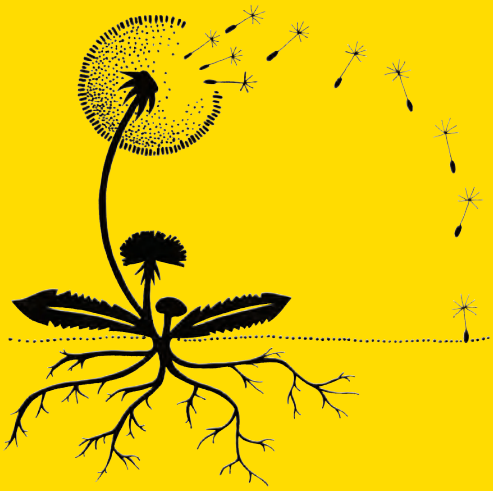
European Commission



Analytical measurement: measurement uncertainty and statistics

Ricardo Bettencourt da Silva,
Ewa Bulska, Beata Godlewska-Żyłkiewicz,
Martina Hedrich, Nineta Majcen,
Bertil Magnusson, Snježana Marinčić,
Ioannis Papadakis, Marina Patriarca,
Emilia Vassileva, Philip Taylor

Editors:
Nineta Majcen, Vaidotas Gegevičius

Joint Research Centre

Analytical measurement: measurement uncertainty and statistics

Editors:
Nineta Majcen
Vaidotas Gegevičius

Authors:
Ricardo Bettencourt da Silva
Ewa Bulska
Beata Godlewska-Żyłkiewicz
Martina Hedrich
Nineta Majcen
Bertil Magnusson
Snježana Marinčić
Ioannis Papadakis
Marina Patriarca
Emilia Vassileva
Philip Taylor

European Commission

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policymaking process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

European Commission
Joint Research Centre
Institute for Reference Materials and Measurements

Contact information
Institute for Reference Materials and Measurements
European Commission
Joint Research Centre
Retieseweg 111
B-2440 Geel
BELGIUM

E-mail: jrc-irmm-trainmic@ec.europa.eu
Tel.: +32 14 571 608
Fax: +32 14 571 863

http://irmm.jrc.ec.europa.eu
http://www.jrc.ec.europa.eu

Luxembourg: Publications Office of the European Union, 2013

JRC 68476

EUR 25207 EN
ISBN 978-92-79-23071-4
ISSN 1831-9424
doi:10.2787/58527

Printed in France

CONTENTS

INTRODUCTION 5

FOREWORD 6

ACKNOWLEDGEMENT 7

ABBREVIATIONS AND ACRONYMS 8

SYMBOLS 9

ABOUT THE AUTHORS 13

CHAPTER 1 19
Measurement uncertainty — Part I Principles

CHAPTER 2 85
Measurement uncertainty — Part II Approaches to evaluation

CHAPTER 3 129
Statistics for analytical chemistry — Part I

CHAPTER 4 179
Statistics for analytical chemistry — Part II

REFERENCES 231

FURTHER READING 233

INDEX 235

Introduction

Dear reader,

Having been involved in the TrainMiC® programme since the very beginning, when it was not even really in place and not known by its current name (i.e. as of 2000/01 onwards), it seems to me as if everything has been explained and said already several times and that there is not a lot more to add. For that reason, I will just briefly mention some thoughts from my personal point of view.

It happened that I was invited to take the position of a visiting scientist at the Institute for Reference Materials and Measurements of the Joint Research Centre of the European Commission during the time when the institute was also executing various activities related to metrology in chemistry in countries candidate to EU accession. Philip Taylor, who was in charge of this task at the institute, was convinced that a harmonised training material on various topics related to EN ISO/IEC 17025, Chapter 5, was very much needed. For that reason, he put Ewa Bulska, Emilia Vassileva, Steluta Duta and myself to work to start developing, under his guidance, training material on traceability, validation, interlaboratory comparison and other related topics. Some time later, Miloslav Suchanek, Ivo Leito, Piotr Robouch and Bertil Magnusson joined and discussions were becoming more and more vivid. Not only about the content, but also about the way the programme should be run. One thing was clear — Philip was right, there was great interest in the topic and it was not easy to handle it in several European countries, being so different, as well as following the requirements of European administration at the same time. More and more countries were joining the programme, more and more colleagues were contributing to the harmonised training material, and some were leaving. I was asked to chair the TrainMiC® Editorial Board in 2007. Being aware of the complexity of this task, I accepted with quite some fear. However, since then, Beata Godlewska-Żyłkiewicz, Bertil Magnusson, Emilia Vassileva, Ewa Bulska, Ioannis Papadakis, Marina Patriarca, Martina Hedrich, Mitja Kolar, Ricardo Bettencourt da Silva and Elizabeth Prichard have done a remarkable job, which has resulted in harmonised training material on various topics. We often had different views on a matter; however, we always succeeded in reaching an agreement. I am, therefore, honoured to have the opportunity to chair a group of colleagues from all around Europe who value expertise as well as dialogue amongst colleagues having different opinions. It is mainly for these reasons that we are now in the position to publish — for the first time — harmonised training material, prepared through the joint effort of the members of this and previous editorial boards. To complete the work, the contributions of Inge Verbist, Lutgart Van Nevel, Tomas Martišius and, especially, Philip were, of course, also essential.

Issuing this book on the occassion of 10 years of the TrainMiC® programme is somehow symbolic — may the next 10 years be at least as productive, useful and kind as the first decade.

Nineta Majcen
TrainMiC® Editorial Board, Chair

Laško, 9 May 2011

Foreword

It gives me great pleasure to write these words on the occasion of the publication of this book. The material presented is the hard work of all the people in the TrainMiC® Editorial Board over many years: people whose roots lie in many countries across Europe and by some strange fortune of faith came together and combined their efforts. I would like to thank them all for their commitment and endurance. Special thanks go to Nineta Majcen, who has been the chair of the Editorial Board since 2007. A daunting task indeed, which she accepted (luckily) without the full knowledge of its complexity. She funnelled the knowledge and the efforts of all the board members into many finished products: this book illustrates her patience, resilience and determination.

This book is about uncertainty and statistics. Oddly enough, it is published at a moment in the history of Europe which is truly loaded with uncertainty. Will the single currency survive? Will the European project come to a grinding halt? Will the demons of the past be released once again from their bottle in which they have been locked for so many decades? Nothing is certain or should be taken for granted.

It seems that humans need a vision to live and thrive. The TrainMiC® vision is to share the common effort of many with people across Europe and to do this in a networked and a non-colonial way, involving the knowledge of many, irrespective of their origin. An endeavour which many question in a day and age where market forces are the standard.

TrainMiC® grew from the deep belief that it is possible for people to work together despite their different insights, history, culture, values. A challenging task indeed, as we see in Europe today. And yes, during our journey we have certainly been able to experience all facets of *la condition humaine*.

I guess that is what makes this book special. In fact, at least about this, I am certain.

Philip Taylor
TrainMiC® Programme Leader

16 July 2011

Acknowledgement

The authors thank all who have in one way or another contributed in the past decade to the content of this book or to the development and deployment of the TrainMiC® programme in general. TrainMiC® would like to especially recognise the contributions of Professor Ivo Leito and Dr. Elizabeth Prichard.

Abbreviations and acronyms

ANOVA	Analysis of variance
BCR	Bureau Communautaire de Référence (Community Bureau of Reference)
BCR-479	Fresh water (low nitrate) certified reference material
CITAC	Cooperation on International Traceability in Analytical Chemistry
CRM	Certified Reference Material
DIN	Deutsche Industrie Norm (German Institute for Standardisation)
EC	European Commission
EN	European standard
EU	European Union
GUM	ISO Guide to the expression of uncertainty in measurement
IEC	International Electrotechnical Commission
ILC	Interlaboratory comparison
JRC-IRMM	Institute for Reference Materials and Measurements of the Joint Research Centre
ISO	International Organisation for Standardisation
JCGM	Joint Committee for Guides in Metrology
LOD	Limit of Detection
LOQ	Limit of Quantification
MU	Measurement uncertainty
PT	Proficiency testing
IQC	Internal Quality Control
VIM3	Third edition of the International vocabulary of metrology

Symbols

a	Intercept
A, A_{sample}	Absorbance
a, b, c, d, a_1, a_2, b_1, d_1, d_2	Input variables of the measurement function
AU	Absorbance Units
b_0	Intercept
b, b_1	Slope
C	Concentration (of the analyte)
C_{LOD}	Concentration of the analyte at the limit of detection
CI	Confidence interval
C_{ref}	Reference value
C_{std}	Concentration of the calibration standard
CV	Coefficient of variation
CV_{PT}	Coefficient of variation in PT studies
CV_R	Reproducibility coefficient of variation
d	Difference between two measurement results
$d_{difference}$	Mean of differences between paired values
df	Degrees of freedom
D	Residual
f	Factor
f_{std}	Unitary factor accounting for the calibration standards uncertainty
F	Value of F-test
F_1, F_2	Influence variables, not included in the original measurement function
G	Value of Grubbs' test
H_0	Null hypothesis
H_1	Alternative hypothesis
I_1, I_2	Influence variables of the measurement function
k	Coverage factor
k, k_a, k_b, k_c, k_d	Constant values of the measurement function
L_c	Decision level
m	Number of replicated analysis used to estimate $\overline{C}_{obs}$
MS	Squared deviation of the mean

$M_{0,1,2...n}$	Mean square values
n	Number of observations at each level
n	Sample size (the number of observations to include in a statistical sample)
N	Total number of observations (results)
P	Probability
p	Number of groups of data (levels)
p_i	Percentage contribution of the uncertainty component *i*
r	Correlation coefficient (linear regression)
R	Correlation coefficient (non-linear regression)
$\bar{R}$	Mean analyte recovery
R-chart	Range chart
RMS_{bias}	Root mean square of different bias values
RSD	Relative standard deviation
s, $s(x_i)$	Standard deviation
S_1	Sum of squares between groups
s^2	Variance
s_a	Standard deviation of the intercept
s_b	Standard deviation of the slope
s_{bl}	Standard deviation of the blank
$s_{c_{\text{obs}}}$	Standard deviation of results of '*m*' replicated analysis used to estimate $\bar{C}_{obs}$
S_0	Sum of squares within groups
s_{PT}	Standard variation of participating laboratories (in PT studies)
s_p	Pooled standard deviation
s_r	Repeatability standard deviation
s_R	Reproducibility standard deviation
SS	Sum of squared deviation about the mean
$s_{y/x}$	Residual standard deviation (standard deviation of the regression line
s_{RW}	Within-laboratory reproducibility
t	Value of t-test
$t_{(0.05,\ n-1)}$	Factor of Student's distribution
$u(x_i)$	Standard uncertainty of the input quantity x_i

u_i	Standard uncertainty associated with variable *i*
$u(C_{ref})$	Standard uncertainty of a reference value
u_d	Standard uncertainty of a difference between two results
u_{inter}	Standard uncertainty associated with the interpolation of a signal in a calibration curve
u(bias)	Uncertainty component due to possible bias
$u(R_w)$	Within-laboratory reproducibility uncertainty component (intermediate precision as defined in VIM3)
u_{Lab1}, u_{Lab2}	Standard uncertainties of measurements reported by Lab 1 or Lab 2
u_c	Combined uncertainty
$u_c(Y)$ or $u(Y)$	Combined standard uncertainty of the output variable
U	Expanded uncertainty
U_i	Expanded uncertainty associated with variable *i*
$U(Y)$ or U_Y	Expanded uncertainty associated with the output variable Y
U_d	Expanded uncertainty of the difference *d*
$V(x_i)$	Variance
w	Mass fraction
w_{Init}	Initially estimated mass fraction
w_{CRM}	Certified mass fraction
$\overline{w}_{obs}$	Mean estimated mass fraction
x	Independent Variable
$\overline{x}$	Mean of sample (the mean of the values for given number of observations, included in a statistical sample)
x_0	Value of input quantity
x_1, x_2	Input quantities
x_a, x_b, x_c	Input quantities of the measurement function associated with input variables *a*, *b* and *c*
x_{crit}	Critical value
x_{bl}	Mean of the blank measures
x_I	Influence quantity of the influence variable *I*
X_{L}	The signal at the limit of detection
X-chart	Mean chart (Shewhart)

x_Y	Output quantity of the measurement function associated with output variable Y
Y	Output variable of the measurement function (dependent variable)
Y_0	Response variable corresponding to blank (signal equal to blank signal)
Y_{LOD}	Signal at the limit of detection
Y_c	Critical value of the response variable
y	Final result
α	Type I error, level of significance
β	Type II error
ΔC_{cont}	Contribution due to possible contamination
	Mean of a population
ν	Degrees of freedom
σ	Standard deviation of a population

About the authors

Ricardo Bettencourt da Silva

Ricardo Bettencourt da Silva completed his BSc in chemistry at the Faculty of Sciences of the University of Lisbon (FCUL), his MSc in bromatology at the Faculty of Pharmacy of the University of Lisbon, and his PhD in analytical chemistry — metrology in chemistry at FCUL. The last two academic degrees were completed in parallel with his full-time professional experience as analyst, in official, public and private laboratories, of the different inorganic and organic analytes in various types of matrices using classical and instrumental methods of analysis.

This analytical experience was focused on the detailed validation of the measurement procedure, test quality control and evaluation of measurement uncertainty. Since 2002, Ricardo has worked regularly as an assessor of the Portuguese Accreditation Body (IPAC) and as a trainer and consultant for the accreditation of chemical laboratories. In 2009, Ricardo was contracted as a researcher by the Centre for Molecular Sciences and Materials of the Faculty of Sciences of the University of Lisbon where he has been continuing his research work on metrology in chemistry while collaborating in teaching at national and foreign universities. Ricardo's research includes the development of approaches for the detailed evaluation of the uncertainty associated with complex measurements and the assessment of the sources of lack of comparability of measurements in some analytical fields.

Ricardo has been a member of the IPAC Accreditation of Chemical Laboratories Working Group since 2006, the Eurachem/CITAC Measurement Uncertainty and Traceability Working Group since 2010, the Portuguese TrainMiC® team since 2008 and the TrainMiC® Editorial Board since 2010.

Ewa Bulska

Ewa Bulska obtained her PhD in analytical chemistry from the University of Warsaw (Poland) where she is currently a professor of analytical chemistry. Her research activity has been devoted mainly to the application of atomic and mass spectrometry in environmental, clinical and industrial fields. She is author and co-author of about 120 scientific papers and reports.

Since 2006, Ewa has been the head of the Polish Centre for Chemical Metrology and she chairs the committee of the postgraduate educational programme in metrology in chemistry for lifelong learning.

Since 2000, Ewa has been closely collaborating with Polish accreditation as a technical assessor for testing and calibration laboratories (EN ISO/IEC 17025). She is also the member of the technical committee on the accreditation of PT providers (EN ISO/IEC 17043). She was elected to be a representative of POLLAB in EUROLAB as well as various working groups (Proficiency Testing and Education and Training) of Eurachem.

Ewa has been involved with TrainMiC® since the beginning of the programme — currently as the Polish TrainMiC® Team Leader, Chair of the TrainMiC® Team Leader Council and member of the TrainMiC® Management and Editorial Boards. Ewa received special recognition for her contribution to the TrainMiC® programme in 2005 from the JRC-IRMM.

Beata Godlewska-Żyłkiewicz

Beata Godlewska-Żyłkiewicz obtained her PhD in analytical chemistry in 1995 from the University of Warsaw, Poland. She was granted short fellowships at the University of Liverpool (United Kingdom), the University of Genoa (Italy), the Aristotle University of Thessaloniki (Greece) and the University of Oviedo (Spain). Beata completed postgraduate studies in chemical metrology at the University of Warsaw in 2008.

Currently, Beata works at Institute of Chemistry of the University of Białystok as an associate professor. She lectures in analytical chemistry, environmental monitoring and chemical metrology. She has published over 50 papers in refereed journals devoted to different methods for the preconcentration and separation of trace elements, including solid phase extraction, electrolysis and biosorption and the application of atomic spectrometry in clinical and environmental analysis.

Beata has been involved with the TrainMiC® programme since 2004 as an authorised trainer of the Polish TrainMiC® team and member of the TrainMiC® Editorial Board.

Martina Hedrich

Martina Hedrich majored in chemistry and received her PhD degree from the Berlin Free University (FUB), Germany. Early research work included X-ray structure analysis of single crystals and the determination of trace elements in human body tissues and fluids with ET-AAS and ICP-OES. Joining BAM Federal Institute for Materials Research and Testing in 1989, Martina developed an affinity towards reference materials and characterised them with spectroscopic and classical methods of inorganic chemistry. She has been the head of working groups dealing with nuclear techniques, gas analysis, chemometrics and metrology and was involved in setting up a quality management system. Thus, Martina's areas of competence include analytical chemistry, reference materials, quality management and metrology in chemistry.

Currently, Martina is the Quality Manager of her Institute, convener of BAM's Certification Committee for Reference Materials, lecturer at FUB and involved in education and training programmes on the national and international level.

Since 2006, Martina has been a member of the TrainMiC® Editorial Board and — as a national team leader — coordinates TrainMiC® activities in Germany.

Nineta Majcen

Nineta Majcen started as a researcher at the University of Ljubljana (Slovenia), where she obtained her PhD on the validation of newly developed methods and chemometrics. She continued her analytical work in the quality control laboratories in industry before stepping into metrology activities at the national and European level.

In metrology, Nineta has been mainly involved in topics related to metrology in chemistry, issues related to metrological infrastructure and knowledge transfer activities. She also closely collaborates with accreditation and standardisation bodies and lectures as a guest lecturer at universities, postgraduate summer schools and at other knowledge transfer events. She is author of more than 200 bibliographic units in both research and expert areas.

