassung

Die metrologische Rückführung auf international akzeptierte Bezugspunkte (nationale Normale) ist eine Voraussetzung für die Vergleichbarkeit, Richtigkeit und Zuverlässigkeit chemisch-analytischer Messergebnisse. Die Entwicklung von nationalen Normalen und deren Bereitstellung zum Anschluss von Referenz- und Routinemessverfahren ist die Aufgabe der metrologischen Staatsinstitute. Für den Anschluss der großen Anzahl von Laboratorien aus dem Bereich der Umweltanalytik werden effiziente Rückführungsstrukturen benötigt.

In dem Vortrag werden mögliche Rückführungsstrukturen vorgestellt und diskutiert.

**Possible ways for
the realization of**

traceability of chemical measurement results in environmental protection

Detlef Schiel

Physikalisch-Technische Bundesanstalt

PTB - German Metrology Institute

- **Tasks :**

  Development and provision of the basis for traceability of measurement results (Realization and dissemination of the SI units)

- **Objectives :**

  Improvement of

  - Accuracy
  - Reliability
  - Comparability
  - Global acceptance

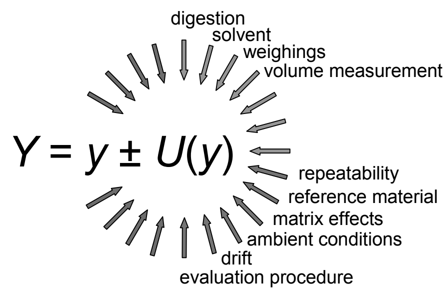
  of measurement results

Uncertainty

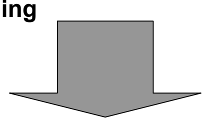
Reliable measurement result corresponding to the **Metrolgical Concept** includes:

Uncertainty
+
Traceability

Sample preparation



Sampling **Measurement**



National Standards (SI)

Traceability

Sample X

n (Cu, X)
Amount of copper
in a sample material X

$\frac{n(\mathrm{Cu, X})}{n(\mathrm{Cu, Y})} = \ldots$ Comparison by a routine method

Reference material Y

n (Cu, Y)
Amount of copper
in the RM Y

$\frac{n(\mathrm{Cu, Y})}{n(\mathrm{Cu, Z})} = \ldots$ Comparison by a reference method

Primary standard Z (national standard)

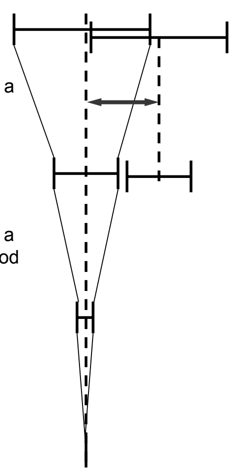
n (Cu, Z)
Amount of copper in a well
characterized material (solution) Z

$$n(\mathrm{Cu, Z}) = w_{\mathrm{pur}} * \frac{1}{M_{\mathrm{Cu}}} * m$$

SI

The mol is that amount-of-substance which contains as many entities as there are in 12 g ^{12}C. [*n*] = 1 mol

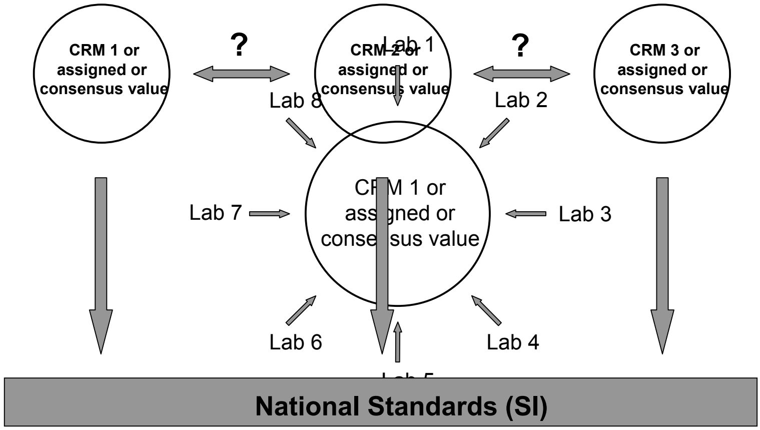
Uncertainty



Traceability *(after VIM):*
Property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties.

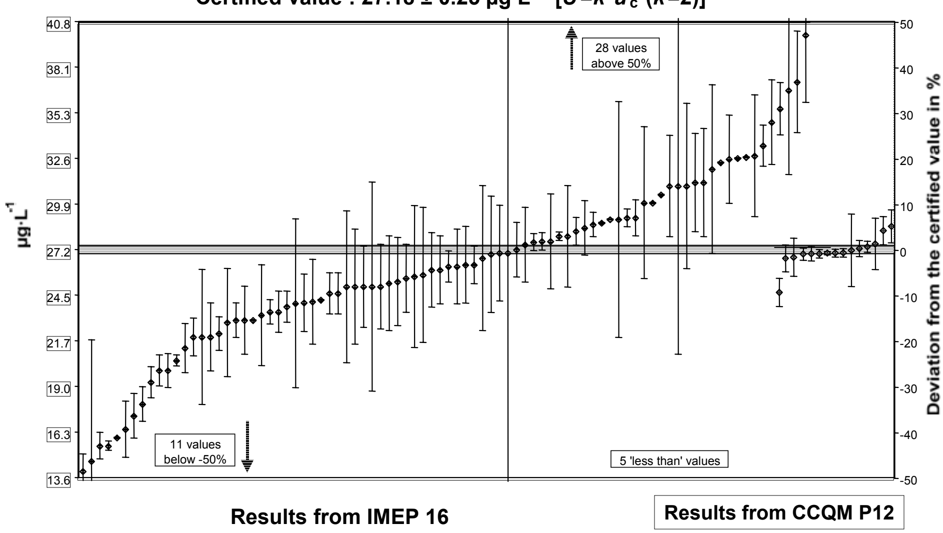
Physikalisch-Technische Bundesanstalt PTB

Comparability

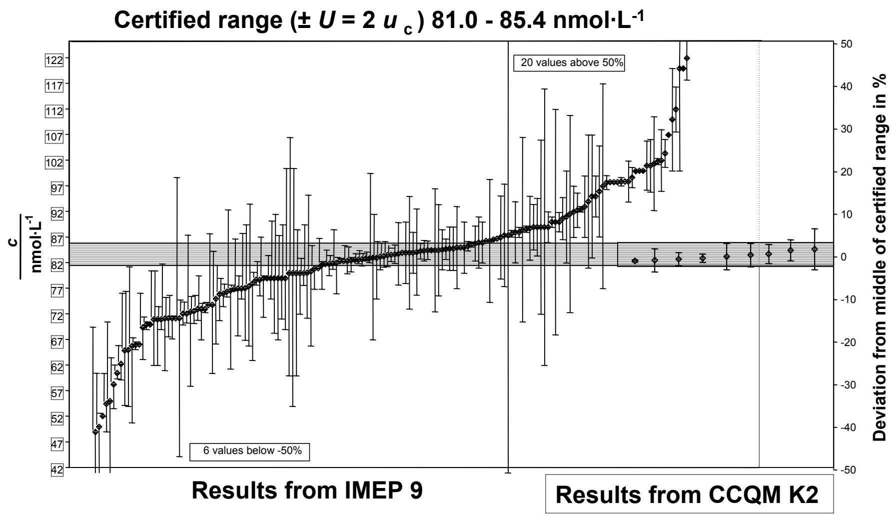


Pb in wine

Certified value : 27.18 ± 0.25 $\mu g \cdot L^{-1}$ [$U = k \cdot u_c$ (k=2)]



Cd in river water



Direct link up

Routine level

Certified reference materials

- At any time available for calibration or verification

But unfortunately:
- At present only a few reliable materials are on offer
- Expensive
- Are NMI's able to satisfy the demand?
- Logistic problem

National Standards

Direct link up

Routine level

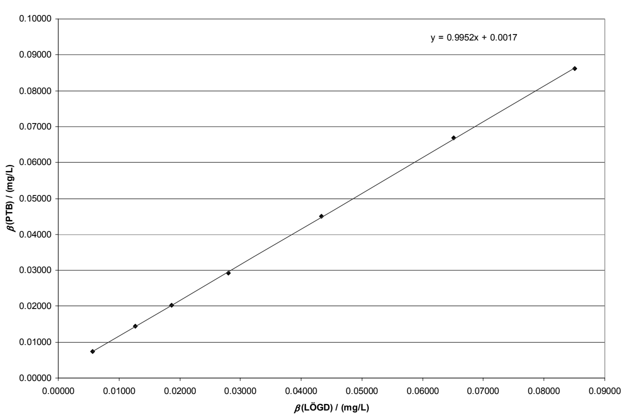
<u>Comparison measurements</u>
(with a reference value)

Efficient dissemination
- Almost all field laboratories can be reached
- improvement of the quality
- Use of established QS structure,
  metrological completition,
  an additional feature

But:
- Large certification work
- Rational methods are necessary

National standards

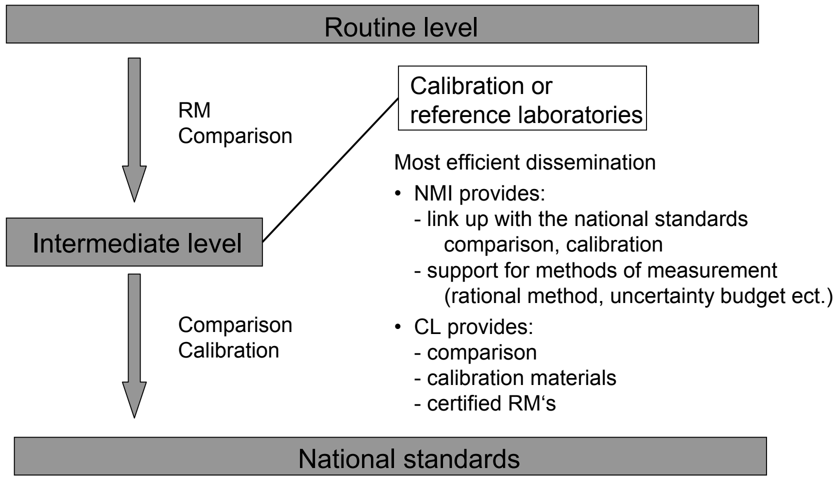
Example for the use of comparison measurements for the link up with national standards (Ni in drinking water)



Proposal for getting traceable values:

- metrological validation of the gravimetrical preparation procedure of the samples
- measurement of the blanks by IDMS

Link up via an intermediate level



Conclusion

Metology can
complete traceability structures
in environmental protection

- Need not to be expensive
- Available as a service of the NMIs

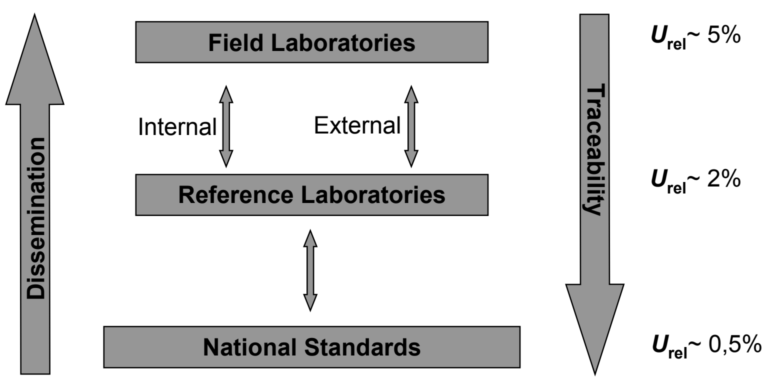
assures:

- National and international acceptance
- and comparability
- Reliability
- Improvement of the accuracy
- Strengthening of the confidence in environmental measurements