Several international conferences, workshops, seminars and high-level events have been organised under Nineta's leadership — the Eurachem workshop on proficiency testing (2006), the European Association of National Metrology Institutes' (EURAMET) European Metrology Research Programme launch event (2008), Quality for South-Eastern European countries (2008), the TrainMiC® Convention (2009), Measurement Science in Chemistry summer school (2009), just to mention a few.

Nineta has proactively contributed to the TrainMiC® programme since the start of the iniative in 2001 and received special recognition for her contribution to the TrainMiC® programme in 2005 from the JRC-IRMM. She is the Slovenian TrainMiC® Team Leader, member of the TrainMiC® Management Board and chairs the TrainMiC® Editorial Board. She is the Slovenian representative in Eurachem and is a member of the working group for training and education.

Nineta is currently working as a EuCheMS secretary-general, where she is also in charge of policy development issues in the field of chemistry.

Bertil Magnusson

Bertil Magnusson started as a marine chemist looking for traces of metals in the oceans, rivers, lakes and rain in the 1970s. At that time, clean rooms and clean sampling was something totally new in the chemical laboratory. After completing his PhD, Bertil joined the chemical company Eka Chemicals within AkzoNobel and worked there as a specialist in analytical chemistry mainly with spectroscopy (XRF, XRD, ICP) and wet chemistry. The work included support for all laboratories within the company in Europe and America.

In 2002, Bertil joined SP Technical Research Institute of Sweden and is currently working in quality in measurements, metrology in chemistry, a research area on international comparability and traceability of chemical measurement results. His main work here is to participate in international cooperation between national metrology institutes and a major part is teaching and writing guidelines and research papers regarding measurement quality.

An important part of Bertil's work is education for analytical laboratories in QA/QC. In Nordic cooperation, Bertil has written the *Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories* (Nordtest Report 537), and *Internal Quality Control — Handbook for Chemical Laboratories* (Nordtest Report 569). In European cooperation, he participates in Eurachem and EUROLAB work with a focus on guidelines on uncertainty. In international cooperation, Bertil represents Sweden on the Consultative Committee for Amount of Substance — Metrology in Chemistry (CCQM) and, in this context, he is using isotope dilution ICP-MS, XRF and conductivity.

Since 2005, Bertil has participated in the TrainMiC® programme as National Team Leader for Sweden and and as a member of the TrainMiC® Editorial Board.

Snježana Marinčić

Snježana Marinčić works at the Institute of Public Health Dr Andrija Štampar in Zagreb, Croatia, and holds the position of Quality Manager. As Quality Manager she is involved in the QA/QC of measurements in the field of testing of samples from environmental, food and common use objects.

Snježana's long experience in analytical chemistry includes testing in the field of water examination, mainly wet chemistry and the application of liquid chromatography methods in the field of environmental and food analysis.

Snježana lectures and is a trainer on issues related to laboratory accreditation and metrology in chemistry such as the evaluation of measurement uncertainty and QA/QC measures. She actively collaborates with the Croatian Accreditation Agency where she is a member of the Working Group Interlaboratory Comparisons, and the Croatian Metrology Society where she is a member of Management Board.

Snježana is a member of the TrainMiC® Editorial Board and is the TrainMiC® National Team Leader for Croatia.

Ioannis Papadakis

Ioannis Papadakis studied earth sciences at the Aristotle University of Thessaloniki, Greece. He continued his education obtaining a scholarship from the European Commission at the JRC-IRMM mainly working on inorganic analysis using ICP-MS, resulting in a PhD in analytical chemistry, 'Introducing Traceability on Analytical Measurements', from the University of Antwerp, Belgium.

Ioannis then continued his work at the JRC-IRMM, organising interlaboratory comparisons in the framework of the IMEP programme and Key Comparisons and Pilot Studies in support of the BIPM MRA.

In 2001, Ioannis joined the Chemistry Department of the University of Cyprus in Nicosia as Visiting Assistant Professor, where he taught Environmental chemistry and quality of measurements.

In 2002, Ioannis returned to Greece and started his involvement in quality management. Since 2003, he has managed the accredited certification body International Quality Certification as its Chief Executive Officer.

Since 2002, Ioannis has supported the Hellenic Accreditation Council in various posts, currently chairing a technical committee responsible for accreditation of laboratories in various disciplines (e.g. measurements on petroleum products, clothes, leather, NDTs) according to EN ISO/IEC 17025.

Since 2007, Ioannis has been teaching and supervising theses on the postgraduate programme 'Quality Assurance' of the Hellenic Open University.

Ioannis has been involved with TrainMiC® since the beginning of the programme and is currently a TrainMiC® authorised trainer, member of the TrainMiC® Editorial Board and a team leader in the Greek TrainMiC® team.

Marina Patriarca

Marina Patriarca gained her PhD in chemistry from the Sapienza University of Rome (Italy) and her MSc in medical sciences from the University of Glasgow (United Kingdom). She joined the Italian National Institute of Health (Istituto Superiore di Sanità) in 1981, where she is still currently working as a senior research scientist. Her research activity has been devoted mainly to the application of atomic spectrometry and has involved the development and validation of analytical methods, including the estimate of uncertainty of measurement; population surveys for risk factors, including environmental exposure to metals; studies of the metabolism of copper and nickel in humans; development and organisation of external quality assessment schemes and assessment and certification of reference materials.

Marina is the author of more than 80 papers and a member of the Atomic Spectrometry Updates Editorial Board. Currently, she supports the quality system at her home institution by providing advice on metrology issues related to the implementation of the technical requirements of EN ISO/IEC 17025. Together with Enzo Ferrara, she represents Italy in Eurachem.

Marina has gained considerable experience in training practitioners, by lecturing in more than 50 courses and seminars for staff of public and private laboratories, devoted to aspects of quality assurance, implementation of ISO standards in testing laboratories and uncertainty of measurement. Recently, she has been involved in training activities for laboratory staff in developing countries.

Since 2006, Marina has been an authorised TrainMiC® trainer and, jointly with Antonio Menditto, coordinated the TrainMiC® activities in Italy as TrainMiC® Team Leader. She is also a member of the TrainMiC® Editorial Board.

Philip Taylor

Professor Philip Taylor has been in analytical chemistry since 1982. He completed his PhD at the University of Gent (Belgium) and started his career in R&D in industry before moving to the metrology institute of the European Commission, the Institute for Reference Materials and Measurements which is part of the Joint Research Centre.

At the JRC-IRMM, Philip heads a unit dealing with reference measurements and training related to metrology in chemistry (TrainMiC®, Measurement Science in Chemistry) to support the European Measurement Infrastructure. He has about 200 research papers to his name.

Philip is keen to ensure that metrology is relevant to today's needs in society, for instance in helping to implement European legislation. This involves training and education activities. He has also been very involved in technical assistance projects related to the enlargement of the EU. He has been rewarded for his endeavours through awards from the the Polish Chemical Society and the University of Maribor.

Philip initiated the TrainMiC® programme and chairs the TrainMiC® Management Board.

Emilia Vassileva

Emilia Vassileva is a research scientist and inorganic chemistry group leader in the IAEA-Environmental Laboratories in Monaco. She gained her master's degree in environmental analytical chemistry at the University of Geneva (Switzerland) and her PhD at the University of Sofia (Bulgaria), where she started as an assistant professor. Her main research interests are in the area of trace and ultra-trace isotope and elemental analysis using ICP-Mass Spectrometry and other advanced instrumental techniques. An important part of her research activities is devoted to reference measurements, development and validation of analytical methods, including the estimate of uncertainty of measurement. She is author and co-author of more than 100 scientific papers and reports.

Currently, Emilia supports the quality system at her home institution by acting as Contact Quality Point on all issues related to the implementation of the technical requirements of EN ISO/IEC 17025. She is actively involved in QA/QC training activities for laboratory staff in developing countries.

Emilia has contributed to the TrainMiC® programme since 2001 and has received special recognition for her contribution to the TrainMiC® programme in 2005 from the JRC-IRMM. She is a member of the TrainMiC® Editorial and Management Boards and is the TrainMiC® National Team Leader for Bulgaria.

Chapter 1

Measurement uncertainty — Part I Principles

Measurement uncertainty is an important EN ISO/IEC 17025 requirement. Two TrainMiC® presentations are dedicated to the uncertainty of measurement.

The first presentation (Principles) focuses on the general understanding of the uncertainty concept, highlighting that the aim of evaluation of uncertainty is to be able to make reliable decisions.

The second presentation (Approaches to evaluation) explains and demystifies the approach of the ISO-GUM (Guide to the expression of uncertainty in measurement) [5] to estimate and report the uncertainty of a measurement result obtained following a specific measurement procedure. A clear description of all the steps needed in the evaluation of uncertainty is presented with respective examples. The modelling approach for the estimation of measurement uncertainty is compared with single laboratory validation and interlaboratory validation approaches. This presentation gives guidance on the selection of the appropriate approach for different purposes and draws attention to the critical issues when applying the various approaches.

Uncertainty of measurement — Part I Principles

TrainMiC
Training in Metrology in Chemistry

Uncertainty of measurement

Part I Principles

Last updated - January 2011

TrainMiC
Training in Metrology in Chemistry

Aim

To familiarise users with the measurement uncertainty concept including its meaning, relevance, impact and evaluation principles.

This presentation aims to explain the measurement uncertainty concept.

TrainMiC
Training in Metrology in Chemistry

Modules

Uncertainty of measurement: Principles:

- Principles of the evaluation of the measurement uncertainty (MU);
- Modelling approach for the evaluation of the MU.

Uncertainty of measurement: Approaches to evaluation:

- Modelling approach (revision);
- Empirical approach based on interlaboratory data;
- Empirical approach based on INTRAlaboratory data.

This presentation is the first of two presentations dedicated to the evaluation of measurement uncertainty in analytical sciences.

The current presentation (MU-I) presents the internationally accepted principles of the evaluation of measurement uncertainty including relevant definitions and conventions. The application of these principles is illustrated with the use of the modelling approach for the evaluation of the uncertainty associated with measurements of the mass fraction of nitrate in fresh waters.

The second presentation (MU-II) goes further in presenting and comparing the three most popular approaches for the evaluation of the measurement uncertainty, namely the modelling approach, the empirical approach based on interlaboratory data and the empirical approach based on intralaboratory data.

TrainMiC
Training in Metrology in Chemistry

Overview

1. Introduction
2. Principles
3. Example
4. Highlights

 Uncertainty Principle 4.03 Slide 4

The overview of the presentation includes an explanation of the meaning and relevance of the measurement uncertainty concept (Introduction), the description of the principles of the evaluation of measurement uncertainty (Principles), the application of the presented evaluation of measurement uncertainty principles to the measurement of the mass fraction of nitrate in fresh water (Examples) and the most relevant message from this presentation (Highlights).

Introduction

TrainMiC
Training in Metrology in Chemistry

1. Introduction

TrainMiC
Training in Metrology in Chemistry

Overview

1. Introduction

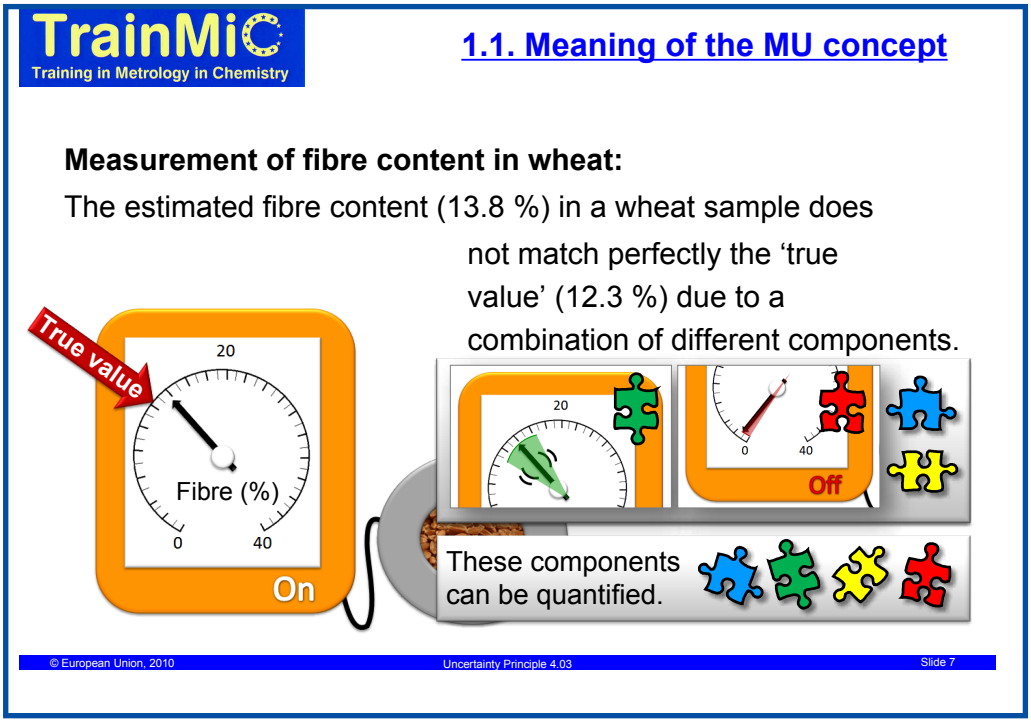
1.1. The meaning of the MU concept

1.2. Why do we need the MU concept?

1.3. Relevance of the MU concept

1.4. When should MU be evaluated?

Section 1 is divided into the subsections shown in the slide.



The meaning of the measurement uncertainty concept is illustrated with the result of the measurement of fibre content in a wheat sample. The measured quantity value ([1]) (i.e. the best estimation of the true value: 12.3 % (w/w)) does not match perfectly the 'true value' of the quantity ([2]) due to a combination of different reasons. These reasons could be (i) the concentration of extraction solutions, (ii) the time of extraction, (iii) the assigned value of used standards, (iv) the limited knowledge about the effect of the sample matrix on analyte extractability, etc. The uncertainty components can be quantified using measurement equations.

[1] **Measured quantity value** (VIM3: Entry 2.10): measured value of a quantity; measured value: **quantity value** representing a **measurement result** [1].

[2] **True quantity value** (VIM3: Entry 2.11): true value of a quantity; true value: **quantity value:** consistent with the definition of a **quantity** [1].

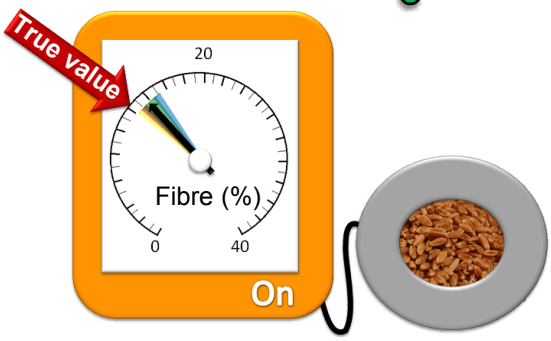
TrainMiC
Training in Metrology in Chemistry

1.1. Meaning of the MU concept

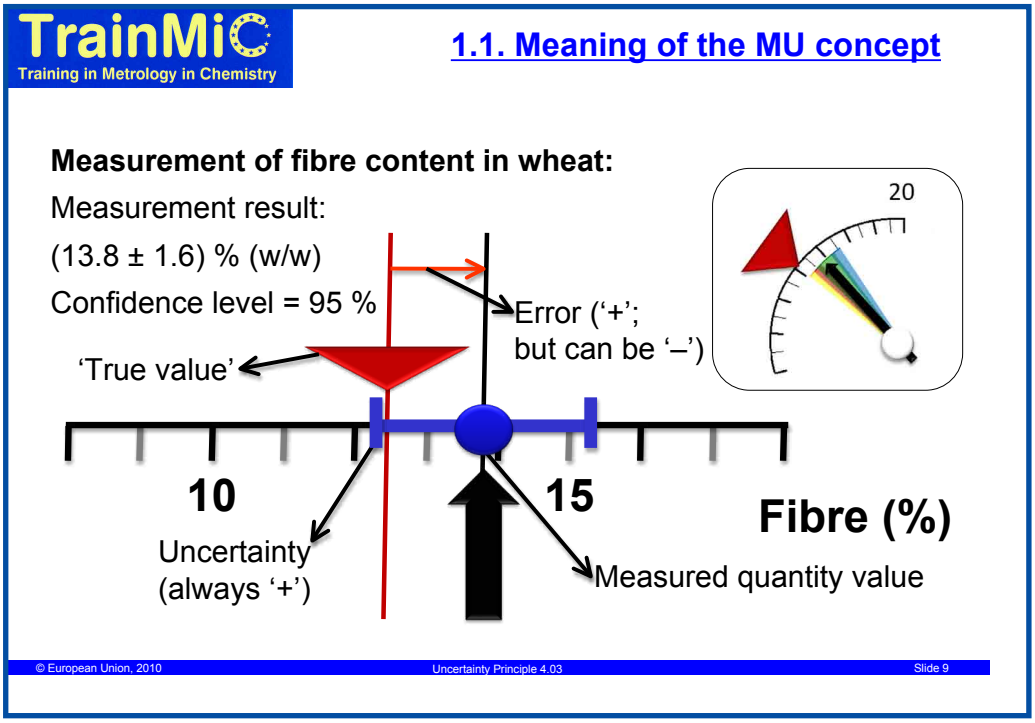
Measurement of fibre content in wheat:

(...) the quantified components can be combined in the measurement uncertainty that estimates a range of values that should encompass the 'true value' with a known probability.

Measurement result:
(13.8 ± 1.6) % (w/w)
Confidence level = 95 %



The quantified uncertainty components can be combined, using uncertainty model equations, aiming to estimate the measurement uncertainty that quantifies the range of values that should encompass the 'true value' of the measurand with known probability (the confidence level of the measurement uncertainty).