Thank's for your attention

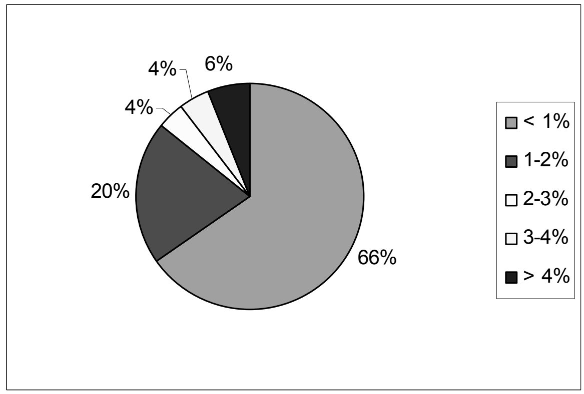
Physikalisch-Technische Bundesanstalt

Traceability structure in clinical chemistry

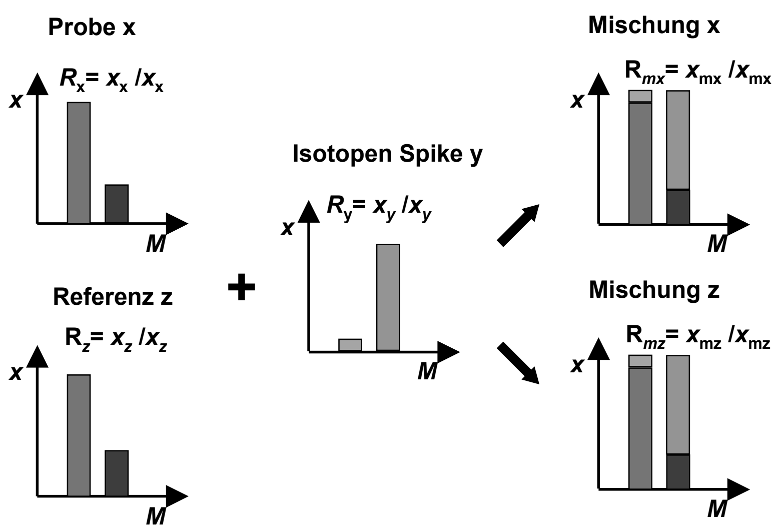


Reliability of element calibration solutions

$$Difference = \frac{(LNE\_value - Certified\ value)}{LNE\_value} \times 100$$



Principle of inverse IDMS (exact matching)



National Standards

Primary standard:	Standard having the highest metrological quality
Realized by:	National Metrology Institutes
By means of :	Primary materials and methods of measurement e.g. IDMS
Assured by:	International comparison measurements (CCQM)
Providing:	National and international comparability and acceptance (MRA)

**For you,
the authorities,
the field laboratories, etc.**

German metrological network

National reference level for chemical measurements

PTB
(coordinator)

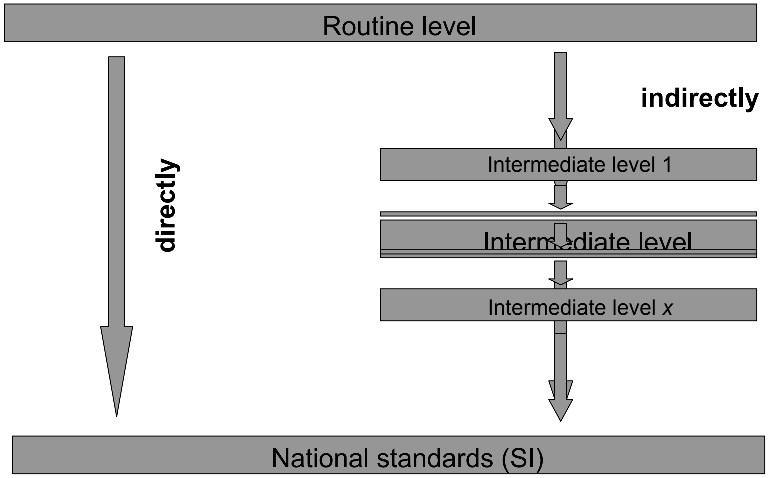
• organic analysis
• element analysis
• electrochemistry
• gas analysis
• clinical chemistry

BAM
• pure primary substances
• primary standards for gas analysis
• certified reference materials

DGKL
• clinical chemistry

UBA
• gas analysis for environmental protection

Possible traceability structures



Direct link up

Routine level

<u>Use of primary methods</u>

- Not recommendable for routine purposes

because:
- Expensive
- Time consuming
- Difficult
- No national and international recognition
- Task of the NMI's
  (better is to use the service of the NMI)

SI

Proficiency testing in the OSPAR framework

Patrick Roose
Management Unit of the North Sea Mathematical Models (MUMM)

Abstract

Scientific knowledge of the seas is the indispensable basis for all marine management. The 1992 OSPAR Convention for the Protection of the Marine Environment of the North-East Atlantic is the current instrument guiding international cooperation on the protection of the marine environment of the North-East Atlantic. To monitor environmental quality throughout the North-East Atlantic, a Joint Assessment and Monitoring Programme (JAMP) has been established. However well a monitoring and assessment programme is defined and executed, if there is no assurance that the information gathered is of good quality and if it is not assessed appropriately, the total exercise will have little value. For this reason, OSPAR has adopted a quality assurance (QA) policy, which acknowledges the importance of reliable information as the basis for effective and economic environmental policy and management regarding the Convention area. This policy requires that QA procedures should be applied to the whole chain of JAMP activities, from programme design, through execution, evaluation and reporting to assessment. In this overall QA framework, the QUASIME proficiency-testing scheme (PTS) plays an important role, particularly in the implementation of collective OSPAR monitoring.

Kurzfassung

Wissenschaftliche Kenntnisse der Meeresumwelt sind unentbehrliche Grundlage zur Überwachung und Erhaltung mariner Systeme. Die 1992 verabschiedete OSPAR- Convention bildet das gegenwärtig maßgebliche Instrumentarium bei der internationalen Kooperation zum Schutz der Meeresumwelt des Nordostatlantiks. Um qualitative Veränderungen im gesamten Überwachungsgebiet aufzuzeigen, wurde das Joint Assessment and Monitoring Programme (JAMP) etabliert. Auch wenn ein solches Programm definiert und ausgeführt wird, hat seine Anwendung wenig Wert, wenn nicht sicher gestellt ist, dass die erhobenen Informationen von guter Qualität sind. Aus diesem Grund hat OSPAR ein Konzept der Qualitätssicherung (QA) eingeführt, welches die Bedeutung verlässlicher Informationen als Grundlage eines wirkungsvollen und ökonomischen Umweltmanagements innerhalb des Einzugsgebiets der OSPAR hervorhebt. Dieses Konzept erfordert, die Anwendung qualitätssichernder Maßnahmen innerhalb der gesamten Kette der JAMP- Aktivitäten, d.h. angefangen bei der Aufstellung von Monitoring- Programmen, über die Durchführung, Auswertung und Darstellung, bis hin zur Bewertung der gewonnenen Daten. In diesem Gesamtrahmen der QA spielen die QUASIME Ringversuche (PTS), insbesondere bei der Durchführung des gemeinsamen OSPAR-Monitorings, eine wichtige Rolle.

MUMM

OSPAR

Proficiency testing in the OSPAR framework

Patrick Roose

RBINS

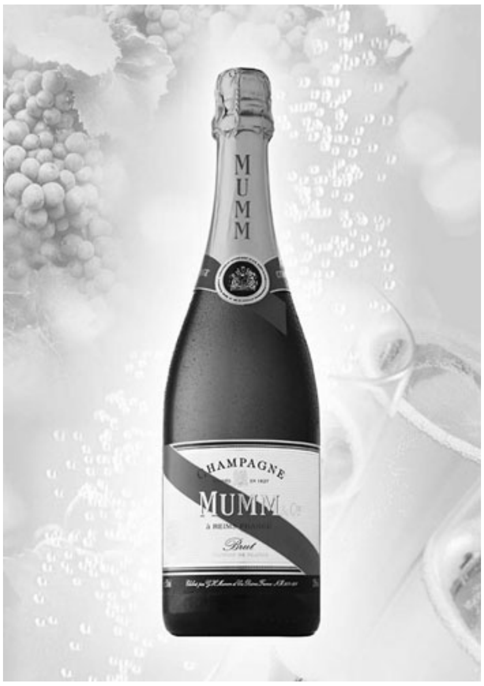
MUMM

OSPAR

What's MUMM?

RBINS

MUMM



OSPAR

RBINS

MUMM

Management Unit of the North Sea Mathematical Models (MUMM)

or

the Department for Management of the Marine ecosystem of the

Royal Belgian Institute of Natural Sciences (RBINS)

http://www.mumm.ac.be

OSPAR

RBINS

MUMM

OSPAR

RBINS

OSPAR?

MUMM

OSPAR

RBINS

OSPAR :

The convention for the protection of the marine environment of the north-east Atlantic

http://www.ospar.org

OSPAR area and contracting parties

The 1992 OSPAR Convention requires Contracting Parties to "*take all possible steps to prevent and eliminate pollution and take the necessary measures to protect the maritime area against adverse effects of human activities so as to safeguard health and conserve marine ecosystems and, where practicable, restore marine areas which have been adversely affected*".

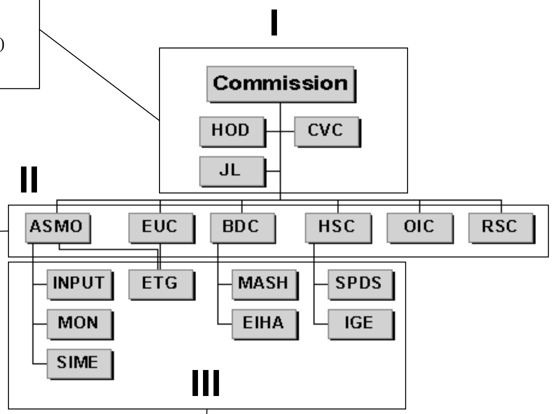
MUMM

Management and advice

- •Committee of Chairmen and Vice-Chairmen (CVC)
- •Group of Jurists/ Linguists (JL)
- •Heads of Delegation (HOD)

Second tier level

- •Environmental Assessment and Monitoring Committee (ASMO)
- •Eutrophication Committee (EUC)
- •Biodiversity Committee (BDC)
- •Hazardous Substances Committee (HSC)
- •Offshore Industry Committee (OIC)
- •Radioactive Substances Committee (RSC)



Third tier level

- •Working Group on Inputs to the Marine Environment (INPUT)
- •Working Group on Monitoring (MON)
- •Working Group on Concentrations, Trends and Effects of Substances in the Marine Environment (SIME)
- •Eutrophication Task Group (ETG)
- •Working Group on Marine Protected Areas, Species and Habitats (MASH)
- •Working Group on the Environmental Impact of Human Activities (EIHA)
- •Working Group on Substances and Point and Diffuse Sources (SPDS)
- •Informal Group of DYNAMEC experts (IGE)

OSPAR

RBINS

MUMM

OSPAR recognizes that scientific knowledge of the seas is the indispensable basis for all marine management.

OSPAR

RBINS

MUMM

OSPAR

RBINS

The OSPAR Convention requires the Contracting Parties, amongst other things, to "*cooperate in __carrying out monitoring programmes__*", *to __develop quality assurance methods__, and assessment tools*.

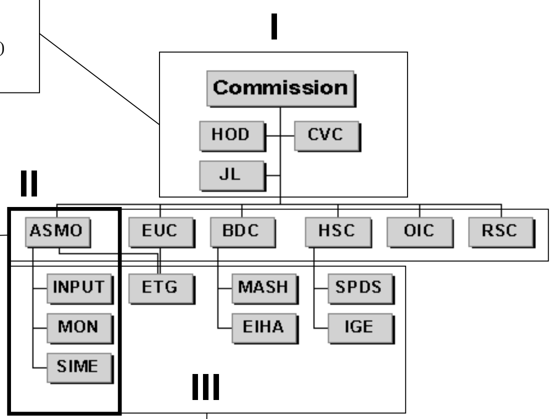
MUMM

OSPAR

RBINS

Management and advice
- Committee of Chairmen and Vice-Chairmen (CVC)
- Group of Jurists/ Linguists (JL)
- Heads of Delegation (HOD)

Second tier level
- <u>Environmental Assessment and Monitoring Committee</u> (ASMO)
- <u>Eutrophication Committee</u> (EUC)
- <u>Biodiversity Committee</u> (BDC)
- <u>Hazardous Substances Committee</u> (HSC)
- <u>Offshore Industry Committee</u> (OIC)
- <u>Radioactive Substances Committee</u> (RSC)



Third tier level
- Working Group on Inputs to the Marine Environment (INPUT)
- Working Group on Monitoring (MON)
- Working Group on Concentrations, Trends and Effects of Substances in the Marine Environment (SIME)
- Eutrophication Task Group (ETG)
- Working Group on Marine Protected Areas, Species and Habitats (MASH)
- Working Group on the Environmental Impact of Human Activities (EIHA)
- Working Group on Substances and Point and Diffuse Sources (SPDS)
- Informal Group of DYNAMEC experts (IGE)

MUMM

The OSPAR monitoring programmes:

JAMP (marine environment), RID (riverine input and discharges) and CAMP (the atmosphere)

OSPAR

RBINS

MUMM

The JAMP

OSPAR

RBINS

MUMM

OSPAR

RBINS

To monitor environmental quality throughout the North-East Atlantic, a Joint Assessment and Monitoring Programme (JAMP) has been established.

MUMM

OSPAR

RBINS

The main objectives of the JAMP:

a. **the preparation of environmental assessments of the status of the marine environment of the maritime area or its regions, including the exploration of new and emerging problems in the marine environment;**
b. **the preparation of contributions to overall assessments of the implementation of the OSPAR Strategies, including in particular the assessment of the effects of relevant measures on the improvement of the quality of the marine environment. Such assessments will help inform the debate on the development of further measures;**

supported by:

c. **the implementation of collective OSPAR monitoring, including the development of the necessary methodologies;**
d. **The preparation of environmental data and information products needed to implement the OSPAR Strategies.**

MUMM

OSPAR

RBINS

Why emphasize quality?

MUMM

OSPAR

RBINS

During the 1998 OSPAR assessment of trends in the concentrations of some metals, PAHs and other organic compounds in the tissues of various fish species and mussels (OSPAR, 1999), about 1280 Time Series (TS) of three or more years for metals and organic contaminants in blue mussel and fish were assessed by an OSPAR working group, for the period 1978-1996.

MUMM

OSPAR

RBINS

A lot of the data was rejected because of the lack of, or bad quality assurance (QA) information. From 67% up to 100% of the data from individual countries were rejected simply due to a lack of essential QA information. This led to a distinctive shortening or loss of time series.

MUMM

OSPAR

RBINS

Table 1. Trace metal observations rejected due to a lack of QA information reported to the ICES database

Country	Trace metals		
	Submitted [1]	Accepted [2]	% Loss
Belgium	7 083	6 935	2
Denmark	9 329	8 326	11
France	4 625	4 139	11
Germany	11 556	9 925	14
Iceland	2 075	1 592	23
Netherlands	4 005	2 925	27
Norway	16 084	13 288	17
Spain	[3]	1 093	[3]
Sweden	18 416	18 156	1
United Kingdom	5 547	5 148	7

1 Number of observations submitted per country per contaminant group.
2 Number of observations accepted after QA assessment.
3 Unavailable due to uncertainties noted by the Spanish delegation just prior to publication.

MUMM

Table 2. Organic observations rejected due to a lack of QA information reported to the ICES database

Country	Organic contaminants				
	Submitted [1]	Accepted [2]	% Loss	Group 2 [3]	Not QA assessed [4]
Belgium	6 090 (7 390)	1 625	73	3 365	1 300
Denmark	85 (176)	0	100	0	91
France	2 532 (2 738)	193	92	1 699	251
Germany	13 447 (18 423)	2 480	82	4 405	4 976
Iceland	2 362 (2 992)	596	75	1 097	630
Netherlands	8 488 (11 760)	767	91	2 245	3 272
Norway	36 805 (47 096)	10 078	73	18 419	10 291
Spain	3 924 (5 380)	1 286	67	346	1 456
Sweden	17 614 (20 562)	0	100	12 402	2 948
United Kingdom	4 672 (5 376)	0	100	4 672	704

[1] Number of observations submitted per country per contaminant group. The number in parenthesis gives the total number of observations (including those that were not subject to QA assessment-see note 4).

[2] Number of observations accepted after QA assessment.

[3] Number of organic 'group 2' observations, i.e. observations that were rejected by the QA assessment, but which were later on included in the assessment (α-HCH,γ-HCH, pp'-DDT, pp'-DDE, pp'-TDE, dieldrin, TNONC.

[4] Number of organic observations that were not subject to QA assessment (HCB, CB 28, CB 52, ΣPCB_7).

OSPAR

RBINS

MUMM

"The rejection of numerous data through Quality Assurance (QA) procedures motivated a need to find means to include "poorer" quality data without significantly affecting the power to detect trends. Further investigation of the means to address this problem is encouraged. "

"Improvements in any similar future assessment exercise are discussed and include the need to adhere to agreed guidelines and submission deadlines, and to have prior agreement of QA procedures and better automation of the generation of related tables and figures."

OSPAR

RBINS

MUMM

OSPAR

RBINS

However well a monitoring and assessment programme is defined and executed, if there is no assurance that the information gathered is of good quality and if it is not assessed appropriately, the total exercise will have little value.

MUMM

OSPAR

RBINS

OSPAR has adopted a quality assurance (QA) policy which acknowledges the importance of reliable information as the basis for effective and economic environmental policy and management regarding the Convention area.

MUMM

OSPAR

RBINS

This policy requires that:

QA procedures should be applied to the whole chain of JAMP activities, from programme design, through execution, evaluation and reporting to assessment.

QA should be fit for purpose (i.e. the assessment or monitoring activity to which it relates)

MUMM

OSPAR

RBINS

The CEMP

MUMM

OSPAR

RBINS

The CEMP can be described as that part of monitoring within the JAMP where the national contributions overlap and are coordinated .

MUMM

OSPAR

RBINS

Three elements are essential for the realisation of the CEMP:

a. **Guidelines**
b. **Quality assurance tools**
c. **Assessment tools**

MUMM

Contracting parties are obliged to participate in proficiency testing schemes such as QUASIMEME in order to assure the quality of their data.

OSPAR

RBINS

MUMM

CEMP

Key parameters:

Mercury, cadmium and lead in biota and sediments
PCBs in biota and sediments
PAHs in biota and sediment
Nutrients in seawater
Direct and indirect Eutrophication effects
PAH- and Metal specific Biological effects
Organotins in sediments
TBT specific effects

OSPAR

RBINS

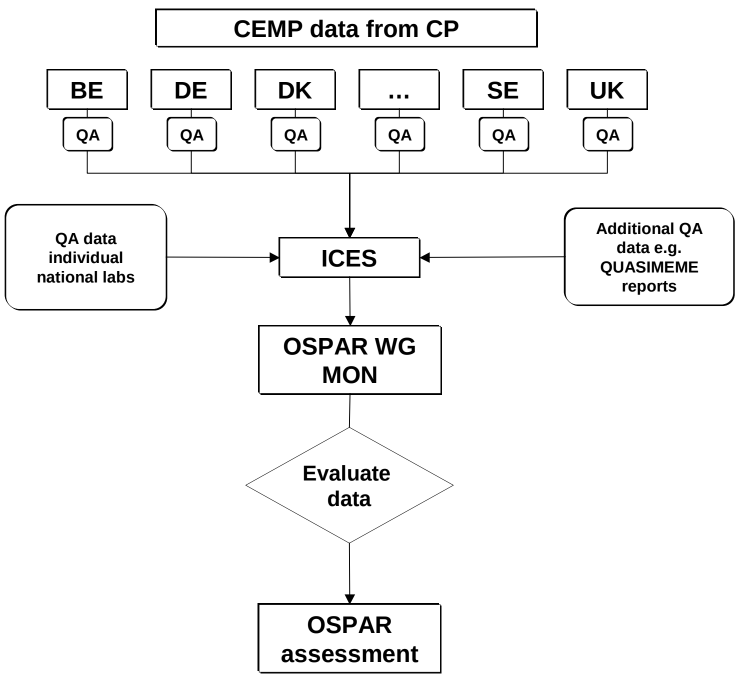
MUMM

Data reporting and evaluation

OSPAR

RBINS

MUMM



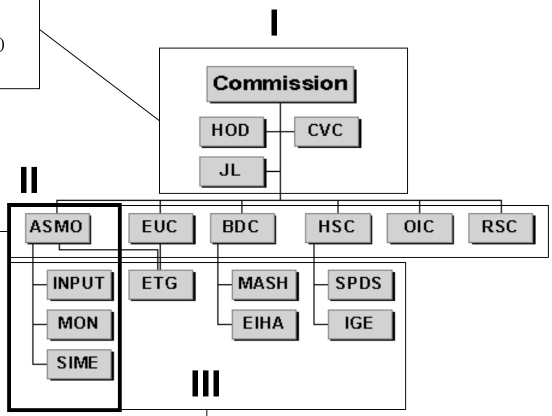
OSPAR

RBINS

MUMM

Management and advice
- Committee of Chairmen and Vice-Chairmen (CVC)
- Group of Jurists/ Linguists (JL)
- Heads of Delegation (HOD)

Second tier level
- Environmental Assessment and Monitoring Committee (ASMO)
- Eutrophication Committee (EUC)
- Biodiversity Committee (BDC)
- Hazardous Substances Committee (HSC)
- Offshore Industry Committee (OIC)
- Radioactive Substances Committee (RSC)



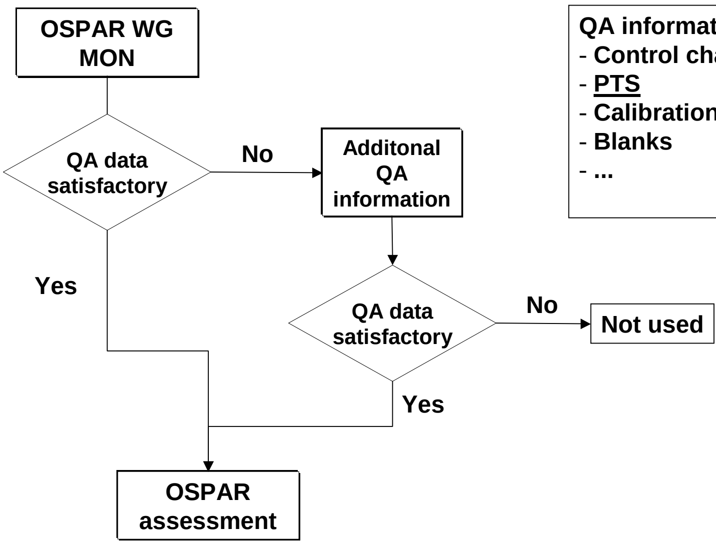
Third tier level
- Working Group on Inputs to the Marine Environment (INPUT)
- Working Group on Monitoring (MON)
- Working Group on Concentrations, Trends and Effects of Substances in the Marine Environment (SIME)
- Eutrophication Task Group (ETG)
- Working Group on Marine Protected Areas, Species and Habitats (MASH)
- Working Group on the Environmental Impact of Human Activities (EIHA)
- Working Group on Substances and Point and Diffuse Sources (SPDS)
- Informal Group of DYNAMEC experts (IGE)

OSPAR

RBINS

MUMM

QA data evaluation 1998 assessment

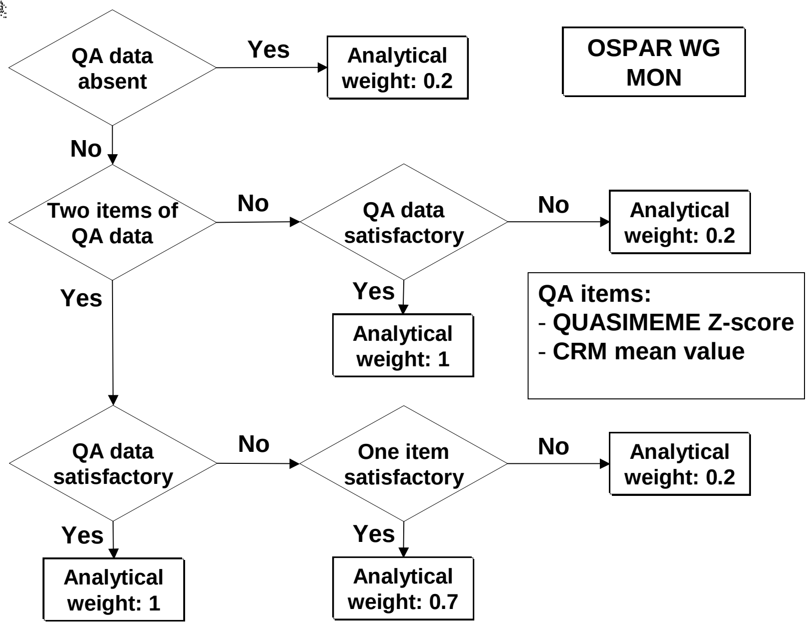


QA information:
- Control charts
- PTS
- Calibration
- Blanks
- ...