The difference between the measured quantity value, x, (13.8 % (w/w)) and the 'true value' of the measurand, T, is the measurement error (x − T). The error can be either a positive or a negative value depending on the relative positioning of x and T. The measurement uncertainty, MU, is a positive value that, in fact, should be larger than the modulus of the error with a probability equivalent to the confidence level of the reported measurement uncertainty.

TrainMiC
Training in Metrology in Chemistry

1.1. Meaning of the MU concept

Measurement uncertainty (VIM3 [1]: Entry 2.26):
Non-negative parameter characterising the dispersion of the **quantity values** being attributed to a measurand, based on the information used.

Measurand (VIM3 [1]: Entry 2.3):
quantity intended to be measured.

VIM3: JCGM 200:2012 — *International vocabulary of metrology — Basic and general concepts and associated terms* (http://www.bipm.org) [1].

The definition of measurement uncertainty, MU, in the latest version of the VIM (VIM3) [1], states the ambition of the measurement uncertainty concept of, together with the measured quantity value, producing intervals that should encompass the true value of the measurand (quantity values being attributed to a measurand). This definition also makes clear that the estimated measurement uncertainty depends on the available information about the measurement performance, quality of used references, model of uncertainty components combination, etc. For the same measurement, different measurement uncertainty values can be reported depending on the quality of uncertainty components evaluation and details of uncertainty combination models used.

The evaluation of the measurement uncertainty does not aim to estimate the 'best/ smallest' measurement uncertainty value. In many cases, pragmatic and simplified models for the evaluation of the measurement uncertainty are fit for the intended use of the measurement.

The measurement uncertainty concept is intimately related to the measurand concept since the measurand defines the quantity intended to be measured.

TrainMiC
Training in Metrology in Chemistry

1.1. Meaning of the MU concept

Defining the measurand is not trivial!

Measurand (VIM3 [1]: Entry 2.3):
quantity intended to be measured.

Clear definition of (a) the analysed item;
(b) the studied parameter.

Example A:

Measurand: Mass fraction of folpet pesticide in (...)

a.1) 2 tonnes of apples or a.2) 200 g apple sample

Before going further with the uncertainty concept, the measurand concept must be presented and discussed.

The defined measurand depends on the 'analysed item' and on the 'studied parameter'. The way the analysed item contributes to the definition of a measurand is illustrated with the determination of folpet fungicide in apples (Example A).

Measurement of the mass fractions of folpet fungicide in a 200 g or a 2 tonnes sample of apples are different metrological challenges since different items are involved. The measurement of folpet in 2 tonnes of the fruit must involve the study of the variability of the fungicide mass fraction in the large amount of apples and the modelling of the impact of the sampling procedure on the ability to estimate the mean folpet mass fraction in 2 tonnes of apples. Therefore, in this case, measurement uncertainty due to sampling must be included in the uncertainty budget. For the measurements of the mass fraction of folpet in 200 g of apples, only analytical steps affect the reliability of the measurement result *r* (i.e. sampling uncertainty is not to be considered).

TrainMiC
Training in Metrology in Chemistry

1.1. Meaning of the MU concept

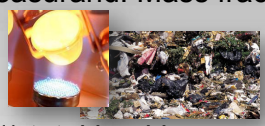
Defining the measurand is not trivial!

Measurand (VIM3 [1]: Entry 2.3):
quantity intended to be measured.

Clear definition of (a) the analysed item;
(b) the studied parameter.

Example B:

Measurand: Mass fraction of (....)

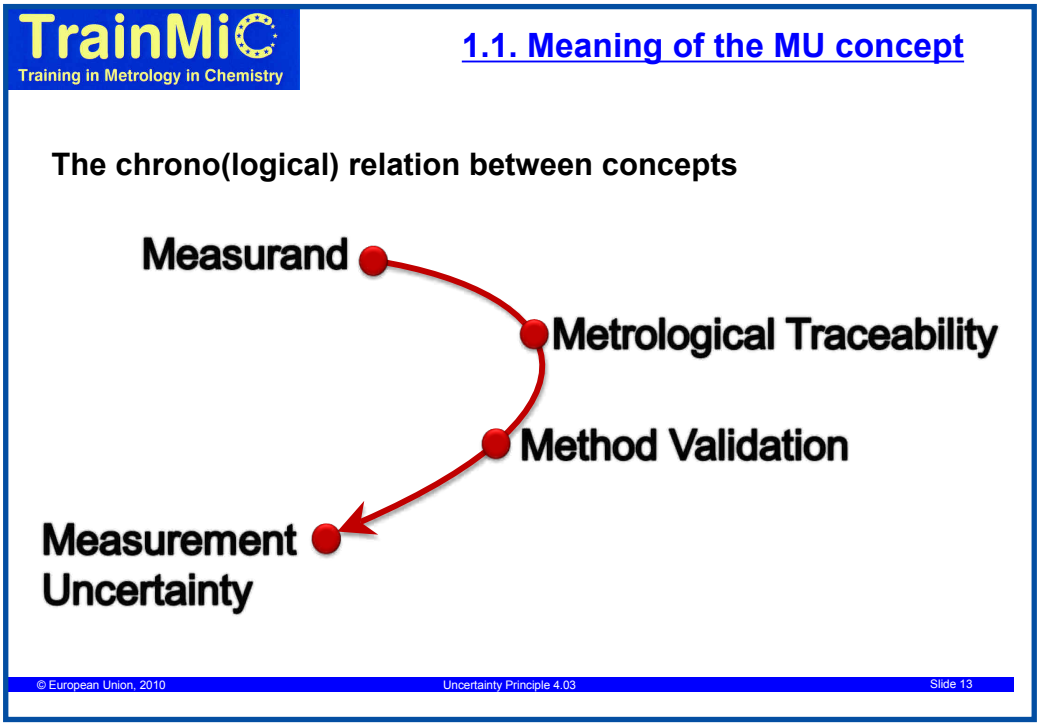


b.1) total lead in an industrial residue sample.

b.2) water soluble lead according to DIN 38414 standard [3] in an industrial residue sample.

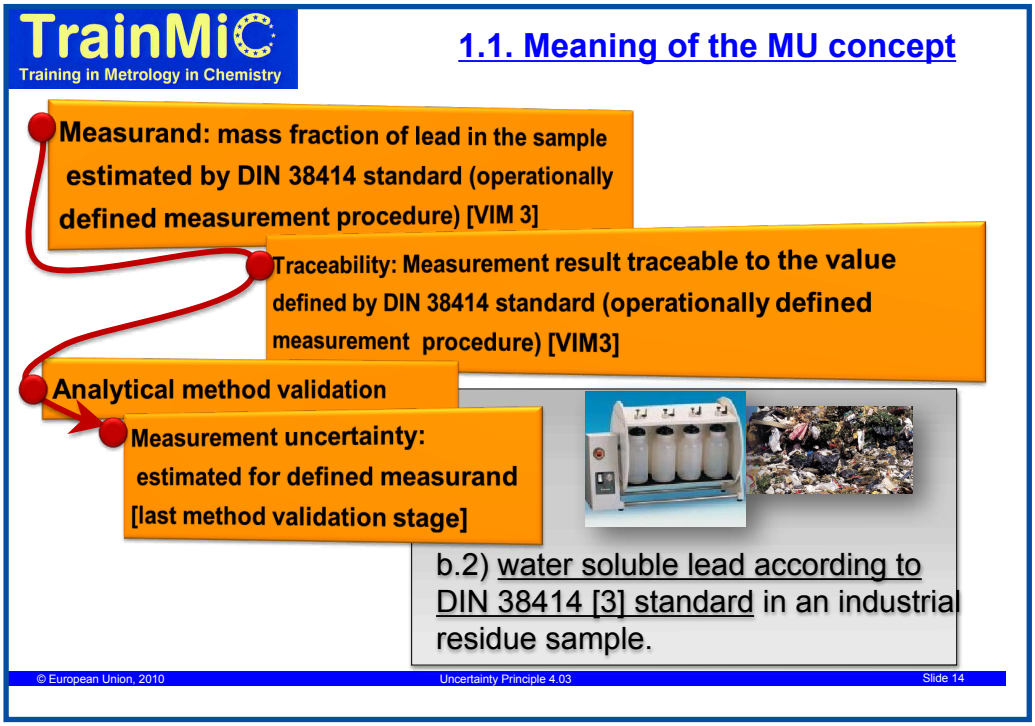
The way the studied parameter contributes to the definition of a measurand is illustrated with the the measurement of the content of lead in an industrial residue sample (Example B). The customer can be interested in either the total lead content or the water soluble lead content. The total lead content is the target if an efficient lead recycling protocol is to be implemented. On the other hand, it is relevant to check the water soluble lead fraction if the residues are to be stored in a solid waste landfill from which lead can leach from rain water into the soil. In both these cases, the same analyte and sample is associated with different parameters.

The definition of the measurand is not trivial and is essential before we start analysing a sample. It must be linked to the aim of the analysis.



The concepts 'measurand', 'metrological traceability', 'method validation' and 'measurement uncertainty' are intimately related in a 'logical' and chronological way [2].

The analytical process should start with the definition of the measurand. The metrological traceability of the result is defined when the reference for the measurement is selected and its role in the measurement equation decided (e.g. correcting relevant bias). After the measurement procedure is selected, its validation is performed and collected validation data are used for the evaluation of the measurement uncertainty of the result.



The previously described chronological relationships can be illustrated with an example.

1. Measurand: the mass fraction of water soluble lead in a sample with reference code XY determined according to DIN 38414 standard [3];
2. Metrological traceability statement: the measurement result is traceable to the reference value as defined by DIN 38414 standard;
3. Validation: validation of the measurement procedure includes the estimation of the performance parameters of the procedure and the assessment of fitness of the measurement procedure for the intended use;
4. Evaluation of uncertainty: measurement uncertainty is estimated from the data available and collected, mostly, from measurement procedure validation.

TrainMiC
Training in Metrology in Chemistry

1.2. Why do we need the MU concept?

- **It is an intrinsic part of the measurement result** (measured quantity value ± measurement uncertainty) units (...).
- **It allows the objective interpretation of the measurement result** (e.g. sample compliance evaluation with a legislation).
- **It allows for checking of the quality of the performed measurement considering its intended use:** MU should be smaller than the target MU (VIM3 [1]: Entry 2.34).
- **It can support the optimisation of measurement procedures for cost and performance** (in particular in the modelling approach).

Measurement uncertainty needs to be estimated since it is an intrinsic part of the measurement result (not an addend to the measurement result). Its value allows an objective and independent interpretation of the measurement result and can be used to check quality and prove the adequacy of the measurement for its intended use. A detailed uncertainty budget can also be used for the optimisation of the measurement procedure aiming at cost and/or uncertainty magnitude reduction.

TrainMiC
Training in Metrology in Chemistry

1.3. Relevance of the MU concept

• **EN ISO/IEC 17025:2005 — *General requirements for the competence of testing and calibration laboratories*** [4]

This international standard for the accreditation of testing laboratories defines that competent laboratories should evaluate their measurement uncertainty (...).

• **EU Legislation for official control:** some EU legislation states that official control must be performed in accredited laboratories (e.g. Article 12, Regulation (EC) No 882/2004).

• **EU Legislation for contaminants in food:** some EU legislation states than foodstuff compliance with contamination limits must be judged considering estimated MU (e.g. Regulation (EC) No 401/2006).

 Uncertainty Principle 4.03 Slide 16

The relevance of the evaluation of measurement uncertainty for a competent presentation of the measurement result is evident from EN ISO/IEC 17025 [4] as well as from legislation. An accredited laboratory must be able to report quantitative measurement results with uncertainty and guide clients on the interpretation of results considering measurement uncertainty. Some EU legislation specifies how measurement uncertainty must be considered in its enforcement — two examples now follow.

(a) Regulation (EC) No 882/2004 of the European Parliament and of the Council of 29 April 2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules,
Article 12: '2. However, competent authorities may only designate laboratories that operate and are assessed and accredited in accordance with the following European standards' (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2004R0882:20060525:EN:PDF).

(b) Commission Regulation (EC) No 401/2006 of 23 February 2006 laying down the methods of sampling and analysis for the official control of the levels of mycotoxins in foodstuffs: 'Acceptance of a lot or sublot — acceptance if the laboratory sample conforms to the maximum limit, taking into account the correction for recovery and measurement uncertainty' (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:070:0012:0034:EN:PDF).

TrainMiC
Training in Metrology in Chemistry

1.4. When should MU be evaluated?

- **When a new analytical procedure is introduced**
- **When the scope of an analytical method is extended** (e.g. Method scope extended to samples with more complex matrix)
- **When relevant changes are introduced in the analytical procedure** (e.g. new equipment, analysts with different expertise, new analytical steps introduced or removed, etc.)
- **When analytical method performance variations are observed** (e.g. Increasing or decreasing precision observed on control charts)

Measurement uncertainty must be evaluated in the situations specified in the slide.

Measurement uncertainty does not need to be estimated for each measurement if pragmatic models estimate measurement uncertainty for the whole analytical range independently of between-day measurement performance variations.

TrainMiC
Training in Metrology in Chemistry

2. Principles

TrainMiC
Training in Metrology in Chemistry

Overview

2. Principles

2.1. The ISO-GUM (Guide to the expression of uncertainty in measurement) [5]

2.2. The Eurachem/CITAC [6], Nordtest [7] and EUROLAB [8] guides

2.3. Steps in the evaluation of the MU

2.4. How results should be reported

2.5. How results should be compared

2.6. Alternative approaches for the evaluation of the MU

The following section presents the internationally accepted principles of the evaluation of the measurement uncertainty.

This section is divided into the subsections shown in the slide.

TrainMiC
Training in Metrology in Chemistry

2.1. The ISO-GUM

The ISO-GUM (Guide to the expression of uncertainty in measurement) presents the internationally accepted principles for the evaluation of the MU [5].

(...) Evaluation effort should not be disproportionate:

'The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement procedure', ISO-GUM, para. 3.4.8 (*).

(*) **ISO-GUM** — JCGM 100:2008 — *Evaluation of measurement data – Guide to the expression of uncertainty in measurement* (http://www.bipm.org) [5].

The internationally accepted principles of the evaluation of measurement uncertainty are presented in the ISO Guide to the expression of uncertainty in measurement (GUM) [5]. The GUM is applicable to any type of quantitative measurements (e.g. chemical, physical, biological measurements). This guide states that the evaluation effort should not be disproportionate considering the intended use of the measurement.

TrainMiC
Training in Metrology in Chemistry

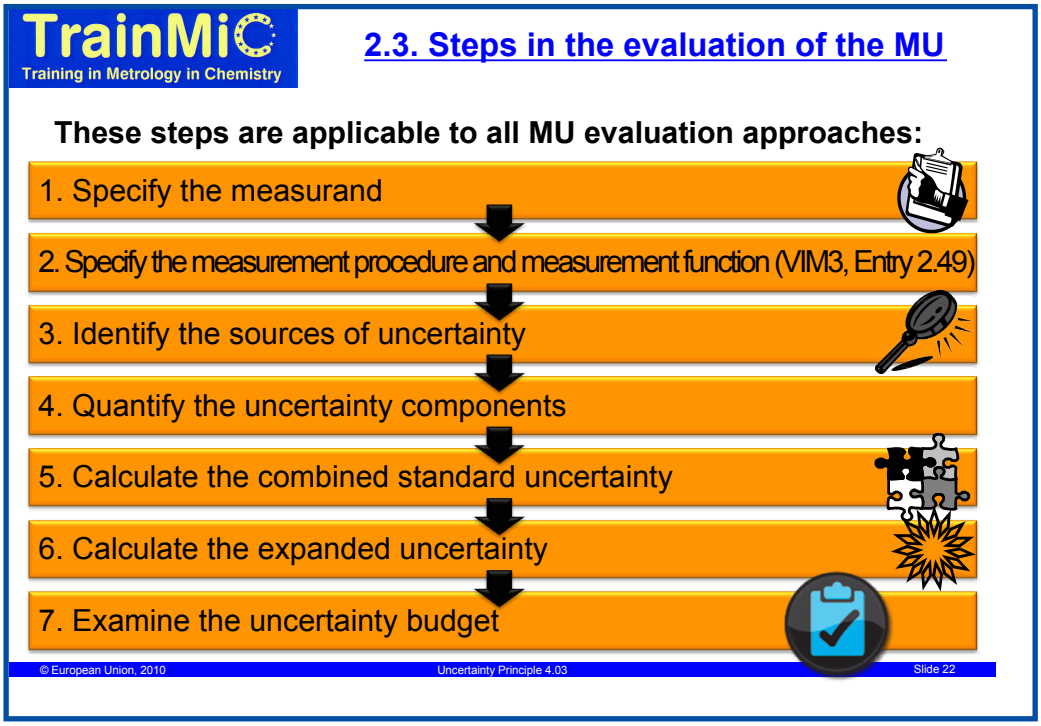
2.2. The Eurachem/CITAC, Nordtest and EUROLAB guides

The application of the GUM general principles to measurements in chemistry is described in these guides:

- Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement*, Second edition, 2000 (http://www.eurachem.org) [6];
- Nortdtest TR537, *Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories*, 2004 (http://www.nordicinnovation.net) [7];
- EUROLAB Technical Report No 1/2007, *Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation* (http://www.eurolab.org) [8].

There are several guides (based on the GUM principle) available for the evaluation of uncertainty of measurements in chemistry: these guides can be downloaded for free from the Internet.

- The Eurachem/CITAC Guide CG4 [6] presents a comprehensive description of detailed and pragmatic approaches for the evaluation of measurement uncertainty.
- The Nordtest TR537 guide [7] presents, in a simple way, the most pragmatic approaches for the evaluation of measurement uncertainty.
- The EUROLAB Technical Report No 1/2007 [8] discusses how various approaches for the evaluation of measurement uncertainty should be applied.