OSPAR

RBINS

QA data evaluation 2005 CEMP assessment



MUMM

OSPAR

RBINS

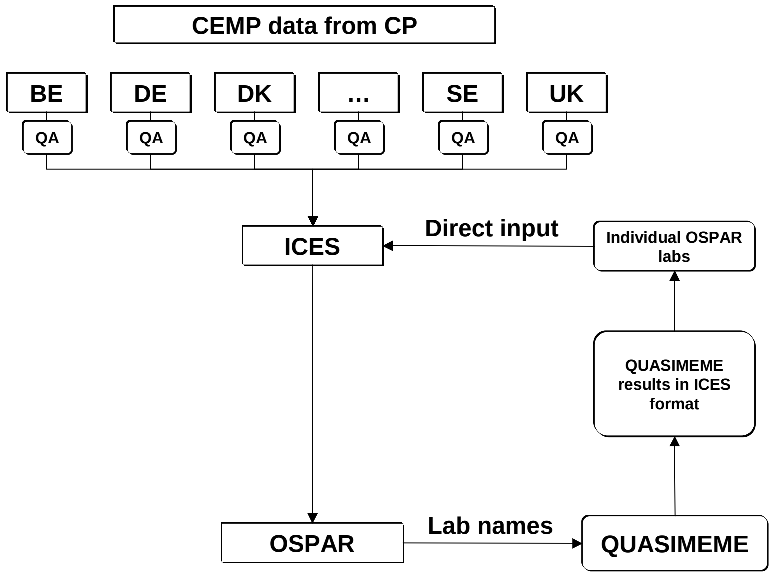
Future

MUMM

OSPAR

RBINS

Future data and QA data reporting



Future

- Facilitated transfer of QUASIMEME QA data to ICES
- Further development of data filters
- Continued attention to data quality
- Expanded lists of contaminants with their specific quality requirements
- Improved follow-up of data and QA reporting
- Biological effects and their specific QA (BEQUALM)

Proficiency testing of field laboratories - Experience from the soil sector

Roland Becker
Federal Institute for Materials Research and Testing (BAM)

Abstract

The proficiency testing scheme „contaminated sites“ is operated at BAM since 1996 and covers the determination of organic and inorganic priority pollutants in soil. Originally being one prerequisite for the accreditation of field laboratories for measurement campaigns on federal estates with a history of industrial or military use the scheme meanwhile attracts up to 170 participants per round mostly on a voluntary basis.
The available technological capabilities for the production of large sets of soil reference materials and their characterisation are the basis of any sound assessment of measurement results done on them and are therefore briefly outlined. The operation of the scheme and the determination of assigned values are discussed with regard to the requirements set by ISO Guide 43 and (inter)national standards DIN 38402 and ISO 13528. The use of asymmetric tolerance intervals for the assessment of participant performance and the robust evaluation of repeatability and comparability standard deviations according to DIN 38402 are contrasted to traditional procedures. The dependence of the comparability standard deviations on the nature and mass content of selected analytes is briefly discussed.
Attempts to support the “once measured-everywhere-accepted" idea by comparing different PT schemes and by international key comparisons of metrology institutes are reviewed.

Kurzfassung

Seit 1996 veranstaltet die BAM Altlasten-Ringversuche zur Bestimmung von organischen und anorganischen Schadstoffen in Böden. Das Ringversuchssystem wurde ursprünglich als eine Grundlage für die Akkreditierung von Prüflaboratorien für Messkampagnen auf vormals industriell oder militärisch genutzten Bundesliegenschaften eingerichtet und erreicht inzwischen bis zu 170 Laboratorien, zum großen Teil auf freiwilliger Basis.
Die vorhandenen technologischen Voraussetzungen für die Herstellung größerer Serien an Bodenreferenzmaterialien und deren Charakterisierung werden als unverzichtbare Grundlage für eine Bewertung von Analysenergebnissen kurz beschrieben. Die Durchführung der Ringversuche und die Ermittlung der Sollwerte werden hinsichtlich der Forderungen des ISO Guide 43 sowie der Ringversuchsnormen DIN 38402 and ISO 13528 diskutiert. Die Verwendung von asymmetrischen Toleranzintervallen zur Bewertung der Laboratorien und die robuste Ermittlung von Wiederhol- und Vergleichsstandardabweichung gemäß DIN 38402 werden traditionellen Verfahren gegenübergestellt. Der Zusammenhang zwischen Analytgehalt und Vergleichsstandardabweichung wird kurz diskutiert.
Aktivitäten zur Förderung der gegenseitigen Anerkennung von Analysenergebnissen auf internationaler Ebene durch Vergleich von Ringversuchsystemen und durch Ringversuche von metrologischen Staatsinstituten (key comparisons) werden beleuchtet.

Aspects of proficiency testing – experience from the soil sector

R. Becker, H.-G. Buge, W. Walther

Federal Institute for Materials Research and Testing (BAM), Berlin

Workshop on Laboratory Proficiency Testing – Experiences in Data Assessment

Umweltbundesamt, Berlin, 7 October 2004

Overview

- Background and motivation of the PT scheme “Contaminated Sites”
- Technical prerequisites for the development of the reference materials
- Operation of a PT round
- Evaluation and assessment of results
- Specific aspects of environmental matrix reference materials

Federal Institute for Materials Research and Testing (BAM)

History:

1871 Prussian Royal Mechanical Technical Testing Institute
1920 Chemical-Technical State Institute
1954 Federal institute under the authority of the Federal Ministry of Economies and Labour (BMWA)
1200 Employees

Results:

700 Publications
1080 Lectures, courses
6000 Test certificates, approvals, consultancy reports

Cooperation in:

1300 National and international committees
145 Rules/regulations
615 Standards, rules for quality assurance
540 Technical-scientific societies and associations

www.bam.de

2

BAM

The Departments of BAM

- I Analytical Chemistry; Reference Materials
- II Chemical Safety Engineering
- III Containment Systems for Dangerous Goods
- IV Environmental Compatibility of Materials
- V Materials Engineering
- VI Performance of Polymeric Materials
- VII Safety of Structures
- VIII Materials Protection; Non- Destructive Testing
- S Interdisciplinary Scientific and Technological Operations
- Z Administration and Internal Services

Safety and Reliability in Chemical and Materials Technology

BAM

Types of interlaboratory comparisons

1. Method validation

Experienced laboratories

ISO 5725 (Part 2)

2. Proficiency testing (PT)

Any laboratory interested in or obliged to participation

Validated method(s), in-house methods

ISO 13528
DIN 38402-45

3. Certification of reference materials

Proficient laboratories

Validated method(s)

ISO Guide 35
BCR Guides

4. ILC for „Special occasions"

e. g. CCQM key comparisons

PT Scheme „Contaminated Sites“ - Background

- Contaminated sites on Federal estates formerly under industrial and military use
- Formation of many new field laboratories
- No nation-wide PT scheme for soil analysis

Motivation for the involvement of BAM

- To specify requirements for sampling in the field
- To comprise determination methods of organic and inorganic priority pollutants
- To provide independent assessment of the proficiency of field laboratories
  => Operation of a soil PT scheme
  => Development of certified reference materials (CRMs)

5

Objectives

The PT scheme “Contaminated Sites”

- To cover the priority pollutants in soil
- To reach as many field laboratories as possible
- To improve the degree of equivalence among participating laboratories
- To yield added values (method development, standardisation)

The single PT round

- Selection of matrix/analyte combination
- Preparation and characterisation of representative Reference Materials
- Development or distribution of analytical methods
- To provide “fair conditions” to all participants
- Processing and assessment of results

6

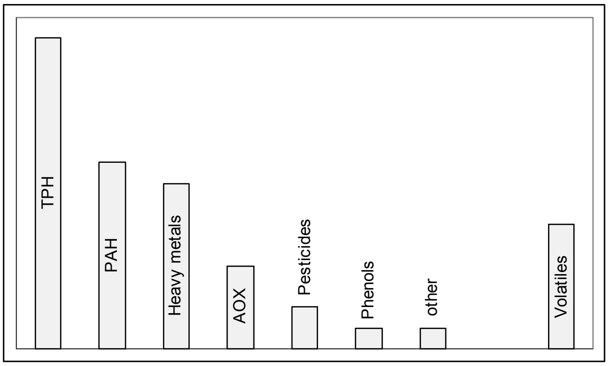
Priority pollutants on former industrial and military sites

Relative abundance on some Federal estates

- Hydrocarbons
- Organo chlorine compounds
- Heavy metals

Source:

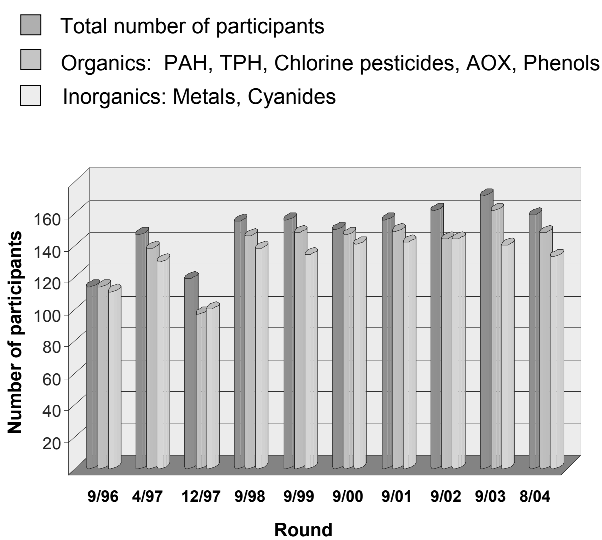
- Fuel, lubricants
- Gasworks
- Production and storage sites

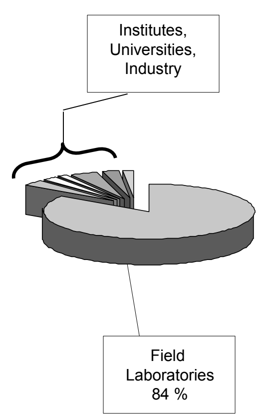


Selection of matrix/analyte combinations:

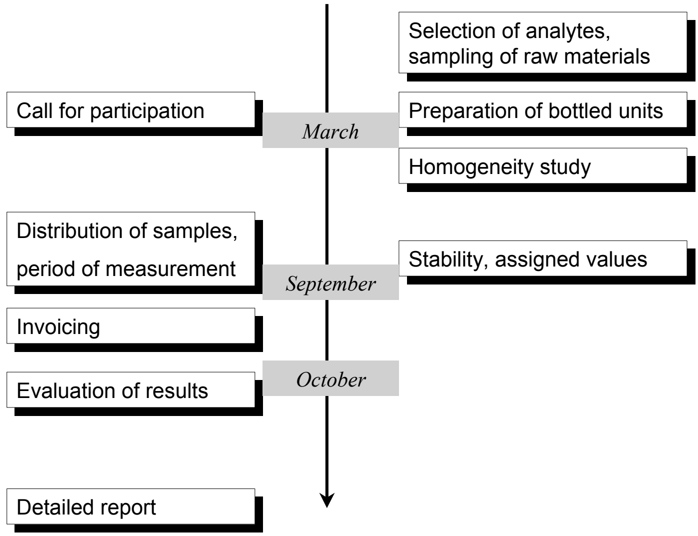
- Regulatory aspects
- Abundance
- Availability of materials
- Stability

PT Scheme „Contaminated Sites“





Organisation of a single round



Preparation of Soil Reference Materials

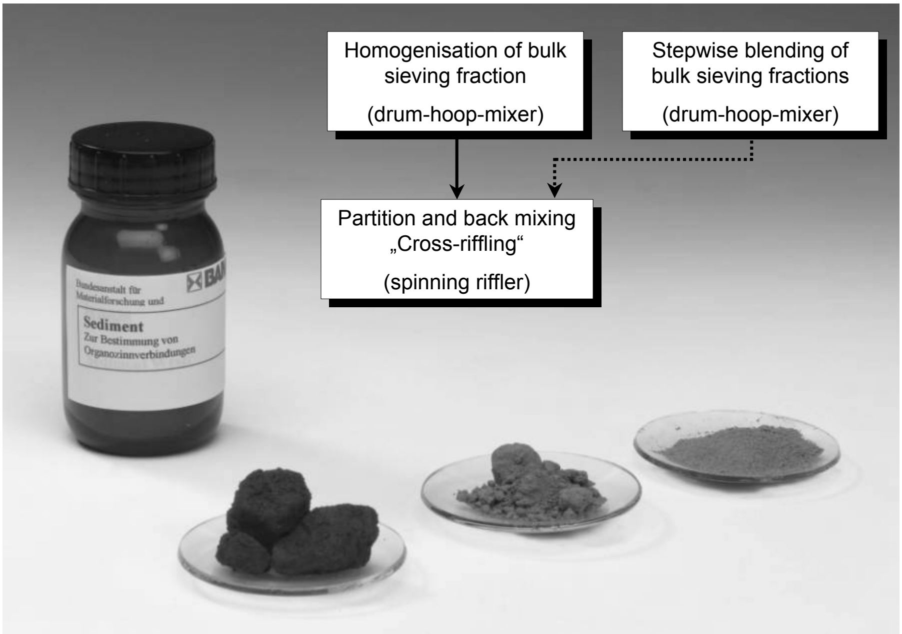
Selection of starting materials

- Specification of matrix/analyte combination
- Sampling of respective “real-world-materials”
- Level of analyte content typical for polluted soils
- No spiking of measurands

Processing into reference materials

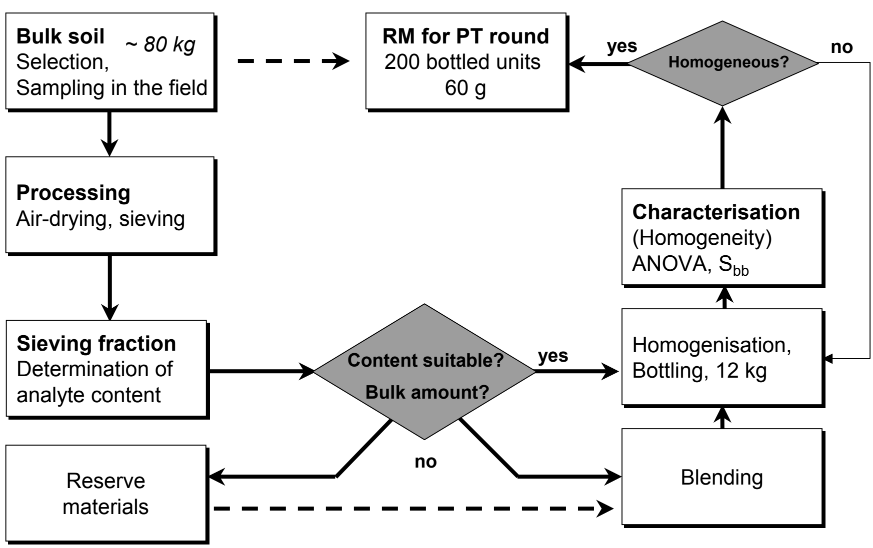
- Homogeneous batch of units
- Several batches with different levels of mass concentration
- Technical facilities for sieving, blending, bottling

Bottling via statistical homogenisation



11

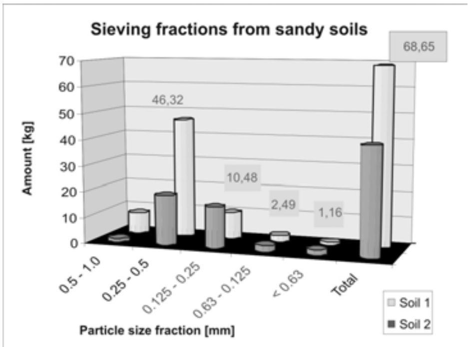
Preparation of Soil Reference Materials



12

Preparation of Reference Materials, Sieving

Fraktion	Particle size [mm]
I	1.00 – 0.500
II	0.500 – 0.250
III	0.250 – 0.125
IV	0.125 – 0.063
V	0.063 – 0.125
VI	< 0.063



Preparation of Reference Materials, Bottling

- 4-step procedure for sample partition and statistical homogenisation
- „Cross-riffling“-procedure using a spinning riffler

Example: 20.5 kg bulk soil

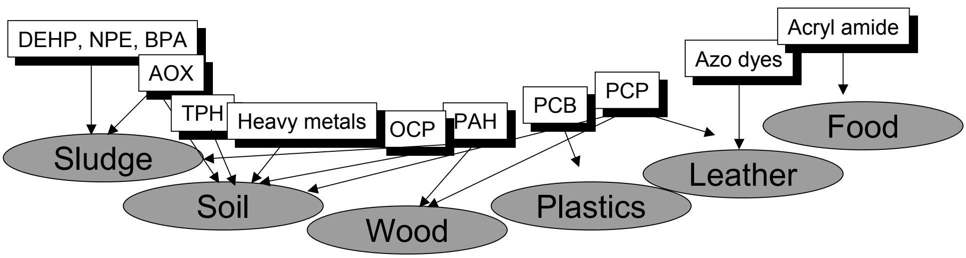
1. Partition	**8 x 2.56 kg**
2. Partition	**64 x 320 g**
Cross-Riffling	**8 x 2.56 kg**
3. Partition	**64 x 320 g**
4. Partition	**512 x 40 g**

BAM

Matrix reference materials in recent projects at BAM

- To be applied in PT rounds
- Method development
- Certified Reference Materials (CRM) (small fraction)
- Standardisation (ISO, CEN, DIN)

- Over 60 soil materials
- Many other matrix/analyte combinations
  (100 – 500 units per batch)



BAM

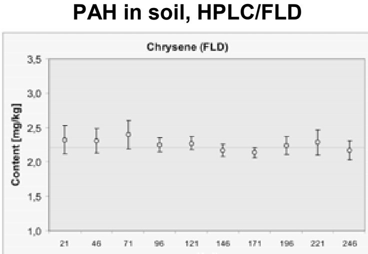
Homogeneity testing

Objective

- To estimate the variability of analyte content within and between the units
- To decide on the acceptance as RM

Adjustment of the analytical method → Determination of minimum sample intake → Selection of units, e.g. 10 out of 250 → 10 * 4 determinations, repeatability conditions → F-Test, ANOVA, Assessment

PAH in soil, HPLC/FLD



ANOVA

Source of variation	*Sum of Squares (SS)*	*dof*	*Mean Squares (MS)*	*Test value (F)*	*P-value*	*Critical F-Value*
Among groups	0,222175	9	0,024686	1,013470	0,451504	2,210697
Within groups	0,730742	30	0,024358			
Total	0,952917	39				

Measurement results

Grand mean: 2.26 mg/kg

Variability between bottles: 6.96 %

Variability within bottles: 6.92 %

Standard uncertainty of mean

$u_X \leq 0,3\hat{\sigma}$

acceptable

yes

Decision on acceptance as RM

Operation of the PT round

Fair competition

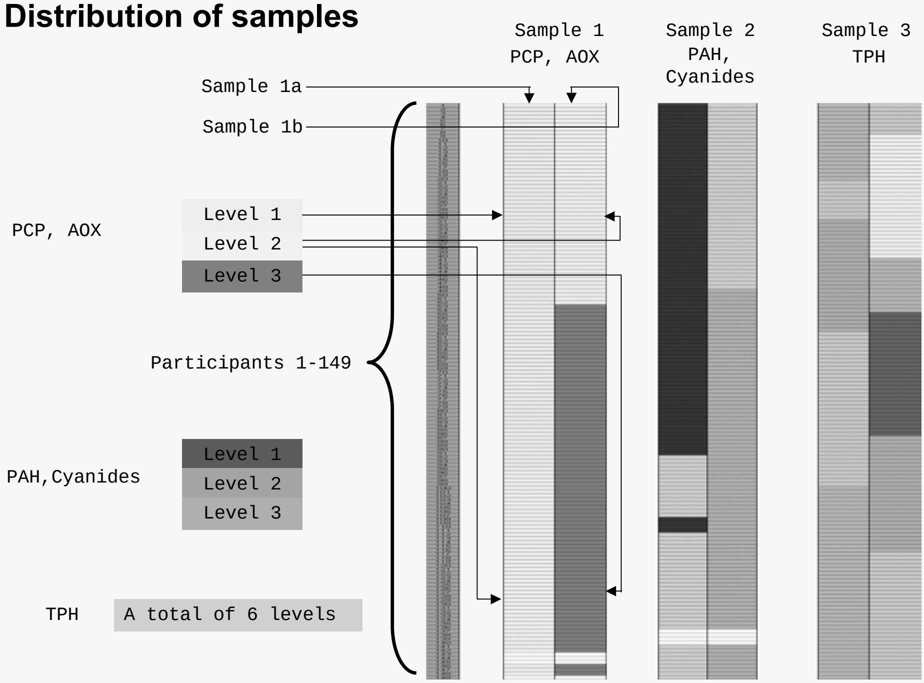
- Impede “telefoneability”
- Set tight period for reporting of results
- Distribute samples with different mass concentrations of measurands
- Documentation of procedures (e. g. chromatograms)

Basis for assessment

- Assigned values for accuracy
- Assigned values for reproducibility standard deviation
- ISO 13528 - Statistical methods for use in proficiency testing by interlaboratory comparisons
- DIN 38402-45 - Interlaboratory comparisons for proficiency testing of laboratories

17

Distribution of samples



Evaluation of results and performance assessment

ISO 13528

DIN 38402-45

Consensus values

- Consensus mean ***x**** *(robust average, Hampel* estimator)
- Consensus repeatability standard deviation (q-estimator)
- Consensus reproducibility standard deviation (q-estimator)

Performance assessment

- Assigned value ***X***
- Assigned reproducibility standard deviation $\hat{\sigma}$
- Modified z-score