Independently of the approach for the evaluation of measurement uncertainty used, the following sequence must be followed. The time and effort needed for each step depends on the applied approach for the evaluation of measurement uncertainty and on the available data.

'Measurement function' is a new term (VIM3 [1]: Entry 2.49) for 'Model equation'.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

1. Specify the measurand

Differences in the defined measurand are often responsible for the incompatibility of measurement results.

Uncertainty Principle 4.03
Slide 23

The definition of a measurand is the first step in the evaluation of the measurement uncertainty. This step is not trivial, as has been previously explained. Misunderstandings and mistakes in defining or properly informing about the considered measurand are often reasons for the incompatibility of measurements obtained by different laboratories. If the same item is analysed by two laboratories for supposedly the same parameter (e.g. see DIN 38414 example) the laboratories will obtain incompatible results ([3]). In fact, the measurands are different and, thus, the results are not comparable.

([3]) **Metrological compatibility of measurement results** (VIM3: Entry 2.47): metrological compatibility; property of a set of **measurement results** for a specified **measurand**, such that the absolute value of the difference of any pair of **measured quantity values** from two different measurement results is smaller than some chosen multiple of the **standard measurement uncertainty** of that difference [1].

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

2. Specify the measurement procedure and measurement function

The measurement procedure is selected considering the defined measurand and other criteria (e.g. target MU and costs).

On many occasions, the measurement function ($Y = f(x_a, x_b, x_c, x_d)$; where Y is the output quantity and x_a, x_b, x_c and x_d are the input quantities from variables a, b, c and d, respectively), established together with the measurement procedure, is updated after the following stage (Identify the sources of uncertainty).

Measurement function:
$Y = f(x_a, x_b, x_c, x_d)$

After defining the measurand, the measurement procedure is selected and the measurement function is written. The measurement procedure must also be chosen considering the target measurement uncertainty ([4]) [1], the available resources, the cost of analysis, etc.

On many occasions, the measurement function must be updated after the identification of the sources of uncertainty. The addition of unitary multiplying factors needed to model the impact of relevant uncertainty components on the measurement is frequently applied.

([4]) **Target measurement uncertainty** (VIM3: Entry 2.34): target uncertainty; **measurement uncertainty** specified as an upper limit and decided on the basis of the intended use of **measurement results** [1].

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

3. Identify the sources of uncertainty

The identified sources of uncertainty must reflect all effects that affect measurement result.

The impact of possible mistakes and blunders is not to be considered since their occurrence and impact is unpredictable.

The identification of the sources of uncertainty is another demanding step in the evaluation of the measurement uncertainty. Understanding which effects can affect measurement quality is not trivial, even for simple measurements.

The impact of mistakes and blunders, such as performing molecular spectrophotometric measurements on solutions with suspended particles, on measurement quality is not to be considered in the evaluation of the measurement uncertainty. This type of mistake is easily detected and should be overcome by repeating the measurement.

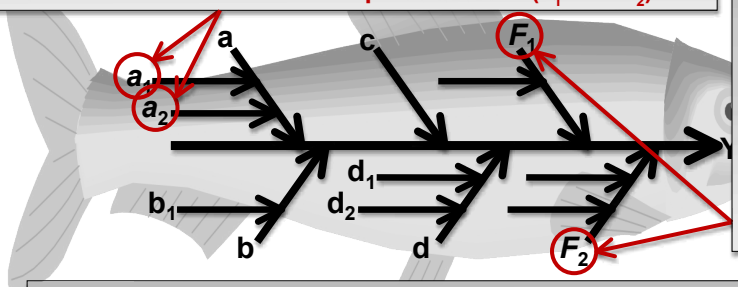
TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

3. Identify the sources of uncertainty

Cause/effect diagrams can be used to avoid double-counting or forgetting relevant effects or correlations between variables.

Input variable *a* represents a sources of uncertainty that can be divided into two input variables (a_1 and a_2)



F_1 and F_2 represent two influencing variables, not included in the original measurement function, that reflect relevant effects (...)

Updated measurement function: $Y = f(x_{a_1}, x_{a_2}, x_{b_1}, x_{d_1}, x_{d_2}, x_c, x_{F_1}, x_{F_2})$

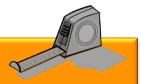
The elaboration of cause/effect diagrams (also known as fishbone diagrams) can help analysts in avoiding forgetting or double-counting relevant uncertainty components. These diagrams have a major vector, converging to the measurand, to which secondary vectors, representing sources of uncertainty, converge. The secondary vectors can also be fragmented in uncertainty components that reflect specific effects on secondary uncertainty components. Whenever useful for the quantification and combination of the uncertainty components, the uncertainty components (represented by a single vector) can be combined in the same vector. This combination is frequently used for components reflecting the precision of single steps, since the global method precision quantifies the combined effect of all these components.

Although cause/effect diagrams are useful tools, analysts still need to be extremely careful in defining the problem to be solved during the evaluation of measurement uncertainty.

Sometimes, it is necessary to take into account unitary (i.e. equal to 1) influence variables (traditionally not used in the calculation of the 'measured quantity value') to cover relevant sources of uncertainty. An example: the impact of temperature oscillations in the estimation of the water soluble lead, at 20 ºC, from an aliquot of a sample of an industrial residue.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU



4. Quantify the uncertainty components

The quantification of the uncertainty components is divided into:

Type A evaluations (VIM3) [1]: 'evaluation of a component of measurement uncertainty by a statistical analysis of measured quantity values obtained under defined measurement conditions'.

In this case, the uncertainty component is quantified in ideal conditions since all information for the reliable estimation of that effect on measurement uncertainty are available (...)

The identified sources of uncertainty are subsequently quantified using the developed measurement function. Well-known models for the quantification of the uncertainty associated with volumetric, gravimetric and instrumental quantification steps are available in the bibliography [5, 6].

The quantification of uncertainty components is divided into two types.

Type A evaluations, performed in conditions where all information about the magnitude of an effect has been provided by experiments in your laboratory.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

4. Quantify the uncertainty components

The quantification of the uncertainty components is divided into:

Type B evaluations (VIM3) [1]: 'evaluation of a component of measurement uncertainty determined by means other than a Type A evaluation of measurement uncertainty'.

> These evaluations are performed when no resources or data needed to gather all objective information about the magnitude of a source of uncertainty are available.

GUM [5] proposes harmonised and pragmatic solutions to overcome these limitations.

Type B evaluations, where approximations to deal with the lack of objective information about the magnitude of the component must be considered.

GUM [5] presents conventions to harmonise type B evaluations for the most frequent scenarios.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

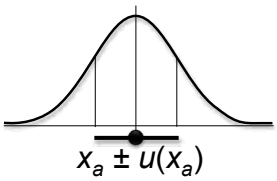
4. Quantify the uncertainty components

(...)

The uncertainty components are quantified as 'standard uncertainties' (u).

Considering variable a, $u(x_a)$ defines an interval ($x_a \pm u(x_a)$) that should encompass the 'true value' of that variable with a confidence level of 68 %.

The uncertainty components are quantified as needed for their combination (propagation of uncertainty law).



The uncertainty components are quantified as standard uncertainties (u) needed for their combination using the law of propagation of uncertainty. The interval built from the best estimation of the input quantity, a, and its standard uncertainty, u_a, ($a \pm u_a$) should encompass the true value of the variable with a confidence level of 68.3 % resembling a standard deviation.

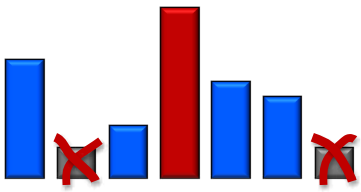
TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

4. Quantify the uncertainty components

(...)

Usually, minor uncertainty components (with values less than one fifth of the major component) do not need to be quantified.



Uncertainty Principle 4.03
Slide 30

The uncertainty components which have been proven after approximate calculations to be minor (with values less than one fifth that of the major component) do not need to be quantified or combined with the other uncertainty components, since they will not change significantly the estimated measurement uncertainty.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

5. Calculate the combined standard uncertainty

Several approaches can be followed to combine the estimated uncertainties:

- law of propagation of uncertainty;
- numerical methods:
  - — Kragten method (easily implemented in a spreadsheet) ([1]);
  - — Monte Carlo method (needs dedicated software) ([2]).

([1]) Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement*, Second edition, 2000, Appendix E.2 (http://www.eurachem.org) [6].

([2]) JCGM 101:2008 — *Evaluation of measurement data — Supplement 1 to the 'Guide to the expression of uncertainty in measurement' — Propagation of distributions using a Monte Carlo method*, First edition (http://www.bipm.org) [9].

This presentation puts forward the most popular way of combining the uncertainty components: the law of propagation of uncertainty.

There are numerical alternatives to the propagation of uncertainty law such as the numerical Kragten method [6] and the Monte Carlo method [9].

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

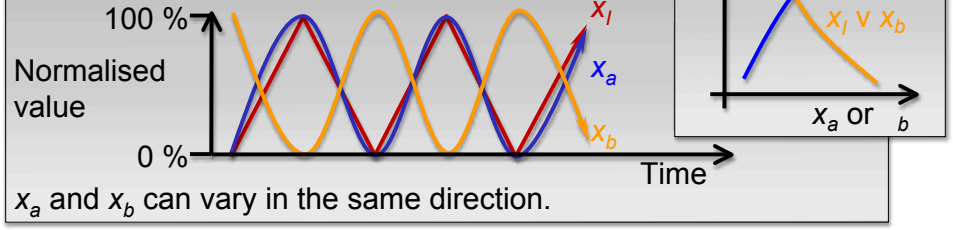
5. Calculate the combined standard uncertainty

5.1. Law of propagation of uncertainty;

When the uncertainty components are independent considering all relevant input and influence quantities:

Example A:

Variables *a* and *b* are correlated considering variable *I*:



x_a and x_b can vary in the same direction.

The simplified version of the law of propagation of uncertainty, presented in the following slides, is only applicable if input quantities are independent considering the variation of influence quantities (i.e. quantities that affect the input quantity value). The slide presents an example of three variables, two input quantities (x_a and x_b) from the measurement function, and one influence quantity (x_i) from which the variation affects input quantities values. The normalised value axes represent the ratio between the value of the variable and the maximum value observed within the studied period of time. Input quantities are correlated since when x_i increases, the x_a value increases and the x_b value decreases affecting the output variable value in a correlated way. The correlated effect can either increase or decrease the uncertainty estimated assuming an independence of input quantities.

Examples of correlated variables:

- variable *a* (solution volume (x_a)): an increase in temperature (x_i) produces an increase in the volume of the solution (x_a);
- variable *b* (molar absorptivity of the analyte (x_b)): an increase in temperature (x_i) produces a decrease in the molar absorptivity of the analyte.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

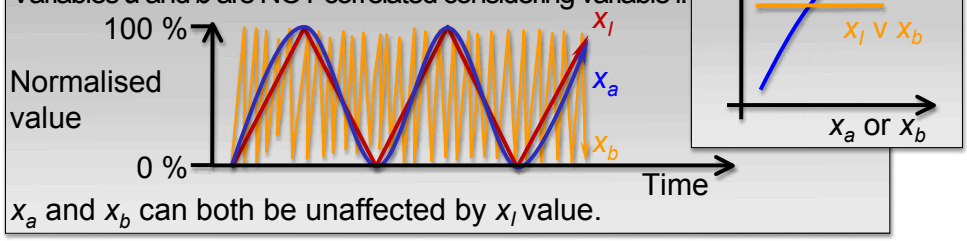
5. Calculate the combined standard uncertainty

5.1. Law of propagation of uncertainty:

When the uncertainty components are independent considering all relevant input and influence quantities:

Example B:

Variables *a* and *b* are NOT correlated considering variable I:



x_a and x_b can both be unaffected by x_I value.

Uncertainty Principle 4.03
Slide 33

When only input quantity x_a is influenced by the x_i value, the input quantities (i.e. x_a and x_b) are not correlated and can be considered independent. In this case, the simplified version of the law of propagation of uncertainty can be used. The observed variation of x_b in the normalised value axes results from the measurement precision.

Examples of not correlated variables:

- variable *a* (solution volume (x_a)): an increase in laboratory temperature (x_i) produces an increase in the volume of the solution (x_a);
- variable *b* (response of GC detector (x_b)): an increase in laboratory temperature (x_i) does not affect GC detector response.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

5. Calculate the combined standard uncertainty

5.1. Law of propagation of uncertainty:

When the uncertainty components are independent considering all relevant input and influence quantities ($Y = f(x_a, x_b, x_c, x_d)$):

$$u(Y) = \sqrt{\left(\frac{\partial Y}{\partial x_a}\right)^2 u(x_a)^2 + \left(\frac{\partial Y}{\partial x_b}\right)^2 u(x_b)^2 + \left(\frac{\partial Y}{\partial x_c}\right)^2 u(x_c)^2 + \left(\frac{\partial Y}{\partial x_d}\right)^2 u(x_d)^2} =$$

$$= \sqrt{\sum \left(\frac{\partial Y}{\partial x_i}\right)^2 u(x_i)^2}$$

Combined standard uncertainty
(notation $u_c(Y)$ is common)

According to the simplified version of the law of propagation of uncertainty for independent variables, the standard uncertainty associated with the output quantity ($u(Y)$) is the square root of the weighted sum of squares of the standard uncertainties associated with the input quantities $(u(x_i))^2$ where the weighted factors are the squares of their partial derivatives $(\partial Y/\partial x_i)^2$.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

5. Calculate the combined standard uncertainty

5.1. Law of propagation of uncertainty (independent variables):

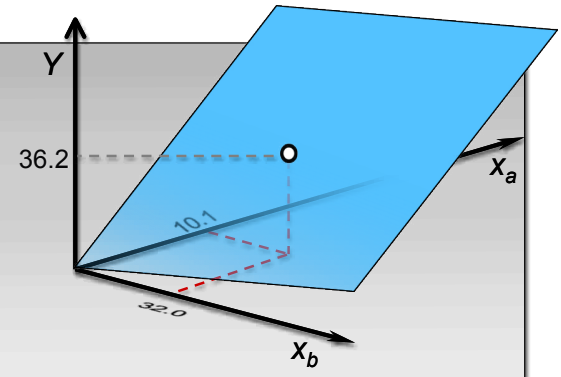
Example A:

$Y = 2x_a + 0.5x_b$:

(...)

if $x_a = 10.1$ and $x_b = 32.0$;

$Y = 36.2$ (units)



This slide illustrates the combination of the uncertainties associated with the input quantities, x_a and x_b, used to calculate the output quantity Y ($Y = 2 \cdot x_a + 0.5 \cdot x_b$). The two input quantities and the output quantity can be represented together in a 3D graph (Y v x_a v x_b). In this graph, it is also represented by the point ($Y/x_a/x_b$: 36.2/10.1/32.0: respective units have been omitted).

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

5. Calculate the combined standard uncertainty

5.1. Law of propagation of uncertainty (independent variables):

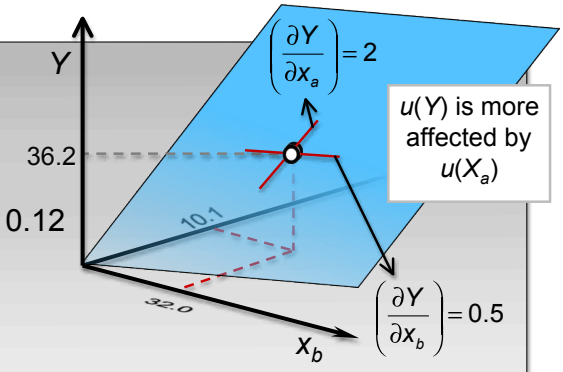
Example A:

$Y = 2x_a + 0.5x_b$:

$$u(Y) = \sqrt{(2)^2 \times u(x_a)^2 + (0.5)^2 \times u(x_b)^2}$$

(...) if $u(x_a) = 0.35$ and $u(x_b) = 0.12$

$$u(Y) = \sqrt{(2)^2 \times (0.35)^2 + (0.5)^2 \times (0.12)^2} = 0.87$$



It is evident from the law of propagation of uncertainty that the contribution of an uncertainty component to the output variable uncertainty depends both on the magnitude of the standard uncertainty ($u(x_i)$) and on the magnitude of the respective partial derivative ($(\partial Y/\partial x_i)$). The partial derivative represents the slope of the tangent of the function Y v x_i. The uncertainty component contribution is larger when the Y value increases more with the increment in the x_i value. In this example, the uncertainty $u(x_a)$ is the major source of uncertainty.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

5. Calculate the combined standard uncertainty

5.1. Law of propagation of uncertainty (independent variables):

Example B:

Particular cases of the law of propagation of uncertainty:

If: $$Y = k + k_a \times x_a + k_b \times x_b + k_c \times x_c + k_d \times x_d$$

(...) where k, k_a, k_b, k_c and k_d are constant values:

$$u(Y) = \sqrt{\left[k_a \times u(x_a)\right]^2 + \left[k_b \times u(x_b)\right]^2 + \left[k_c \times u(x_c)\right]^2 + \left[k_d \times u(x_d)\right]^2}$$

The general law of propagation of uncertainty, for independent input quantities, can be simplified for linear relationships as in the slide.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

5. Calculate the combined standard uncertainty

5.1. Law of propagation of uncertainty (independent variables):

Example C:

Particular cases of the law of propagation of uncertainty:

If: $$Y = \frac{k \times x_a \times x_b}{x_c \times x_d}$$

Relative standard uncertainty

(...) where *k* is a constant value:

$$\frac{u(Y)}{Y} = \sqrt{\left(\frac{u(x_a)}{x_a}\right)^2 + \left(\frac{u(x_b)}{x_b}\right)^2 + \left(\frac{u(x_c)}{x_c}\right)^2 + \left(\frac{u(x_d)}{x_d}\right)^2}$$

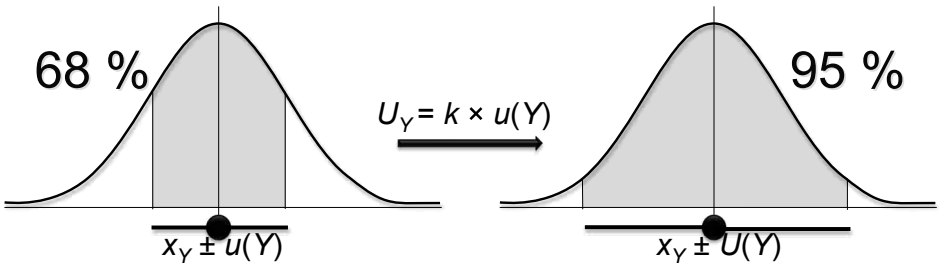
The general law of propagation of uncertainty, for independent input quantities, can be simplified for multiplying relationships as in the slide. When the output quantity is calculated from the multiplication and/or division of the input quantities, the relative standard uncertainty of the output quantity ($u(Y)/Y$) is estimated by the square root of the sum of the squares of the relative standard uncertainties of the input quantities $(u(x_i)/x_i)^2$.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

6. Calculate the expanded uncertainty

This stage aims at expanding the confidence level associated with the 'combined standard uncertainty' ($u(Y)$) from 68 % to 95 % or 99 %. This expansion involves the use of a multiplying factor (coverage factor, k) to estimate an expanded uncertainty (U).