19

Evaluation of results and performance assessment

Consensus values

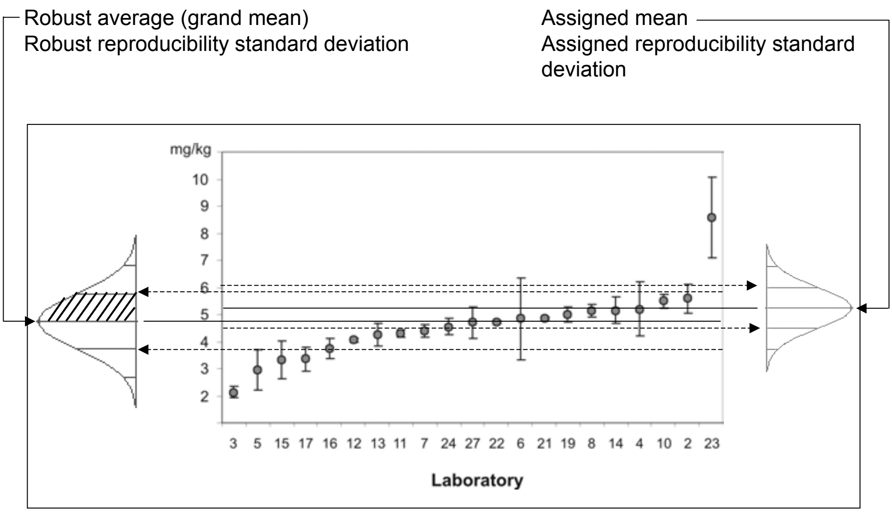
Robust average (grand mean)

Robust reproducibility standard deviation

Assigned values

Assigned mean

Assigned reproducibility standard deviation



20

Evaluation of results and performance assessment

$$Z - score = \frac{x_i - X}{\hat{\sigma}} \qquad x_i : \text{ Laboratory mean}$$

$|Z\text{-}score| \leq 2$: *satisfactory*

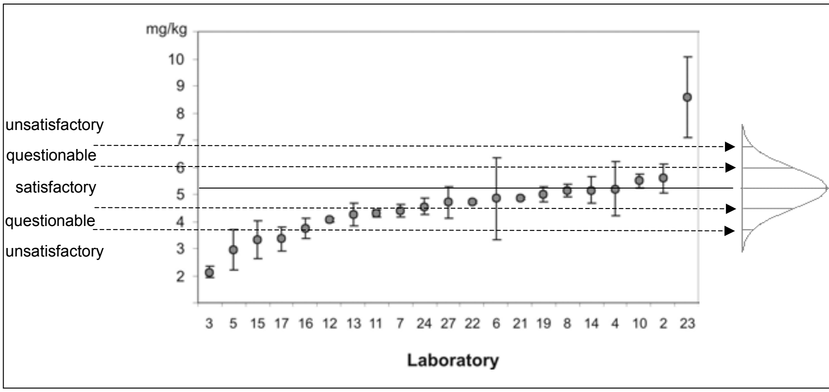
$2 \leq |Z\text{-}score| \leq 3$: *questionable*

$|Z\text{-}score| > 3$: *unsatisfactory*

Assigned values

Assigned mean ***X***

Assigned reproducibility standard deviation $\hat{\sigma}$



Assessment of proficiency using z_u-score

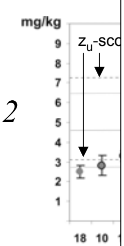
- z_u-score: Modified z-score (DIN 38402-45)

$$z_u = \frac{g}{k_1} \cdot z \quad (for\ z < 0)$$

$$z_u = \frac{g}{k_2} \cdot z \quad (for\ z \geq 0)$$

$$k_2 > k_1; \quad k_1, k_2 : f(\hat{\sigma}) \qquad g = 2$$

- Asymmetric tolerance interval
- Lower limit cannot drop below 0
- Avoids preference for lower recovery



Sample: PAH in soil, Level 1
Parameter: Phenanthrene
S_R: 1,189 mg/kg
Mean: 6,168 mg/kg
S_R rel.: 19,28%
Laboratories: 88
Assigned S_R for PT: 15,00%

Laboratory	Xi [mg/kg]	z-score	z_u- -score
23	2,09	-4,408	-4,711
12	4,04	-2,3	-2,458
2	4,15	-2,181	-2,331
9	4,305	-2,014	-2,152
56	4,35	-1,965	-2,1
31	4,43	-1,879	-2,008
17	4,705	-1,582	-1,69
67	4,86	-1,414	-1,511
43	5,08	-1,176	-1,257
..	..	..	..
..	..	..	..
6	7,5	1,439	1,324
41	7,59	1,536	1,414
26	7,94	1,915	1,762
21	8,17	2,163	1,99
78	9,1	3,168	2,915
11	9,48	3,579	3,293
39	9,935	4,071	3,745

Literature: S. Uhlig, P. Hentschel. Fres. J. Anal. Chem. (1997)

Determining assigned values

Consensus value from BAM laboratories

Example: Heavy metals

- 6 independent operator/equipment combinations
- ICP/MS, ICP/OES, AAS

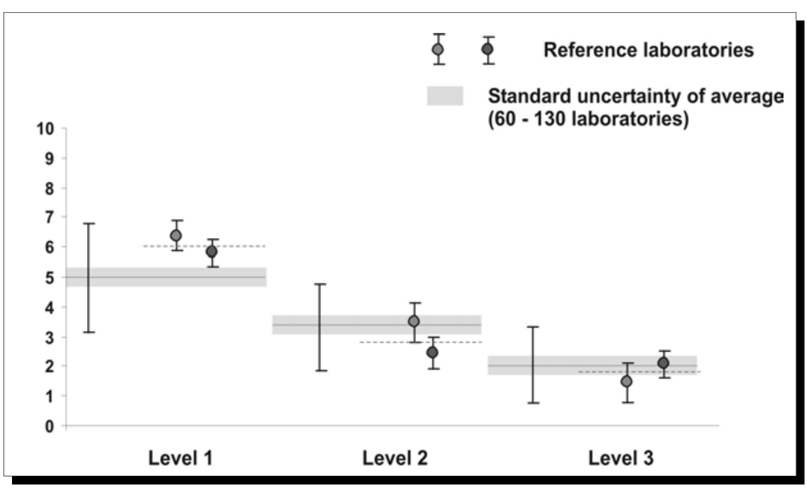
Consensus value from participants

Example: PAH

- 2 independent operator/equipment combinations (HPLC/FLD; GC/MS)
- Higher method variability

Determining assigned values

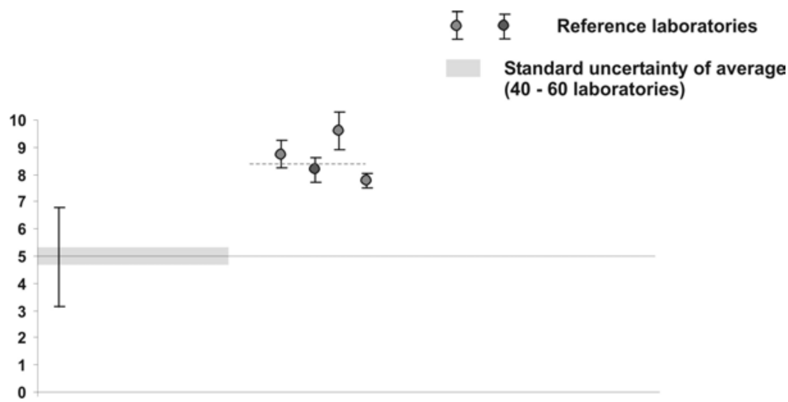
Example: PAH



Determining assigned values

Example: OCP, EOX

- Standard procedures tend to yield incomplete recovery



Determining the assigned standard deviation for proficiency assessment

“By perception“ (ISO 13528) => “fit-for-the-purpose“ statement

- The degree of equivalence that the coordinator wishes to be achieved
- Experience from various ILCs
- Experience from measurement campaigns (e. g. certification studies)
- Dependent on the level of content

$$z_u = \frac{g}{k_1} \cdot \frac{x_i - X}{\hat{\sigma}} \quad (for\ z < 0)$$

$$Z - score = \frac{x_i - X}{\hat{\sigma}}$$

$$z_u = \frac{g}{k_2} \cdot \frac{x_i - X}{\hat{\sigma}} \quad (for\ z \geq 0)$$

$$k_2 > k_1; \qquad k_1, k_2 : f(\hat{\sigma})$$

BAM

Typical assigned standard deviations

Analyte	Content [mg/kg]	Assigned RSD [%]
PAH	0,2 - 2	20 - 25
PAH	2 - 20	12 - 20
TPH	500 - 5000	8 - 10
AOX	10 - 100	10
OCP	1 - 50	18
Cd, Hg	0,1 - 1	10
Heavy metals	10 - 1000	5 - 7

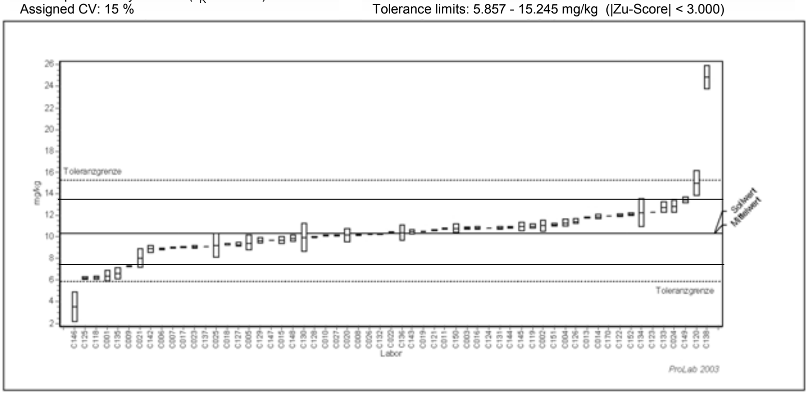
27

BAM

PT – Evaluation of results

Level: PAH 4
Mean: 10.308 mg/kg
Assigned mean: 10.308 mg/kg
Rel. Repeatability std.dev. (s_r rel.): 2.59 %
Rel. Reproducibility std.dev. (s_R rel. = CV): 15.88 %
Assigned CV: 15 %

Analyte: EPA-PAH (Sum)
Material: PAH in soil, level 4
Parameter: Sum of 16 PAH (EPA)
Method: DIN38402 A45
Number of Laboratories: 59
Tolerance limits: 5.857 - 15.245 mg/kg (|Zu-Score| < 3.000)



28

Assessment of proficiency

Single parameters (e. g. TPH)

=> Both samples to be passed

Parameter group (e. g. PAH, metals)

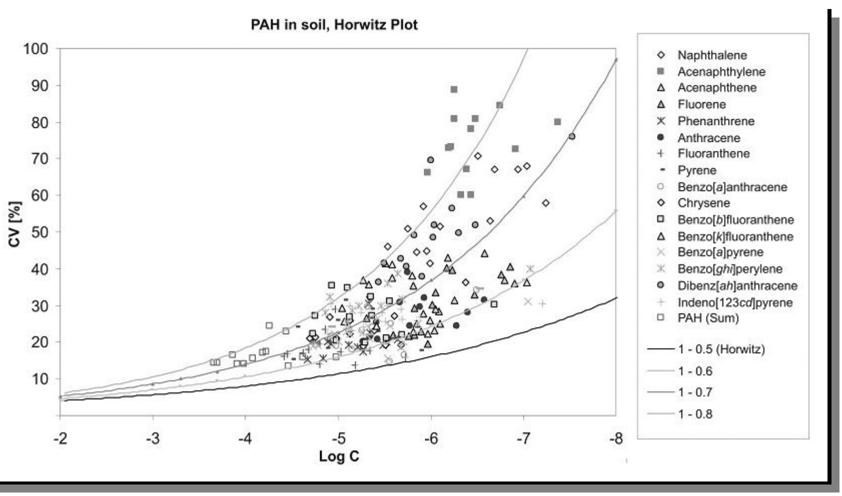
=> 80 % of all parameters in both samples to be passed

Reporting

- Anonymous presentation of results
- Certificate with individual results and performance
- Detailed report

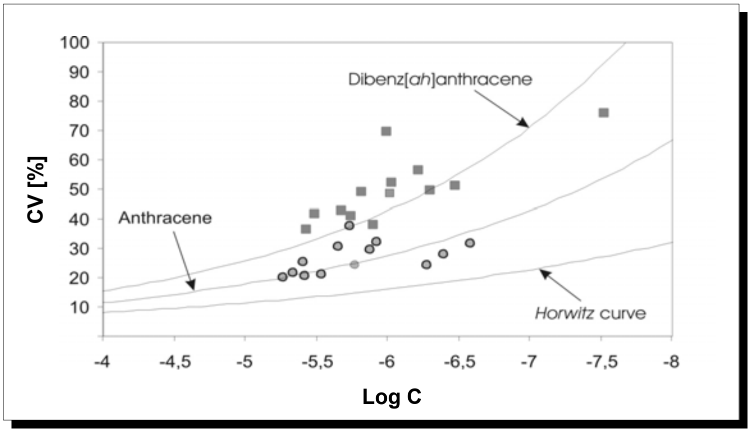
29

PAH in Soil - Interlaboratory variability



30

PAH in Soil - Interlaboratory variability of selected congeners



Comparison of PT schemes

- Information on available PT schemes
- Comparison, benchmarking of PT scheme operation
- Mutual recognition of performance in PT schemes

BAM

The PT scheme directory

eptis

European information system on proficiency testing schemes

www.eptis.bam.de

- Non-commercial directory service
- Joint publication of 16 European partners + USA
- Currently 790 PT schemes listed
- Hosted by BAM
- Open to further input
- Quality criteria according to ISO/IEC Guide 43

Supporting organisations

ilac eurolab Eurachem EA European Co-operation for Accreditation Institute for Reference Materials and Measurements IRMM

- Comparison of Intercomparisons (COEPT)

33

BAM

www.eptis.bam.de

EU CoEPT project

"Intercomparison of intercomparisons"

Objectives

To study the differences and similarities in the operation of PT schemes in selected analytical chemistry sectors:

- Soil (Polycyclic aromatic hydrocarbons – PAH)
- Water
- Milk powder
- Occupational hygiene

Two intercomparisons

- Comparison of statistical protocols (artificial data set)
- Comparison of whole schemes (real data obtained from a common RM)

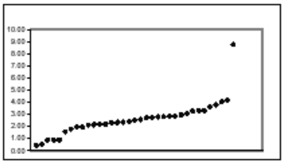
PT providers are invited to benchmark their evaluation protocols and results with CoEPT

34

COEPT

1st intercomparison

"PAH in soil"



Matrix = Soil
Parameter = Benzo[b]fluoranthene
Unit = mg/kg

Table Legend

x = Laboratory Result
X = *Assigned Value*
S = performance standard deviation
Z-score = Z = (x - X) / S

|Z| < 2, satisfactory
2 < |Z| < 3, questionable
|Z| > 3, unsatisfactory

	PT-provider	S01	S02	S04	S05	S06	S07	S09	S10	
		2.496	2.510	2.420	2.511	2.500	2.442	2.496	2.500	X
Lab	x	0.970	0.377	0.940	0.961	0.500	0.366	0.970	0.721	S
39221	0.46	-2.1	-5.4	-2.1	-2.1	-4.1	-5.4	-2.1	-2.8	
39231	0.56	-2.0	-5.2	-2.0	-2.0	-3.9	-5.1	-2.0	-2.7	
39228	0.87	-1.7	-4.4	-1.6	-1.7	-3.3	-4.3	-1.7	-2.3	
39232	0.88	-1.7	-4.3	-1.6	-1.7	-3.2	-4.3	-1.7	-2.2	
39218	0.90	-1.6	-4.3	-1.6	-1.7	-3.2	-4.2	-1.6	-2.2	
39226	1.57	-1.0	-2.5	-0.9	-1.0	-1.9	-2.4	-1.0	-1.3	
39207	1.78	-0.7	-1.9	-0.7	-0.8	-1.4	-1.8	-0.7	-1.0	
39224	1.93	-0.6	-1.5	-0.5	-0.6	-1.1	-1.4	-0.6	-0.8	
39209	1.95	-0.6	-1.5	-0.5	-0.6	-1.1	-1.3	-0.6	-0.8	
39223	2.11	-0.4	-1.1	-0.3	-0.4	-0.8	-0.9	-0.4	-0.5	
39227	2.13	-0.4	-1.0	-0.3	-0.4	-0.7	-0.9	-0.4	-0.5	
39204	2.18	-0.3	-0.9	-0.3	-0.3	-0.6	-0.7	-0.3	-0.4	
39210	2.19	-0.3	-0.8	-0.2	-0.3	-0.6	-0.7	-0.3	-0.4	
39214	2.32	-0.2	-0.5	-0.1	-0.2	-0.4	-0.3	-0.2	-0.2	
39211	2.34	-0.2	-0.5	-0.1	-0.2	-0.3	-0.3	-0.2	-0.2	
39222	2.35	-0.2	-0.4	-0.1	-0.2	-0.3	-0.3	-0.2	-0.2	
39213	2.40	-0.1	-0.3	0.0	-0.1	-0.2	-0.1	-0.1	-0.1	
39225	2.50	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	
39208	2.57	0.1	0.2	0.2	0.1	0.1	0.3	0.1	0.1	
39212	2.70	0.2	0.5	0.3	0.2	0.4	0.7	0.2	0.3	
39229	2.74	0.3	0.6	0.3	0.2	0.5	0.8	0.3	0.3	
39234	2.79	0.3	0.7	0.4	0.3	0.6	0.9	0.3	0.4	
39237	2.81	0.3	0.8	0.4	0.3	0.6	1.0	0.3	0.4	
39215	2.83	0.3	0.8	0.4	0.3	0.7	1.1	0.3	0.5	
39236	2.85	0.4	0.9	0.5	0.4	0.7	1.1	0.4	0.5	
39230	2.955	0.5	1.2	0.6	0.5	0.9	1.4	0.5	0.6	
39203	3.05	0.6	1.4	0.7	0.6	1.1	1.7	0.6	0.8	
39201	3.246	0.8	2.0	0.9	0.8	1.5	2.2	0.8	1.0	
39219	3.3	0.8	2.1	0.9	0.8	1.6	2.3	0.8	1.1	
39206	3.32	0.8	2.1	1.0	0.8	1.6	2.4	0.8	1.1	
39216	3.6	1.1	2.9	1.3	1.1	2.2	3.2	1.1	1.5	
39233	3.79	1.3	3.4	1.5	1.3	2.6	3.7	1.3	1.8	
39217	4.02	1.6	4.0	1.7	1.6	3.0	4.3	1.6	2.1	
39235	4.179	1.7	4.4	1.9	1.7	3.4	4.7	1.7	2.3	
39205	8.76	6.5	16.6	6.7	6.5	12.5	17.3	6.5	8.7	
39202										
39220										

- Comparison of data processing
- Artificial data set (34 laboratories)
- 8 PT providers

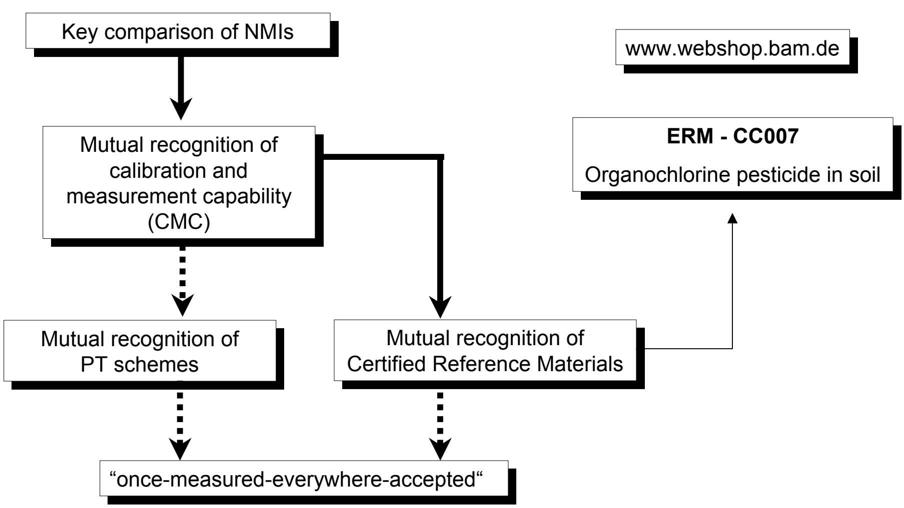
2<|Z|, satisfactory

2<|Z|<3, questionable

|Z|>3, unsatisfactory

35

CCQM – Key comparison



36

Thanks to:

Chr. Jung	**H. Scharf**
H. Weydig	**H.-G. Buge**
A. Hofmann	**W. Walther**

37

Proficiency Testing for Water Analyses in the Environment in Germany

Dr. Michael Koch

Abstract

The Institute of Sanitary Engineering, Universität Stuttgart, is the biggest proficiency test provider for the analysis of drinking water, wastewater and groundwater in Germany.

Wastewater and groundwater PTs are organized on behalf of the environmental authorities in cooperation with the Working Group of the Federal States on water issues (LAWA). Design, evaluation and assessment of these PTs is regulated a technical specification A-3 of LAWA. For every PT there is a list of possible analytical methods and a prescribed limit of determination. The statistical evaluation is done with robust methods (Q-method, Hampel estimator). The mean is used a conventional true value. The assessment of single values is done via Z_U-scores (limits: ± 2) with limitations for the standard deviation.

For a successful participation 80% of the values and 80% of the parameters have to be analysed successfully.

Since 2000 also PTs for rapid tests (cuvette tests) on wastewater in sewage treatment plants are organised with up to 311 participants. Design and evaluation in these PTs are according to DIN 38402-A45 using wherever possible also the variance function described in this standard. With data from both PT systems it is possible to compare the results of rapid test methods with conventional, standardized analyses.

Kurzfassung

Das Institut für Siedlungswasserbau der Universität Stuttgart ist der größte Ringversuchsveranstalter in Deutschland für die Analytik von Trink-, Grund- und Abwasser.

Die Grund- und Abwasserringversuche werden im Auftrag der Umweltbehörden in Kooperation mit der Länderarbeitsgemeinschaft Wasser (LAWA) organisiert. Das Design, die Auswertung und die Bewertung richten sich dabei nach dem Merkblatt A-3 der LAWA. Für jeden Ringversuch gibt es Liste zugelassener Analysenverfahren und Mindestbestimmungsgrenzen. Für die statistische Auswertung werden Verfahren der robusten Statistik angewandt (Q-Methode, Hampel-Schätzer). Der Mittelwert wird als konventionell richtiger Wert festgesetzt. Die Bewertung der Einzelwerte erfolgt über Z_U-Scores (Grenze: ±2) mit einer Begrenzung der Standardabweichung.

Für eine erfolgreiche Teilnahme müssen 80 % der Werte und 80% der Parameter erfolgreich analysiert worden sein.

Seit 2000 werden auch Ringversuche für Betriebsanalytik (Küvettentests) von Abwasser in Kläranlagen mit bis zu 311 Teilnehmern angeboten. Das Design und die Auswertung richten sich hier nach DIN 38402-A45, wobei, wann immer möglich, die in dieser Norm beschriebene Varianzfunktion angewandt wird. Mit den Daten aus beiden Ringversuchssystemen ist es möglich, die Ergebnisse der Betriebsmethoden mit denen der konventionellen, genormten Analytik zu vergleichen.