The confidence level of the combined standard uncertainty ($u(Y)$) (i.e. 68.3 %) must be expanded to a higher level (usually 95 % or 99 %) before measurement of the result being reported or interpreted. The expansion is performed by multiplying the combined standard uncertainty ($u(Y)$) with an adequate multiplying factor (k) called the coverage or expansion factor. The resulting 'expanded uncertainty' is represented by a capital 'U' ($U(Y) = k \cdot u(Y)$).

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

6. Calculate the expanded uncertainty

In most cases, the amount of information combined in the 'combined standard uncertainty' guarantees a high number of degrees of freedom associated with the estimated result.

In these cases, the following approximations can be performed:

- Confidence level of approximately 95 %: $U(Y) = 2 \times u(Y)$ ($k = 2$);
- Confidence level of approximately 99 %: $U(Y) = 3 \times u(Y)$ ($k = 3$).

Example A:

Considering $u(Y) = 0.87$:
The expanded uncertainty ($U(Y)$) is: $2 \times 0.87 = 1.74$ (units) for a confidence level of approximately 95 %.

When a thorough quantification of major uncertainty components is performed, so as to guarantee reliable evaluation of the measurement quality, the estimated combined standard uncertainty is associated with a high number of degrees of freedom. In these cases, coverage factors, k, of 2 or 3 can be used to expand the uncertainty to confidence levels of approximately 95 % or 99 % respectively. The word 'approximately' should be used when stating the confidence level to make it clear that an approximated model was used.

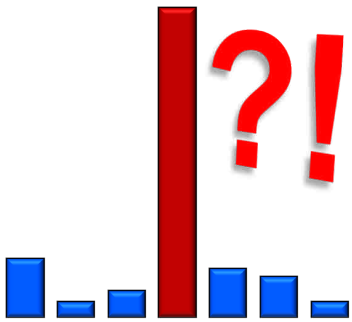
The slide presents the expansion of the standard uncertainty estimated for Example A previously given.

TrainMiC
Training in Metrology in Chemistry

2.3. Steps in the evaluation of the MU

7. Examine the uncertainty budget

The careful examination of the relative magnitude of the uncertainty components allows the detection of mistakes



After estimation of the expanded uncertainty, the uncertainty budget should be checked to identify possible mistakes in calculations or defined assumptions. If an expected minor source of uncertainty is a major uncertainty component, measurement function and corresponding calculations should be checked.

TrainMiC
Training in Metrology in Chemistry

2.4. How results should be reported

• The MU should be reported with a maximum of two significant figures following adequate number rounding rules:

Example A1:

Considering $U(Y) = 1.74$ (units), it should be reported 1.7 (units).

• The measured quantity value should be reported with the same number of decimal places reported in the MU. The units, coverage factor and confidence level should be also clearly reported:

Example A2:

Considering $x_Y = 36.20$ (units are, for instance, mg L^{-1}), it should be reported: (36.2 ± 1.7) mg L^{-1}

For a confidence level of approximately 95 % considering a coverage factor of 2.

The results should be reported in a harmonised way to avoid misinterpretation of their meaning. The slide presents conventions, described in GUM [5], that should be followed for reporting expanded uncertainties.

TrainMiC
Training in Metrology in Chemistry

2.5. How results should be compared

Results from different measurements and/or samples should be compared considering the uncertainty associates with their difference (u_d) and defining a target zero value.

Example A1:

The same wine sample was analysed in two laboratories for procimidone pesticide concentration. The following results were reported by both laboratories:

Lab 1: 26.9 ± 2.7 µg L^{-1} (k = 2; confidence level of 95 %);

Lab 2: 30.7 ± 4.9 µg L^{-1} (k = 3; confidence level of 99 %).

Results from two measurements, for instance obtained from the analysis of the same item by two laboratories or obtained from the analysis of two items by the same laboratory, must be compared taking the respective measurement uncertainty into account. Two measurement results are compatible if the confidence interval of their difference, d, with a high confidence level (typically, approximately 95 % and 99 %) include the target zero value ($d \pm k \cdot u_d$).

Example A1 illustrates the comparison of results from two measurements of procimidone in the same wine sample obtained by two laboratories.

TrainMiC
Training in Metrology in Chemistry

2.5. How results should be compared

Example A1:

- (…) the standard uncertainties from both measurement results are:
  Lab 1: $u_{Lab\ 1}$ = 2.7/2 = 1.35 µg L^{-1} (k = 2; conf. level of 95 %);
  Lab 2: $u_{Lab\ 2}$ = 4.9/3 = 1.63 µg L^{-1} (k = 3; conf. level of 99 %).
- (…) the standard uncertainty, u_d, associated with the difference, d (d = 26.9 – 30.7 = – 3.8 µg L^{-1}) between both results is:

$$u_d = \sqrt{\left(u_{Lab\ 1}\right)^2 + \left(u_{Lab\ 2}\right)^2} = \sqrt{\left(1.35\right)^2 + \left(1.63\right)^2} = 2.12 \text{ µg } L^{-1}$$

- (…) since the respective expanded uncertainty (U_d) for a confidence level of approximately 95 % is 4.24 µg L^{-1} (2 × 2.12= 4.24),and this value is larger than $|d|$ = 3.8, it can be concluded that: ?

(...) the measurement results are compatible (VIM3 [1]: Entry2.47) (i.e. metrologically equivalent).

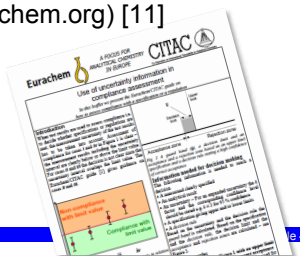
The expanded uncertainties from both measurements must be converted to standard uncertainties before being combined in the standard uncertainty of the difference, u_d. The standard uncertainty of the difference, u_d, is calculated from the equation previously presented for linear relationships between variables (Example B in Section 5.1). A coverage factor of 2 is used to expand the standard uncertainty of the difference. The measurement results are compatible since $|d| < (2 \cdot u_d)$.

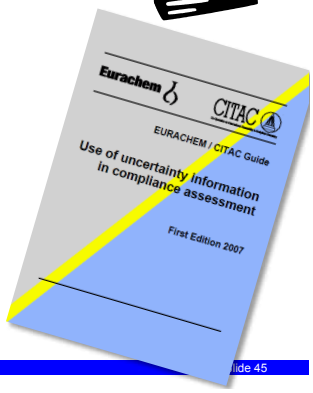
TrainMiC
Training in Metrology in Chemistry

2.5. How results should be compared

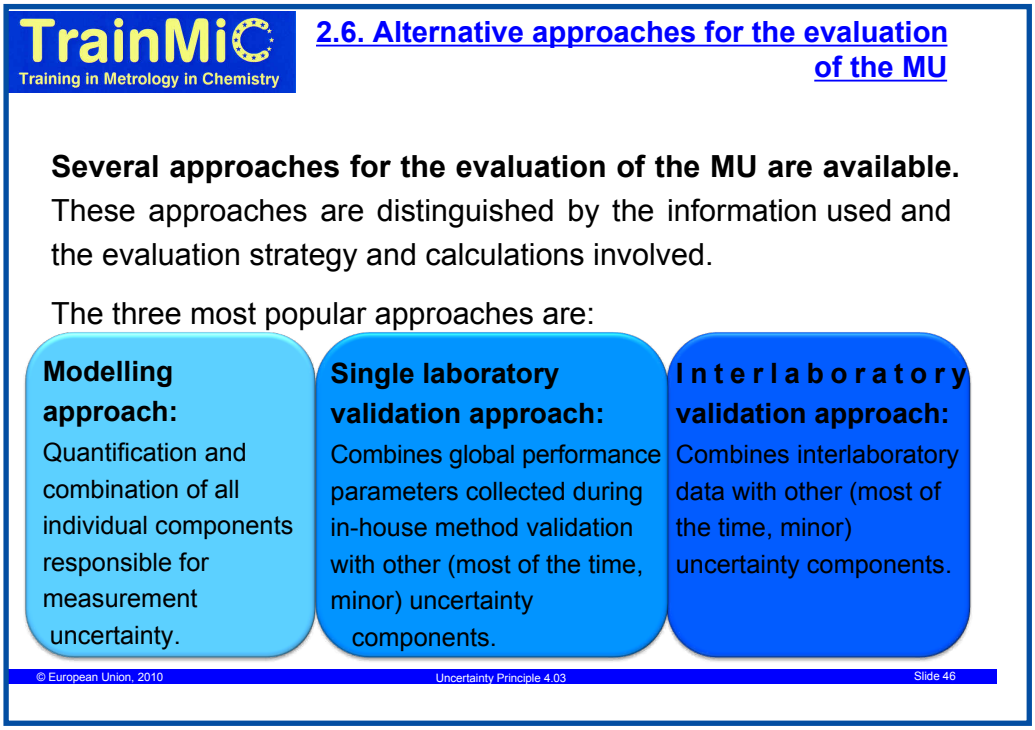
For the evaluation of the compliance of a result with a reference limit, see the following references:

- Eurachem/CITAC Guide, *Use of uncertainty information in compliance assessment,* First edition, 2007 (http://www.eurachem.org) [10]
- Eurachem/CITAC Information leaflet, *Use of uncertainty information in compliance assessment,* 2009 (http://www.eurachem.org) [11]





As shown in the slide, the assessment of the compliance of an analysed sample with reference limits is discussed in references [10] and [11].



Several approaches for the evaluation of measurement uncertainty, based on the general principles of GUM [5], are known. The three most popular approaches are:

1. the modelling approach
2. the single laboratory validation approach and
3. the interlaboratory validation approach.

In the modelling approach (1), individual uncertainty components that contribute to measurement uncertainty are quantified and combined. In the popular single laboratory validation approach (2), intermediate precision and bias are evaluated and their impact on measurement result is quantified as two independent uncertainty components. These uncertainty components are combined, as relative standard uncertainties, with other components (most of the time minor) that also contribute to measurement uncertainty but were not reflected in the previous validation data (e.g. sample heterogeneity).

In the interlaboratory validation approach (3), reproducibility, most of the time reflecting the major sources of uncertainty, is combined with any other uncertainty components, that contribute to measurement uncertainty, but were not reflected on the previous data (e.g. sample heterogeneity).

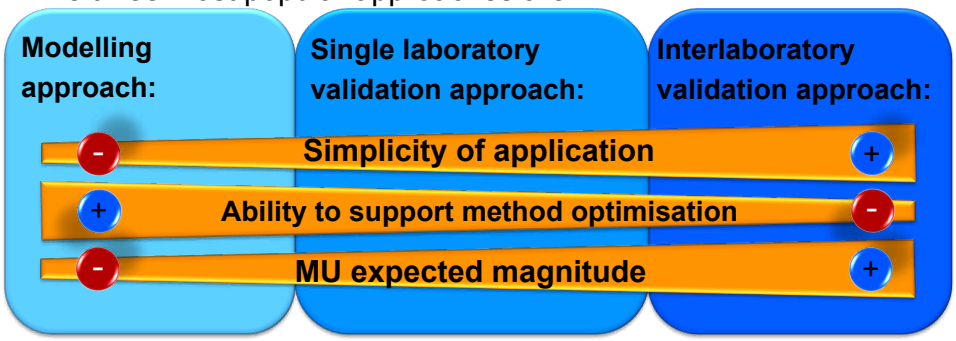
Other valid approaches for the evaluation of measurement uncertainty are available and described in the bibliography.

TrainMiC
Training in Metrology in Chemistry

2.6. Alternative approaches for the evaluation of the MU

Several approaches for the evaluation of the MU are available.
These approaches are distinguished by the information used and the evaluation strategy and calculations involved.

The three most popular approaches are:

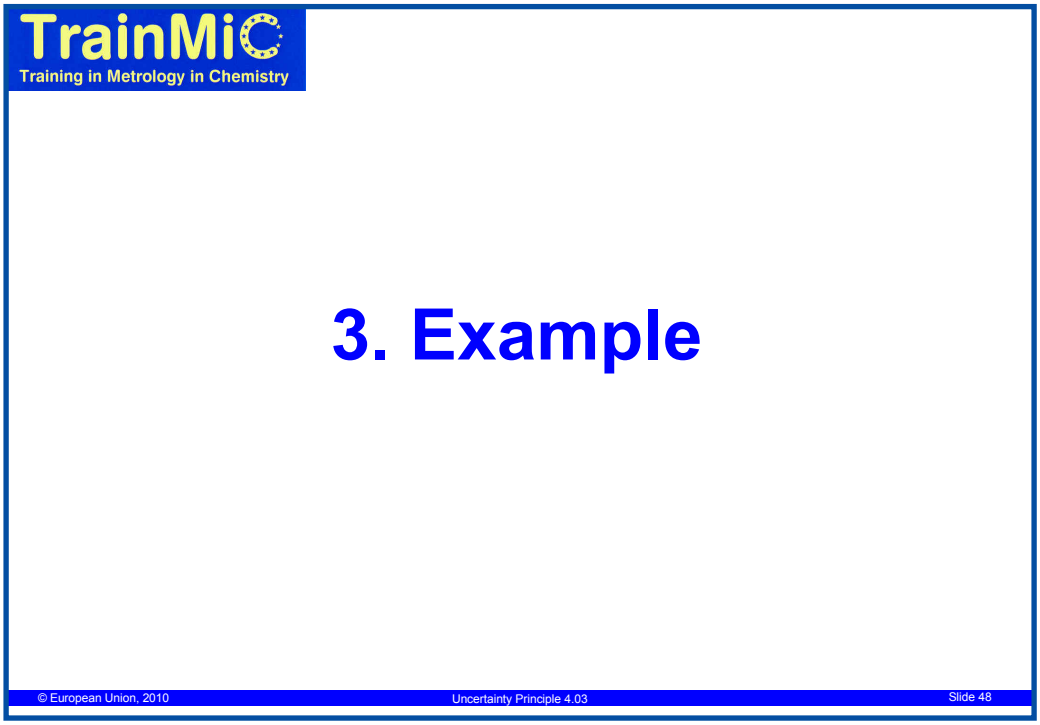


The interlaboratory validation approach is the one that involves simpler algorithms since many uncertainty components are combined in measurement reproducibility.

The modelling approach allows measurement procedure optimisation for cost or measurement uncertainty magnitude reduction.

The modelling approach usually results in estimates of uncertainty which are smaller because detailed models of the measurement performance are developed. The other more pragmatic approaches involve elaboration of simplified models of measurement performance that will not describe, as accurately, how measurement performs for a specific case.

The above mentioned trend is observed in evaluations performed by all three approaches.



This section illustrates the described principles by explaining the evaluation of uncertainty associated with the measurement of the mass fraction of nitrate in fresh waters by ion chromatography.

TrainMiC
Training in Metrology in Chemistry

Overview

3. Example

3.1. Problem description (Measurement of the mass fraction of nitrate in drinking water)

3.2. Metrological traceability

3.3. Measurement procedure validation

3.4. Evaluation of the MU (modelling approach)

3.5. Conclusion

This section is divided into the subsections shown in the slide.

TrainMiC
Training in Metrology in Chemistry

3.1. Problem description

- Measurement of the mass fraction of nitrate in fresh water samples by ion chromatography:

— Definition of the metrological traceability;

— Brief description of the analytical method validation;

— Evaluation of the measurement uncertainty (Modelling approach).

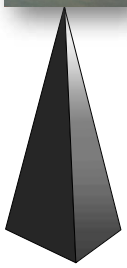
Measurand: Nitrate mass fraction in a specific fresh water sample (e.g. nitrate mass fraction in water sample with reference number 10/1524).

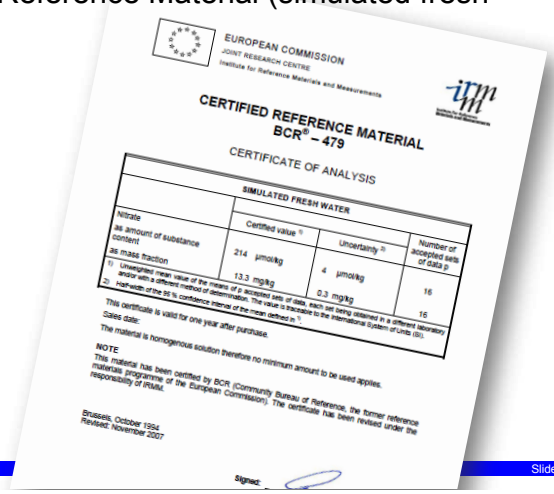
This slide presents the steps (in chronological order) needed for the measurement of the mass fraction of nitrate in fresh water. The last stage of this process is the evaluation of measurement uncertainty. The measurand is defined as previously described.