Universität Stuttgart
Germany

Proficiency Testing for Water Analyses in the Environment

Dr.-Ing. Michael Koch
Institute for Sanitary Engineering, Water Quality and Solid Waste Management, Universität Stuttgart
Dep. Hydrochemistry
Bandtäle 2, 70569 Stuttgart
Tel.: +49 711/685-5444 Fax: +49 711/685-7809
E-mail: Michael.Koch@iswa.uni-stuttgart.de
http://www.iswa.uni-stuttgart.de/ch

PT Provider for

- Drinking water
  - see U. Borchers
- Wastewater and groundwater
  - on behalf of the environmental authorities
- Rapid test methods for wastewater analysis
  - mainly for wastewater treatment plants

Wastewater and groundwater PT

- cooperation in the "Working Group of the Federal States on water issues" (LAWA) with the following other PT providers:
  - Institut für Hygiene und Umwelt, Hamburg
  - Bayrisches Landesamt für Wasserwirtschaft, München
  - Landesamt für Umweltschutz, Saarbrücken
  - Landesumweltamt Düsseldorf
  - Hessische Landesanstalt für Umwelt und Geologie, Wiesbaden
  - Niedersächsisches Landesamt für Ökologie, Hildesheim
  - Staatl. Umweltbetriebsgesellschaft, Radebeul
  - Landesanstalt für Umweltschutz, Halle
  - Landesamt für Natur und Umwelt, Kiel
  - Thüringer Landesanstalt für Umwelt, Jena

Regulations on laboratories for the proof of technical competence and for notification

- Requirements of ISO/IEC 17025
- "Module Water" with additional requirements
- Analysis of wastewater, groundwater and surface water is divided in 9 ranges (7 chemical, 2 biological)
  - 1 - Sampling and general parameters
  - 2 - Photometry, ion chromatography, titrimetry
  - 3 - Analysis of elements
  - 4 - Group and summary parameters (part 1)
  - 5 - Group and summary parameters (part 2)
  - 6 - Gas chromatographical methods
  - 7 - HPLC methods
  - 8 - Microbiological methods
  - 9 - Biological methods, biotests

PTs

Institute for Sanitary Engineering, Water Quality and Solid Waste Management, Universität Stuttgart

PTs

Year	Parameters	number of participants		range
		total	BW	
1998	Heavy metals in WW	495	130	3
1999	Group and summary parameters in WW	540	245	4,5
2000	Heavy metals in WW	539	206	3
2000	Major ions in GW	565	220	2
2001	Pesticides in GW	223	115	6,7
2001	Group and summary parameters in WW	506	133	4,5
2002	Elements in WW	478	168	3
2002	Volatile aromatic organics and volatile halocarbons in WW	326	222	6
2003	Major ions in WW	530	173	2
2003	PAH in GW	153	-	7
2003	Group and summary parameters in WW	448	124	4,5
2004	Elements in WW	394	149	3
2004	Chloropesticides in GW	ca. 160	ca. 90	6
2005	Major ions in WW			2
2005	Group and summary parameters in WW			4,5

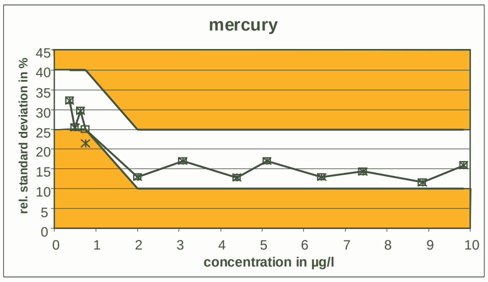
Design and Evaluation

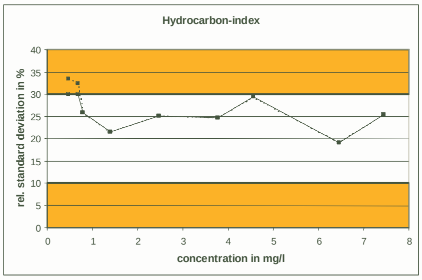
- According to a technical specification of the "Working Group of the Federal States on water issues" (LAWA): A-3
- real matrix, spiked with different amounts of analyte
- distribution with parcel service or collection by participants at decentralized delivery points
- list of allowed analytical methods
- prescribed limit of determination
- statistical evaluation with robust methods (Q-method, Hampel estimator)

Institute for Sanitary Engineering, Water Quality and Solid Waste Management, Universität Stuttgart

Assessment

- Conventional true value from robust mean
- Assessment of single values via Z_U-scores (limits: ± 2) with limitations for the standard deviation





Laboratory assessment

- The participation in the PT is successful if:
  - 80 % of all values are successful. Not successful are values
    - outside the tolerance limits
    - no quantification (no value, or "<…")
    - determined in another laboratory
    - analysed with wrong method
    - delivered after the deadline
  - 80 % of the parameters are successfully analysed
    - at least 50 % of the values of this parameter must be successful

PT for rapid test methods

- For self-checking analyses in waste water treatment plants
- no staff with special analytical skills
- normally cuvette tests from different manufacturers
- parameter: COD, total N, total P, ammonium-N, nitrate-N

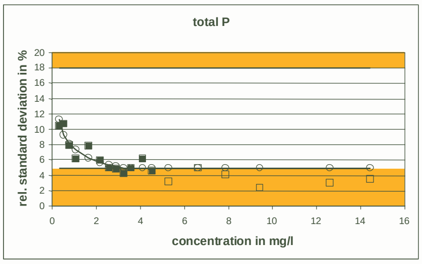
Design and Evaluation

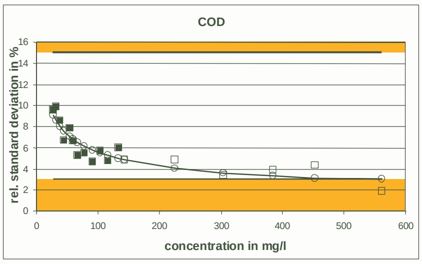
- According to DIN 38402 - A45
- real matrix, spiked with different amounts of analyte
- distribution with parcel service
- statistical evaluation with robust methods (Q-method, Hampel estimator)

Institute for Sanitary Engineering, Water Quality and Solid Waste Management, Universität Stuttgart

Assessment

- Conventional true value from robust mean
- Assessment of single values via Z_U-scores (limits: ± 2) using - wherever possible - the variance function from DIN 38402 - A45 and with limitations for the standard deviation





Laboratory assessment

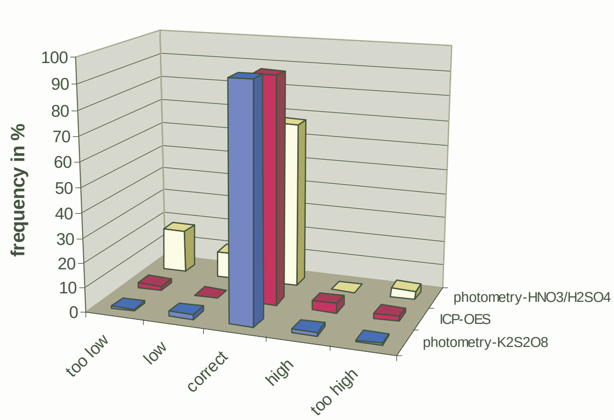
- The participation in the PT is successful if:
  - 80 % of the reported values are successful. Not successful are values
    - outside the tolerance limits
    - no quantification (no value, or "<...")
    - determined in another laboratory
  - Values for at least 60 % of the parameters have been reported (3 out of 5)
  - 80 % of the parameters are successfully analysed
    - at least 50 % of the values of this parameter must be successful

Method specific evaluation

- With all reported values the used analytical method is recorded
- From a histogram of Z_U-scores conclusions about the methods can be drawn
- 5 classes for Z_U:
  - $Z_U<-2$ too low
  - $-2\leq Z_U<-1$ low
  - $-1\leq Z_U\leq+1$ correct
  - $+1<Z_U\leq+2$ high
  - $Z_U>+2$ too high

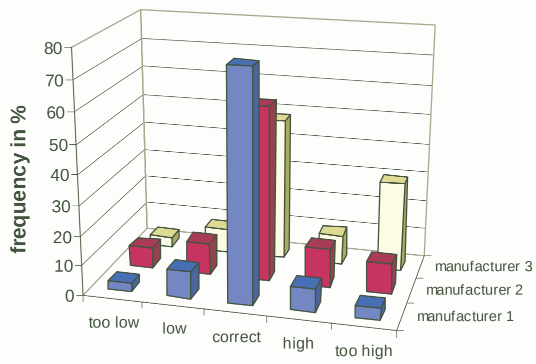
Institute for Sanitary Engineering, Water Quality and Solid Waste Management, Universität Stuttgart

Method specific evaluation – example 1



Institute for Sanitary Engineering, Water Quality and Solid Waste Management, Universität Stuttgart

Method specific evaluation – example 2

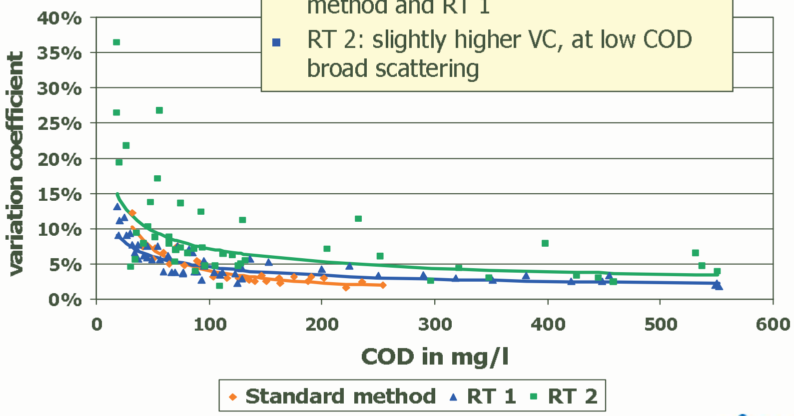


Comparison between conventional analysis and rapid test

- Method for comparison of precision
  - only methods are considered with more than 8 values per level
  - reproducibility standard deviation (robust; for each method separately) of all PTs is depicted vs. concentration
  - variance function is calculated for all values of each method and shown as a "centre line"

Comparison of precision – example

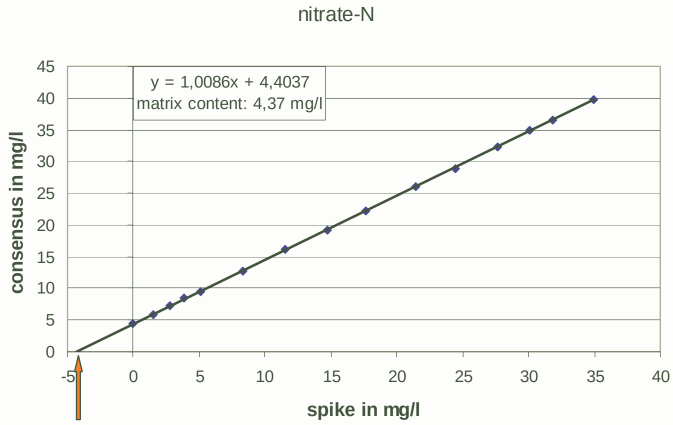
- good comparability between standard method and RT 1
- RT 2: slightly higher VC, at low COD broad scattering



Comparison of trueness

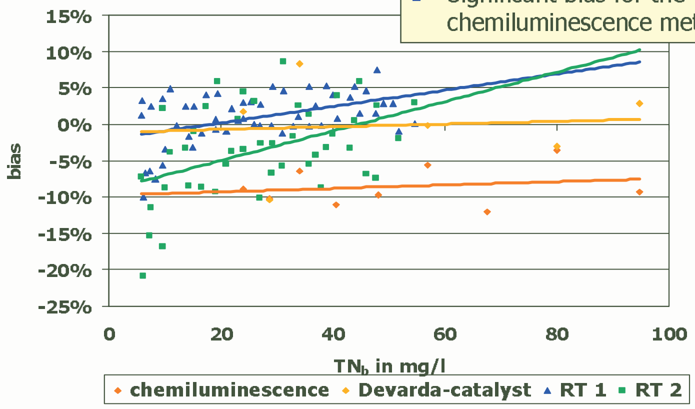
- Method for comparison of trueness
  - robust mean was calculated for each method separately and for all PTs and depicted vs. a conventional true value
  - conventional true value from matrix content plus spiked amount
  - calculation of a linear regression line for better visual comparison

Matrix content - example



Trueness - example

- Acceptable trueness for rapid tests
- Significant bias for the chemiluminescence method



Summary

- PTs for Water Analyses in the environment are well established in Germany
- The statistical methods described in DIN 38402-A45 - Q method, Hampel estimator and the variance function - proved to be very useful
- The comparison of different methods delivers valuable information