TrainMiC
Training in Metrology in Chemistry

3.2. Metrological traceability

Measurement results will be traceable to the nitrate mass fraction of the BCR-479 Certified Reference Material (simulated fresh water).





EUROPEAN COMMISSION
JOINT RESEARCH CENTRE
Institute for Reference Materials and Measurements

CERTIFIED REFERENCE MATERIAL
BCR® – 479

CERTIFICATE OF ANALYSIS

SIMULATED FRESH WATER

	Certified value [1]	Uncertainty [2]	Number of accepted sets of data [3]
Nitrate as amount of substance content	214 µmol/kg	4 µmol/kg	16
as mass fraction	13.3 mg/kg	0.3 mg/kg	16

[1] Unweighted mean value of the means of 4 accepted sets of data, each set being obtained in a different laboratory and/or with a different method of determination. The value is traceable to the International System of Units (SI).
[2] Half-width of the 95 % confidence interval of the mean defined in [1].

This certificate is valid for one year after purchase.
Sales data:
The material is homogenous solution therefore no minimum amount to be used applies.

NOTE
This material has been certified by BCR (Community Bureau of Reference, the former reference materials programme of the European Commission). The certificate has been revised under the responsibility of IRMM.

Brussels, October 1994
Revised: November 2007

Signed:

The metrological traceability of the measurement result was defined after it was decided that measurement results would be corrected for bias obtained from the analysis of the certified reference material BCR-479. The performed correction aims to establish the measurement traceability to the value embodied in the stated certified reference material. Therefore, measurements are traceable to the certified value of the mass fraction of nitrate in BCR-479. The certified value has mass fraction units (mg of nitrate in kg of water).

TrainMiC
Training in Metrology in Chemistry

3.3. Measurement procedure validation

Measurement procedure validation involves the following steps:

- Evaluation of the Limit of Quantification;
- Evaluation of the instrumental response linearity;
- Repeatability test;
- Intermediate precision test;
- Trueness test: Replicated analysis of BCR-479 Certified Reference Material in intermediate precision conditions.

Validation of this measurement procedure involves the following steps:

1. estimation of the limit of quantification;
2. evaluation of the linearity of the ion chromatographer instrumental response;
3. assessment of the measurement repeatability;
4. assessment of the measurement intermediate precision;
5. assessment of the measurement trueness through replicated analysis of the BCR-479 under intermediate precision conditions.

The measurement procedure validation ends with the evaluation of measurement uncertainty that includes an assessment of its fitness for the intended use.

TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.1. Specify the measurand

Measurand: Nitrate mass fraction in a specific fresh water sample.

3.4.2. Specify the measurement procedure and measurement function

Measurement procedure: Direct measurement of the nitrate mass fraction, in a sample aliquot, by ion chromatography after multi-point calibration with calibration standards prepared in pure water with known mass fraction. Initially, the estimated measurement result is corrected for analyte recovery observed on the analysis of the BCR-479 CRM.

The following slides present the application of the previously described steps in the evaluation of measurement uncertainty for the measurement of the mass fraction of nitrate in fresh waters.

The definition of the measurand and the selection of the measurement procedure are the first steps in this evaluation. The measurement procedure specifies that measurement results are corrected for bias as measured in the analysis of BCR-479: the decision for correcting for the bias on measurement result was taken when its traceability was defined.

TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.1. Specify the measurand

Measurand: Nitrate mass fraction in a specific fresh water sample.

3.4.2. Specify the measurement procedure and measurement function

Application example:

$$w = \frac{w_{Init}}{\overline{R}} = \frac{w_{Init} \times C_{CRM}}{\overline{C}_{obs}}$$

w: mass fraction = 2.348 mg kg^{-1};

w_{Init}: initially estimated mass fraction **(2.38 mg kg^{-1})**

$\overline{R}$: mean analyte recovery **(0.987)**

C_{CRM} : certified mass fract. (BCR-479)**(13.3 mg kg^{-1})**

$\overline{C}_{obs}$: mean estimated mass fraction of the CRM **(13.48 mg kg^{-1})**

The measurement function is presented on the left-hand side of the slide together with the bias correction factor (reverse of the mean analyte recovery). On the right-hand side of the slide, and in blue, example results in the measurement of the nitrate mass fraction in a fresh water sample are presented (2.348 mg kg^{-1}).

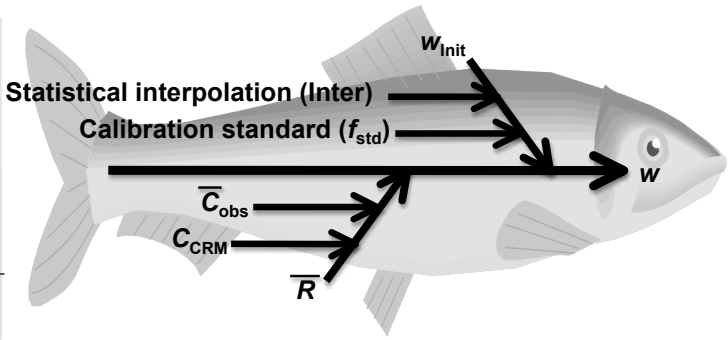
TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.3. Identify the sources of uncertainty

Measurement function Updating:

$$w = \frac{w_{Init} \times f_{std}}{\overline{R}} = \frac{w_{Init} \times C_{CRM} \times f_{std}}{\overline{C}_{obs}}$$



An additional unitary factor (f_{std} = 1) must be considered to allow for the calibration standards' uncertainty.

The demanding identification of the sources of uncertainty is presented in the cause/effect diagram. All input quantities are uncertainty components represented by vectors in the cause/effect diagram. The uncertainty associated with the initially estimated mass fraction (w_{Init}) is affected by two uncertainty components:

1. statistical interpolation uncertainty estimated by the regression model; and
2. uncertainty associated with the concentration of calibration standards.

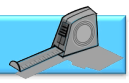
The assumption in the least squares regression model related to standard concentrations is: calibration standards' relative (not absolute) concentrations must be affected by negligible uncertainty ([5]) [6, 12]. A unitary multiplying factor (f_{std}) must be added to the measurement function to allow for the calibration standards uncertainty.

([5]) Eurachem/CITAC Guide CG4 [6], p. 77: in this guide, the following approximation for the least squares regression model is defined: 'Therefore the usual uncertainty calculation procedures for c_0 only reflect the uncertainty in the absorbance and not the uncertainty of the calibration standards, nor the inevitable correlations induced by successive dilution from the same stock. In this case, however, the uncertainty of the calibration standards is sufficiently small to be neglected'.

TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.4. Quantify the uncertainty components



(i) f_{std} The relative standard uncertainty ($u_{f_{std}}/f_{std}$) associated with f_{std} factor is estimated by excess, in a pragmatic way, by the relative standard uncertainty ($u_{C_{std}}/C_{std}$) associated with the concentration of the calibration standard with lowest concentration (highest relative standard uncertainty):

$$\frac{u_{f_{std}}}{f_{std}} = \frac{u_{C_{std}}}{C_{std}} = 0.00841$$

Calculations were performed as presented in Example A.1 in the Eurachem/CITAC CG4 Guide CG4 (http://www.eurachem.org) [6].

The next stage in the evaluation of uncertainty is the quantification of the uncertainty components.

The relative standard uncertainty associated with the factor f_{std} is estimated by excess, in a pragmatic way of the relative standard uncertainty ($u_{C_{std}}/C_{std}$) associated with the concentration of the calibration standard with lowest concentration (highest relative standard uncertainty) [12].

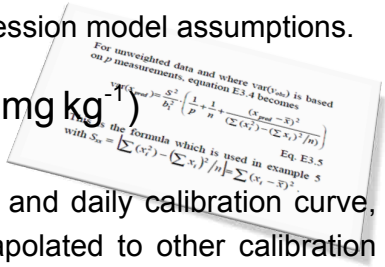
TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.4. Quantify the uncertainty components

(ii) *Inter* The interpolation standard uncertainty (u_{inter}) was estimated using Equation 3.5 from Appendix E.3 of the Eurachem/CITAC Guide CG4 (http://www.eurachem.org) [6]. This equation was applied after checking regression model assumptions.

$$u_{inter} = 0.1743 \text{ mg kg}^{-1} \text{ (for 2.38 mg kg}^{-1}\text{)}$$


For unweighted data and where $\text{var}(y_{obs})$ is based on p measurements, equation E3.4 becomes

$$\text{var}(x_{est}) = \frac{s_r^2}{b_1^2}\left(\frac{1}{p} + \frac{1}{n} + \frac{(x_{est} - \bar{x})^2}{(\sum x_i^2) - (\sum x_i)^2/n}\right)$$ Eq. E3.5

Here, the formula which is used in example 5 with $S_{xx} = \left[\sum (x_i^2) - (\sum x_i)^2/n\right] = \sum (x_i - \bar{x})^2$

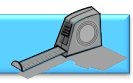
Since u_{inter} varies with the concentration and daily calibration curve, the reported u_{inter} value cannot be extrapolated to other calibration curves and/or concentrations.

The interpolation uncertainty was estimated from the specific multi-point calibration curve and sample signal using equations from the linear regression model [6]. This model was applied after a careful evaluation of the validity of the regression model assumptions.

TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.4. Quantify the uncertainty components



(iii) $\bar{R}$ The standard uncertainty associated with the estimated analyte recovery, $\bar{R}$, results from the combination of the uncertainty associated with $\bar{C}_{obs}$ and C_{CRM}.

Application example:

$$u_{\bar{R}} = \bar{R} \times \sqrt{\left(\frac{s_{C_{obs}}}{\bar{C}_{obs} \times \sqrt{m}}\right)^2 + \left(\frac{U_{C_{CRM}}/k}{C_{CRM}}\right)^2} = \left(\frac{13.48}{13.3}\right) \times \sqrt{\left(\frac{0.8937}{13.48 \times \sqrt{12}}\right)^2 + \left(\frac{0.3/2}{13.3}\right)^2} = 0.02252$$

(...)

u_i is the standard uncertainty associated with variable i

U_i is expanded uncertainty associated with variable i

$s_{C_{obs}}$ is the standard deviation from m replicated analysis used to estimate $\bar{C}_{obs}$

k is the coverage factor associated with $U_{C_{CRM}}$

The relative standard uncertainty of the mean analyte recovery, estimated from the analysis of the BCR-479 under intermediate precision conditions, results from the combination of two components as relative standard uncertainties (terms inside the square root operation): the first term represents the relative standard deviation of the mean recovery reflecting the impact of the measurement precision on mean recovery; the second term represents the relative standard uncertainty of the certified value that estimates how the quality of the certified value affects the quality of the measurements of samples with unknown nitrate mass fraction.

TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.5. Calculate the combined standard uncertainty

Since the measurement function involves the multiplication and division of the input quantifies:

Application example:

$$u_w = w \times \sqrt{\left(\frac{u_{\bar{R}}}{\bar{R}}\right)^2 + \left(\frac{u_{inter}}{W_{Init}}\right)^2 + \left(\frac{u_{f_{std}}}{f_{std}}\right)^2} =$$

$$= 2.348 \times \sqrt{\left(\frac{0.02252}{1.0135}\right)^2 + \left(\frac{0.1743}{2.38}\right)^2 + \left(\frac{0.00841}{1}\right)^2} =$$

$$= 0.1808 \text{ mg kg}^{-1}$$

u_i is the standard uncertainty associated with variable i

All uncertainty components are combined following the particular case of the law of propagation of uncertainty for multiplicative relationships (Section 5.1, Example C).

TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.6. Calculate the expanded uncertainty

The expanded uncertainty is estimated considering a coverage factor of 2 for a confidence level of approximately 95 %:

$$U_w = 2 \times u_w = 2 \times 0.1808 = 0.3616 \text{ mg kg}^{-1}$$

The measurement results is reported as:

$$(2.35 \pm 0.36) \text{ mg kg}^{-1} (*)$$

(*) For a confidence level of approximately 95 % considering a coverage factor of 2.

A coverage factor of 2 is used to expand the standard uncertainty to a confidence level of approximately 95 %.

The coverage factor and confidence level used must be reported together with the result.

TrainMiC
Training in Metrology in Chemistry

3.4. Evaluation of the MU

3.4.7. Examine the uncertainty budget

Since the uncertainty components are combined as relative standard uncertainties:

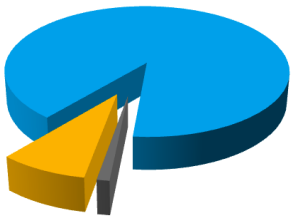
$$\frac{u_w}{w} = \sqrt{\left(\frac{u_{\bar{R}}}{\bar{R}}\right)^2 + \left(\frac{u_{inter}}{W_{Init}}\right)^2 + \left(\frac{u_{f_{std}}}{f_{std}}\right)^2}$$

(…), their percent contribution, p (%), is estimated by:

$$p_{\bar{R}} = \left(\frac{u_{\bar{R}}}{\bar{R}}\right)^2 \Big/ \left(\frac{u_w}{w}\right)^2$$

$$p_{inter} = \left(\frac{u_{inter}}{W_{Init}}\right)^2 \Big/ \left(\frac{u_w}{w}\right)^2$$

$$p_{f_{std}} = \left(\frac{u_{f_{std}}}{f_{std}}\right)^2 \Big/ \left(\frac{u_w}{w}\right)^2$$



- $p_{inter} = 90.5\,\%$
- $p_{\bar{R}} = 8.3\,\%$
- $p_{f_{std}} = 1.2\,\%$

Uncertainty Principle 4.03
Slide 61

The uncertainty budget is examined from the percentage contribution of the uncertainty components point of view. The presented equations for estimating the contribution to the uncertainty budget were derived from the way uncertainty components were combined. This information can be used to detect possible mistakes in calculations and can be further used for optimisation of costs or uncertainty magnitude reduction.

Examples of proposals for optimisation are:

1. decreasing measurement uncertainty: calibrate the chromatography instrument in a linear range associated with a smaller relative repeatability (typically at higher concentrations);
2. cost reduction: use cheaper (more uncertain) calibration standards that would not affect significantly (u_w/w) value.

TrainMiC
Training in Metrology in Chemistry

3.5. Conclusion

• The MU value (0.36 mg kg^{-1}; U_w/w = 15.3 %) is only applicable to the studied mass fraction (i.e. 2.35 mg kg^{-1}) and calibration curve since u_{inter} was estimated in repeatability conditions.

• The developed MU model: $u_w = w \times \sqrt{\left(\frac{u_{\bar{R}}}{\bar{R}}\right)^2 + \left(\frac{u_{inter}}{W_{Init}}\right)^2 + \left(\frac{u_{f_{std}}}{f_{std}}\right)^2}$

(....) is only applicable to undiluted samples with mass fractions within the calibration range and analysed using the studied daily calibration.

• The relative MU value (U_w/w = 15.3 %) is fit for intended use since is smaller than the relative target MU (i.e. 20 %).

The expanded uncertainty estimated by the modelling approach is only applicable to this respective measurement as u_{inter} varies with the concentration level and daily calibration curve.

The developed model for the combination of uncertainty components is not applicable for samples subjected to dilution before ion chromatographic determination. In such cases, the uncertainty associated with sample dilution must be added to the uncertainty budget. Since the relative expanded uncertainty (i.e. 15 %) is smaller than the relative target uncertainty (i.e. 20 %), the measurement is fit for the indented use.

TrainMiC
Training in Metrology in Chemistry

4. Highlights

TrainMiC
Training in Metrology in Chemistry

4. Highlights

• Measurement uncertainty (MU) does not imply doubt about the validity of a measurement; on the contrary, knowledge of the uncertainty implies increased confidence in the validity of a measurement result.

• MU defines a tolerance around the ‘measured quantity value’ that should encompass the ‘true value’ of the measurand with known probability.

• MU is essential for the objective and transparent evaluation of the measurement result meaning.

• Different approaches for the evaluation of the MU based on GUM [5] principles are available depending on the used information.

Chapter 2

Measurement uncertainty — Part II Approaches to evaluation

Part II of the presentation on measurement uncertainty (Approaches to evaluation) explains and demystifies the approach of the ISO-GUM (Guide to the expression of uncertainty in measurement) [5] used to estimate and report the uncertainty of a measurement result obtained following a specific measurement procedure. A clear description of all the steps needed in the evaluation of uncertainty is presented with respective examples. The modelling approach for the estimation of measurement uncertainty is compared with the single laboratory validation and interlaboratory validation approaches. This presentation gives guidance on the selection of the appropriate approach for different purposes and draws attention to the critical issues when applying the various approaches.

TrainMiC
Training in Metrology in Chemistry

Uncertainty of Measurement

Part II Approaches to Evaluation

Last updated - January 2011

The aim of this presentation is to put forward the different approaches to the evaluation of uncertainty. The different approaches are presented mainly according to the EUROLAB Technical Report No 1/2007, *Measurement uncertainty revisited: Alternative approaches to uncertainty evaluatio*n [8].

TrainMiC
Training in Metrology in Chemistry

Aim

- To present the three main approaches to the evaluation of measurement uncertainty (MU) based on the GUM [5] (*) principles
- To give guidance on the selection of approach for different purposes
- To give guidance on reporting uncertainty over the concentration range

* JCGM 100:2008 — *Evaluation of measurement data — Guide to the expression of uncertainty in measurement* (GUM), 2008 (http://www.bipm.org) [5].

Different approaches can be selected for the evaluation of measurement uncertainty, depending on the purpose and available data. In this presentation, three major approaches are considered. In practice, it is often a combination of approaches that is used.

Frequently, the uncertainty varies over the concentration range for which the procedure is applicable: uncertainty can be reported in absolute (e.g. mg kg^{-1}) or in relative units (%).

TrainMiC
Training in Metrology in Chemistry

Overview

- Literature for the evaluation of uncertainty
- Uncertainty as a function of concentration
- Three commonly used approaches:
  - — modelling approach
  - — single laboratory validation & QC approach
  - — interlaboratory validation approach
- Steps in the evaluation of measurement uncertainty
- Evaluation of uncertainty in the measurement of the concentration of ammonium in fresh water using the three approaches
- Conclusions

In this presentation, an example of a procedure is used to illustrate the three approaches. The example is the spectrophotometric measurement of the concentration of ammonium in drinking water *expressed as N* — in the presentation, data and calculations relating to the ammonium example are marked in yellow.

Note: 'Single laboratory validation & QC approach' is an abbreviation for 'Single laboratory validation and quality control data approach'.

TrainMiC
Training in Metrology in Chemistry

Literature for the evaluation of uncertainty

The application of the GUM general principles to measurements in chemistry is described in the guides:

Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement*, Second edition, 2000 (http://www.eurachem.org) [6]

Nordtest TR537, *Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories*, 2004 (http://www.nordicinnovation.net) [7]

EUROLAB Technical Report No 1/2007, *Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation* (http://www.eurolab.org) [8]

 Uncertainty_Approaches-2 Slide 4

These are three important documents detailing some of the various approaches to the evaluation of measurement uncertainty in chemical analyses; this presentation is based on these.

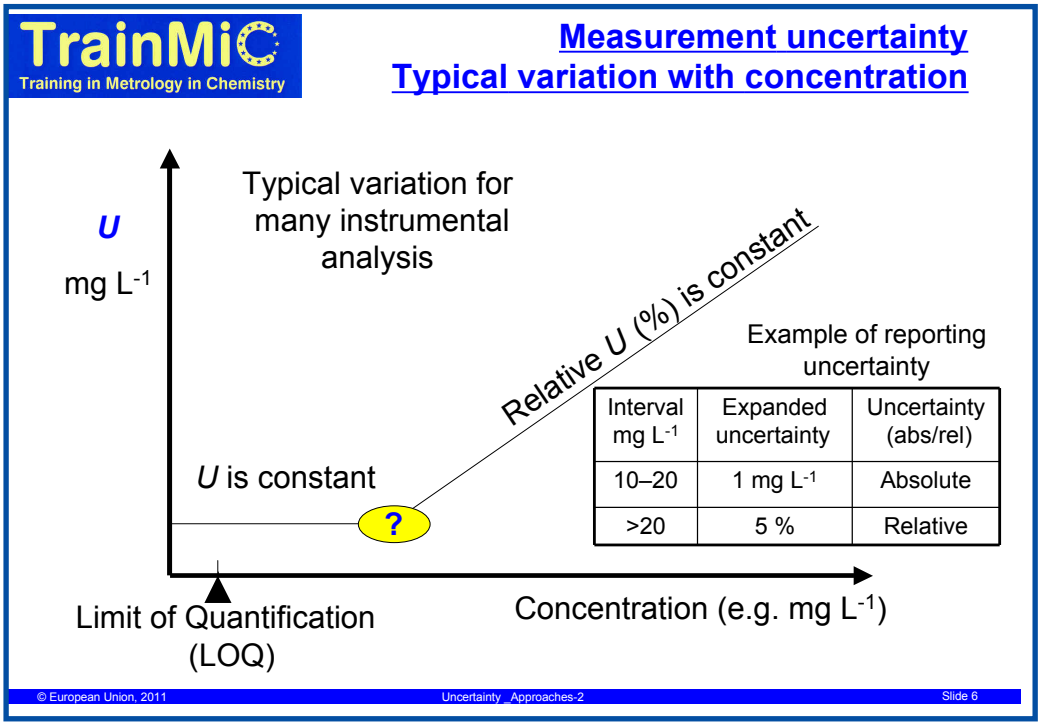
TrainMiC
Training in Metrology in Chemistry

Uncertainty as a function of concentration

- **At low concentrations (near the limit of quantification), use absolute uncertainties**
  - — Uncertainty of results is almost independent of analyte level
- **At higher concentrations, use relative or absolute uncertainties**
  - — Uncertainty of results is, for many instrumental analyses, roughly proportional to analyte concentrations at higher concentrations

Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement*, Second edition, 2000, Appendix E.4 [6]

The uncertainty often varies with concentration: further information can be found in the Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement*, in the appendix 'Useful statistical procedures' [6].



In many instrumental analyses (e.g. GC, ICP, AAS, XRF, UV), the variation shown in this graph is typical of the variation of uncertainty v concentration. In other techniques (e.g. titration, pH), uncertainty is less dependent on the concentration. This has to be taken into account when reporting uncertainty for results obtained according to a given procedure. A proposal is given on how to report uncertainty for the results obtained with an instrumental measurement procedure with a limit of quantification (LOQ) of 10 mg L^{-1}.

TrainMiC
Training in Metrology in Chemistry

Uncertainty by different approaches

- Modelling approach
  — Uncertainty of an **individual** result of a measurement **using a measurement procedure in the laboratory**
- Single laboratory validation & quality control approach
  — Typical uncertainty of results obtained using a **measurement procedure in the laboratory**
- Interlaboratory validation approach
  — Uncertainty of results obtained using the **same measurement procedure in different laboratories**

The uncertainties obtained may refer to different measuring conditions

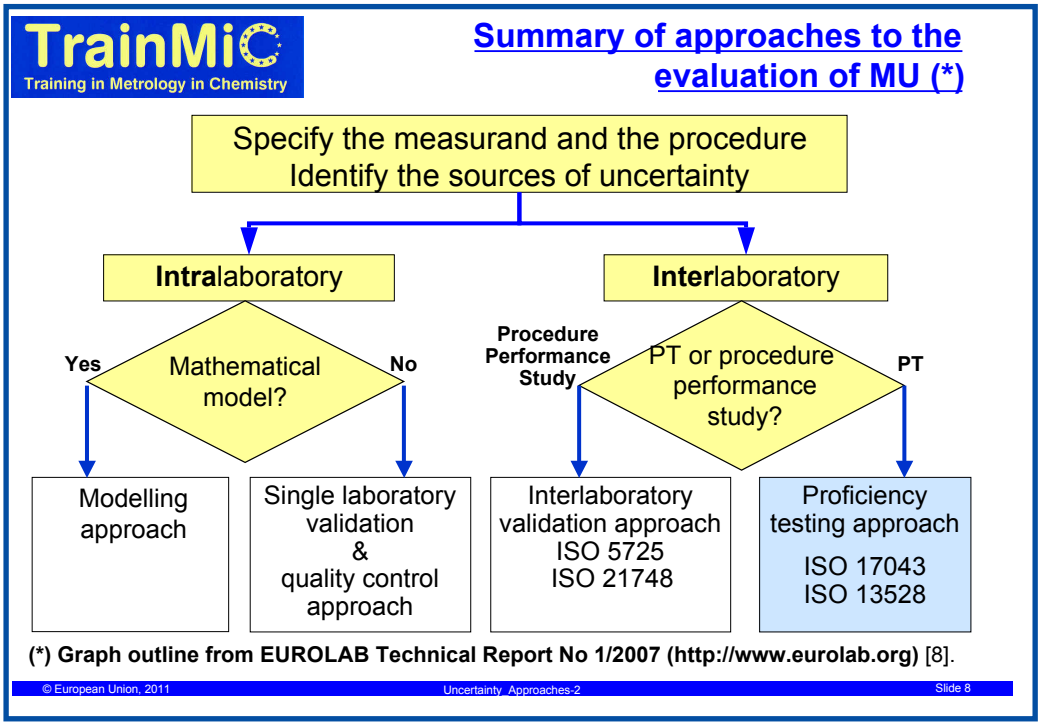
The main difference between the approaches lies in how the uncertainty components are grouped in order to quantify them. The uncertainty estimate obtained may refer to different situations.

In the modelling approach, the components are mostly quantified individually, whereas in the interlaboratory approach, all components are, in general, quantified as one estimate — the reproducibility standard deviation. In the single laboratory validation and quality control data approach, the components are grouped into a few major components.

The modelling approach mainly refers to a particular measurement result. Thus, it is possible to obtain the uncertainty estimate specifically referring to this particular measurement result under repeatability conditions.

The single laboratory validation and quality control data approach uses data gathered over a long period of time in your own laboratory. A preliminary evaluation of measurement uncertainty can be performed using the validation data (mainly including short-term precision and trueness data). The uncertainty can be re-evaluated later after routine use of the procedure, adding in quality control data. An uncertainty estimate is derived from results obtained using this procedure in your laboratory under within-laboratory reproducibility conditions (intermediate precision).

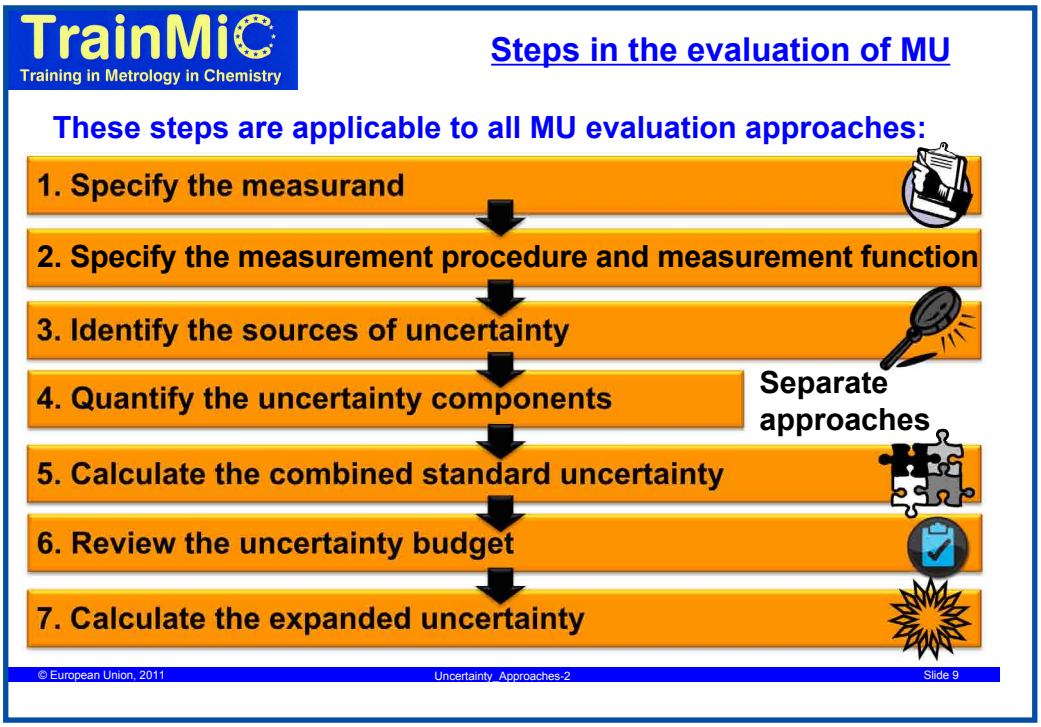
The interlaboratory validation approach uses data gathered from several laboratories on one occasion — resulting in an uncertainty estimate for the results obtained using this procedure by any competent laboratory is calculated. Measurement results must be obtained under reproducibility conditions.



In the EUROLAB report [8], the different approaches are presented graphically, a part of this is shown here. This report also includes the proficiency testing approach (PT), which is not generally recommended since, in most cases, laboratories that participate in a PT use different procedures.

If we choose to evaluate uncertainty in our own laboratory, we use an intralaboratory approach. If we use a standard method exactly according to the scope of the published data from an interlaboratory validation (performance study according to ISO 5725:1994) [13], we could choose to use an interlaboratory approach.

Note that all the approaches have the first steps in common.



The following are the steps involved in the evaluation of measurement uncertainty.

Most of the steps are the same for all approaches — Step 4 is different in the three approaches. Step 6 mainly refers to the modelling approach but, for all approaches, the obtained uncertainty can be compared with the target uncertainty and also with an uncertainty obtained in another laboratory.

TrainMiC
Training in Metrology in Chemistry

Example
Measurement of ammonium concentration

Procedure
EN ISO 11732:2005 — *Water quality — Determination of ammonium nitrogen — Method by flow analysis* (CFA and FIA) *and spectrometric detection* [15]

Scope
This International Standard specifies a suitable method for the measurement of the ammonium nitrogen concentration in various types of waters (such as fresh, ground, drinking, surface and waste waters) in the range 0.1 to 10 mg L^{-1} (undiluted sample)

Absolute or relative uncertainty?
At this low level of 0.2 mg L^{-1}, we will evaluate both relative and absolute uncertainty.

This is the procedure we will use for a comparison of the approaches. We will evaluate uncertainty estimates at low concentrations — 0.2 mgL^{-1}. Close to the limit of quantification (LOQ), we would normally evaluate an absolute uncertainty. In this case, we are not sure which to choose so we will evaluate both absolute and relative uncertainty estimates.

TrainMiC
Training in Metrology in Chemistry

All approaches Step 1 Specify the measurand

Measurand = quantity intended to be measured

Example:

Mass concentration (mg L^{-1}) of ammonium expressed as nitrogen in a laboratory sample of fresh water

(Denoted as C_{sample})

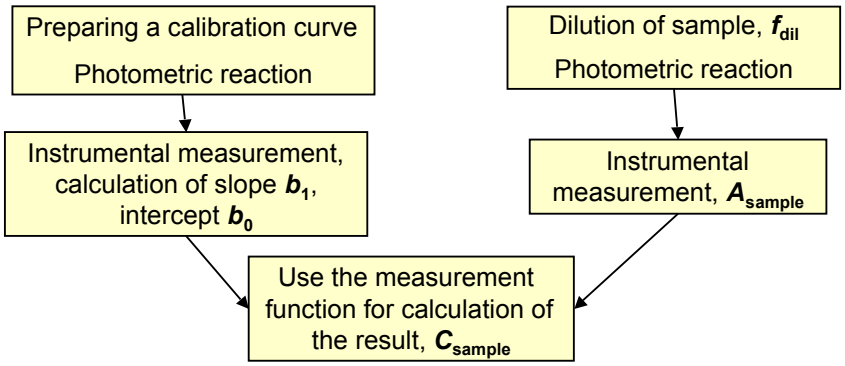
Target uncertainty — if available, state the target uncertainty (e.g. 15 %).

Step 1 Specify the measurand: with a target uncertainty of 15 %, the method is fit for its intended purpose.

TrainMiC
Training in Metrology in Chemistry

All approaches Step 2
Specify the measurement procedure

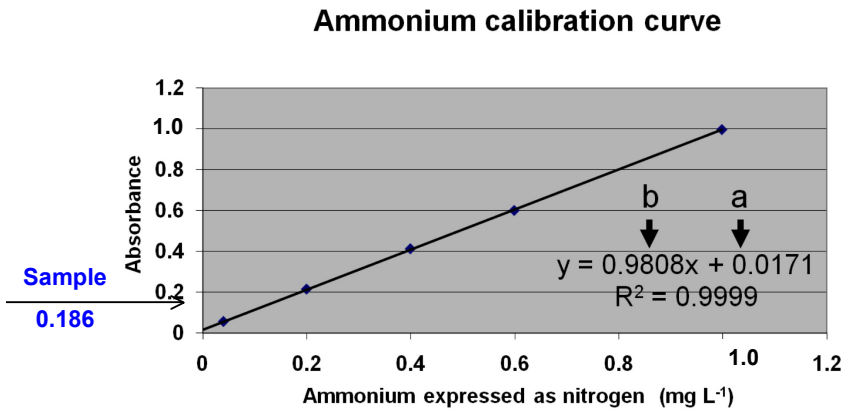
Procedure: EN ISO 11732:2005
— Water quality — Determination of ammonium nitrogen — Method by flow analysis (CFA and FIA) *and spectrometric detection* [15]



Here, the steps in the procedure for the measurement of the concentration of ammonium nitrogen in drinking water are presented.

TrainMiC
Training in Metrology in Chemistry

All approaches Step 2
Measurement model — the calibration curve



The slide shows the calibration curve — absorbance v concentration and also the results of linear regression.

In this example, the limit of quantification (LOQ) is 0.1 mg L^{-1}.

TrainMiC
Training in Metrology in Chemistry

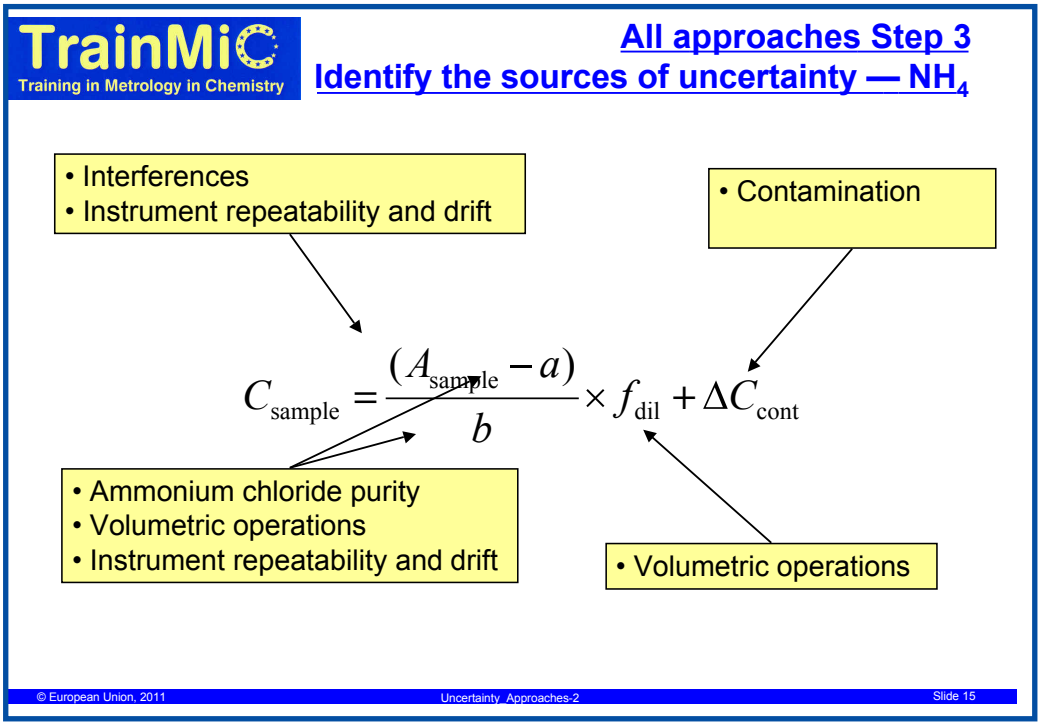
All approaches Step 2
Measurement model

- Measurement model

$$C_{\text{sample}} = \frac{(A_{\text{sample}} - a)}{b} \times f_{\text{dil}} + \Delta C_{\text{cont}}$$

A_{sample} is the absorbance of the dye solution obtained from the sample
b and a are the slope and intercept of the calibration curve
f_{dil} is the dilution factor
ΔC_{cont} is the contribution due to possible contamination

In many cases, a measurement function has to be further developed and extended to take into account uncertainty components that were not taken into account in the initial measurement function. Corrections are assumed to be in the measurement function to take account of all recognised, significant systematic effects. The slide shows the measurement function used to calculate the result extended with a factor (ΔC) to take into account contamination. For the evaluation of measurement uncertainty, the measurement function has to be extended since the model does not take into account contamination — an important source of uncertainty, estimated by variation in the blank results. The uncertainty of ΔC is estimated from analysing a blank sample on different days and calculating the standard deviation in concentration units — this will be the standard uncertainty of ΔC.



The slide shows the different sources of uncertainty that are identified and attributed to the input quantities. Variables *a* and *b* are related to the calibration function:

Absorbance = *b* * concentration + *a*

TrainMiC
Training in Metrology in Chemistry

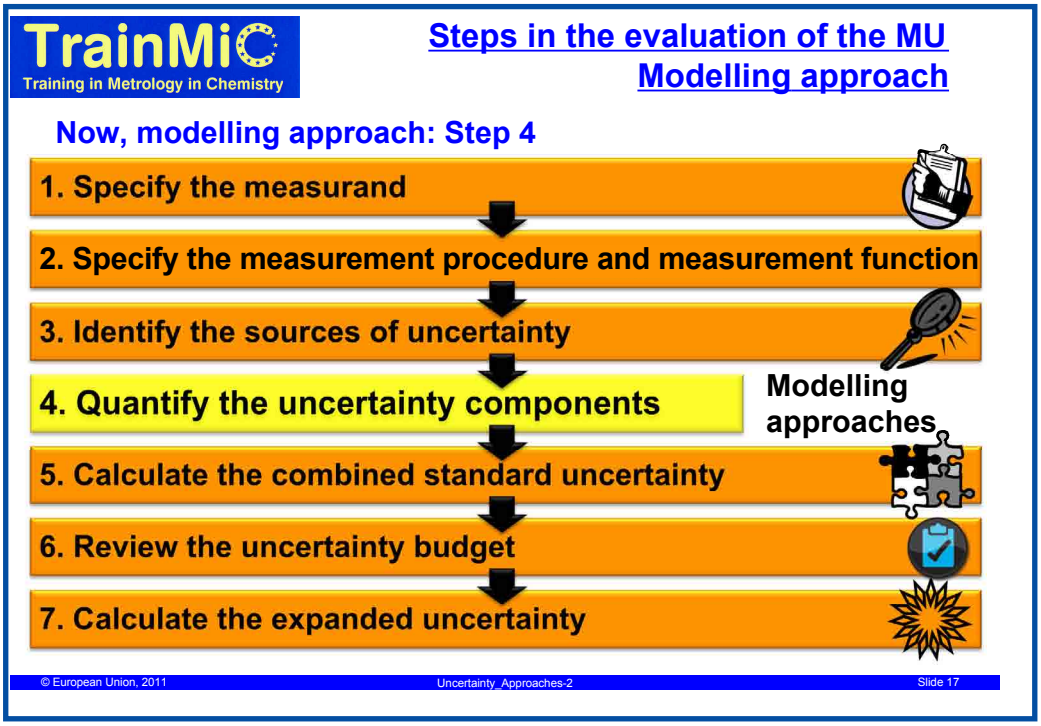
All approaches Step 3
Identify the sources of uncertainty

A general list of possible sources (*) — not necessarily independent:

- Sampling and subsampling
- Storage conditions
  - — Contamination/losses
- Instrument effects
  - — Memory effects
  - — Interferences
- Reagent purity
- Assumed stoichiometry
- Measurement conditions
- Sample effects
  - — Varying recovery
- Computational effects
  - — Selection of calibration model
- Blank correction including uncertainty of the blank
- Operator effects
- Random effects

(*) From Section 6, Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement*, Second edition, 2000 (http://www.eurachem.org) [6].

All approaches need to take into account all sources of uncertainty: the slide shows an exhaustive list of possible sources of uncertainty.



After the first three steps, in Step 4, the uncertainty components are quantified.

TrainMiC
Training in Metrology in Chemistry

Modelling approach Step 4
Quantify the uncertainty components — NH_4

Quantity	Value	*u*	Unit	
A_{sample}	0.1860	0.0033	AU	Sample absorption
a	0.0171	0.0022	AU	Intercept
b	0.9808	0.0120	AU L mg^{-1}	Slope
f_{dil}	1.2500	0.0053	Unitless	Dil factor
ΔC_{cont}	0.0000	0.0040	mg L^{-1}	Factor

To estimate *u*(*a*) and *u*(*b*), see Eurachem/CITAC Guide CG4 [6].

Step 4 with the modelling approach: the table shows the results of the calculation of the uncertainty of the individual components. The contamination issue (0.004 mg L^{-1}) is important in ammonium levels below 0.3 mg L^{-1}.

The slope and intercept are correlated but, in this case, the correlation between the slope and intercept was checked and had no significant influence.

Note: Absorbance is unitless but it is often reported in Absorbance Units (AU).

TrainMiC
Training in Metrology in Chemistry

Modelling approach Step 5
Calculate combined standard uncertainty

Several approaches can be followed to combine the estimated uncertainties:

- Law of propagation of uncertainty
- Numerical methods
  - Kragten method (*) (easily implemented in a spreadsheet) (*)
  - Dedicated software

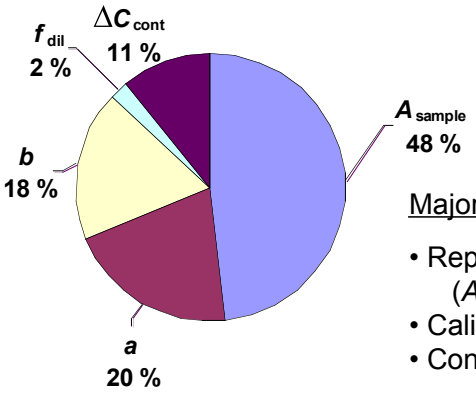
Using, in this case, the dedicated software to calculate combined standard uncertainty $u_c = 0.006$ mg L^{-1}

(*) *Analyst,* 1994, 119, pp. 2161–2166 [17] and Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement,* Second edition, 2000, Appendix E.2 [6].

The calculation of combined standard uncertainty can be carried out in several ways: more information about possible calculation methods is available in Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement* [6].

TrainMiC
Training in Metrology in Chemistry

Modelling approach Step 6
Review the uncertainty budget



Concentration level
0.2 mg L^{-1}

Major contributions :

- Repeatability and interferences (A_{sample})
- Calibration (*a* and *b*)
- Contamination (ΔC_{cont})

$$C_{sample} = \frac{(A_{sample} - a)}{b} \times f_{dil} + \Delta C_{cont}$$

The slide illustrates a summary of the uncertainty budget and shows the individual contribution of each uncertainty component as a percentage of the combined standard uncertainty in the concentration of ammonium at a concentration level of 0.2 mg L^{-1}. The ΔC contribution of 0.004 mg L^{-1} to the overall uncertainty will decrease with increasing concentration.

TrainMiC
Training in Metrology in Chemistry

Problems with modelling applied to chemical measurements

- Often not readily modelled
- Uncertainty contributions not readily quantified

Because:

- it is often difficult to separate the analyte from the matrix;
- interference from other components of the sample;
- sample inhomogeneity.

Failure to correctly account for all significant uncertainty sources leads to underestimation of uncertainty!

Let's move to next approach — single laboratory …

In order to obtain a reliable measurement uncertainty, all major contributing components need to be included. Using the modelling approach, you need to know the procedure in detail in order to be able to include all major components.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 — different scenarios

Scenario 1	Scenario 2
New procedure is introduced in the laboratory	Procedure already in place
Similar approach as scenario 2 but no QC data are available (the validation data is used)	Estimation of uncertainty or update of an existing uncertainty estimate using validation and quality control data

The following example is based on scenario 2 → Example shown in detail in Nordtest 537 (2004) (http://www.nordicinnovation.net) [7]

Here, we have two scenarios — a new procedure or a procedure already in place. In this example, we use data from a procedure which has been in use at the laboratory for several years — scenario 2.

A robust uncertainty estimation using this approach needs a large amount of data. A first evaluation of an uncertainty estimate can be made using validation data and then, subsequently, this value can be updated when the procedure has been in use in the laboratory for a longer time.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

In this approach, the sources of uncertainty are grouped into:

The whole range of variation encountered during the typical use of the measurement procedure:
range of expected values and sample types within the scope of the procedure during routine use. The **within-laboratory reproducibility (*) s_{Rw} can be obtained from** quality control data at different concentration levels (scenario 2).

The overall bias under within-laboratory reproducibility (*) conditions:
the use of certified reference materials (CRM) comparison with reference procedures, spiking and PT can be used to evaluate the component of uncertainty associated with **trueness** (i.e. *u*(bias)).

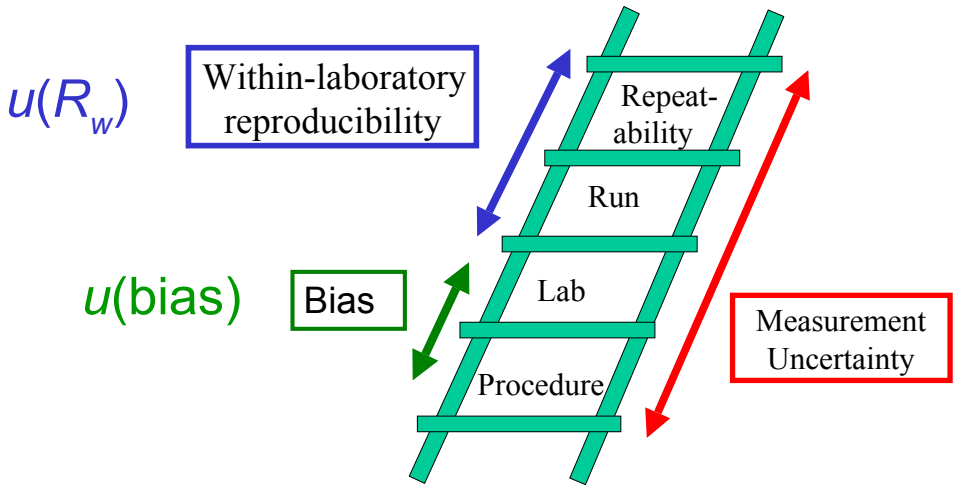
(*) Note: The VIM3 [1] term for within-laboratory reproducibility is 'intermediate precision'.

In this approach, the sources of uncertainty are grouped into two major components: precision and trueness. For both components, the laboratory must investigate the variation in the size of the components ensuring the scope of the procedure is fully covered (i.e. concentration range and different matrices). Anything that changes the results should be varied representatively.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

Ladder illustrating the uncertainty components



The two major components in the single laboratory validation & QC approach are the within-laboratory reproducibility and bias.

- R_w includes repeatability and between-days (runs). In the repeatability standard deviation, the sample inhomogeneity is included.
- Bias — both, laboratory and procedural bias are included.

Note: Throughout this part of the presentation, the original notation as in the Nordtest handbook [7] is used.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

- $u(R_w)$ is the uncertainty component that takes into account long-term variation of results — within-laboratory reproducibility (s_{Rw})
- Ideally, for one laboratory using one procedure:
  - — different days (longer time will give a more robust estimation)
  - — different technicians
  - — different reagent batches
  - — all instruments (several may be used within the laboratory)
  - — sample similar to test samples (matrix, concentration, homogeneity)
  - — …

Important:

Repeatability <	Within-laboratory reproducibility <	Combined uncertainty
s_r <	s_{Rw} <	u_c

For a reliable estimate of within-laboratory reproducibility, it is necessary to look at the performance of the procedure in routine use in the laboratory over a long time period. The information from the quality control data produced for the applied internal quality control supplements the validation data.

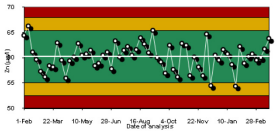
TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

If the QC sample covers the whole procedure

$u(R_w) = s_{Rw}$ at this level

- the warning limits (2*s*) of X-chart
  - — using a stable control sample covering the whole measurement procedure



Normally, at both low and high concentration but here, at low concentration

The control sample analysis has to cover the whole analytical process

Ammonium in fresh water
From the X-chart at 0.2 mg L^{-1}: control limits are set to ± 3.34 %

Thus $u(R_W)$ = 3.34 %/2 = 1.67 %

With quality control in place, the control limits will have been established. At this concentration level, the uncertainty component for the within-laboratory reproducibility is obtained from the X-chart by dividing the setting of the warning limit (±) by two.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

- **Potential bias can be estimated:**
  - from analysis of the several samples with a reference procedure
  - from analysis of ***several*** certified reference materials (CRM)
  - several PT rounds
  - from spiking experiments

- **Potential bias may have to be estimated separately for different sample types and different concentration levels**

Ideally, several reference measurements, several certified reference materials, several PTs, etc., because the bias will, in many cases, vary with matrix and concentration.

There are several ways to obtain an estimate of the bias within the scope of the procedure. A reliable estimate of the trueness of a laboratory measurement procedure can be obtained by analysing the test samples using a reference procedure and comparing the results. However, in most cases, this is not possible. If this is not possible, other ways of estimating bias within the scope of the procedure are proposed.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

$$u(bias) = \sqrt{RMS_{\text{bias}}{}^2 + \overline{u}\,(C_{\text{ref}})^2}$$

This component accounts for the uncertainty of the average bias of the laboratory results from the C_{ref}

This component accounts for the average uncertainty of the reference value C_{ref}

Uncertainty_Approaches-2
Slide 28

In this case, the bias is estimated from several different types of samples. The uncertainty component related to bias is then the combined uncertainty of reference values $u\left[\left(\overline{C}_{ref}\right)\right]$ and the root mean square of the different bias uncertainty estimates obtained. The RMS_{bias} formula is slightly different if only one certified reference material is used.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

Bias and uncertainty of reference value are expressed as a relative value

$$bias\ (\%) = \frac{(\bar{x}_i - x_{\text{ref}})}{x_{\text{ref}}} \times 100 \qquad u(C_{\text{ref}},\%) = \frac{U(C_{\text{ref}})}{2x_{\text{ref}}} \times 100$$

"Averaging" is done using the root mean square:

$$RMS_{\text{bias}} = \sqrt{\frac{\sum (bias_i)^2}{n_{\text{CRM}}}} \qquad \bar{u}(C_{\text{ref}}) = \sqrt{\frac{\sum u(C_{\text{ref}\,i})^2}{n_{\text{CRM}}}}$$

NOTE: *n* refers to the number of CRMs used (e.g. number of different bias estimates).

The calculations can be performed to provide an uncertainty estimate expressed as an absolute or relative uncertainty, depending on the variation of the uncertainty of the result with the concentration (see the slide 'Measurement uncertainty — Typical variation with concentration'). All components in the evaluation of measurement uncertainty should be expressed in the same way: absolute or relative and in the same unit.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

For ammonium, no CRMs are available: therefore, PT results are used.

PT Exercise	Nominal value x_{ref}	Laboratory result x_i	"Bias"	CV_R	Number of labs
Year	µg L^{-1}	µg L^{-1}	%	%	
1999 1	81	83	2.5	10	31
2	73	75	2.7	7	36
2000 1	264	269	1.9	8	32
2	210	213	1.4	10	35
2001 1	110	112	1.8	7	36
2	140	144	2.9	11	34
		Mean value	**+ 2.20**		34
		RMS	**2.26**	8.9	—

 Uncertainty, Approaches-2 Slide 30

In the ammonium example, no reference measurements or certified reference materials are available. Therefore, the results from participation in proficiency testing are used to evaluate uncertainty. The drawback is, of course, that the assigned values are not always traceable. However, in many cases, where there is considerable experience with the measurement procedure, the median of the results from a proficiency testing can be considered a good estimate of the true value.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 4 Quantify the uncertainty components

Ammonium in drinking water

Root mean square of bias

$$RMS_{\text{bias}} = \sqrt{\frac{2.5^2 + 2.7^2 + \cdots + 2.9^2}{n_{\text{PT}}}} = 26.2\ \%$$

Uncertainty of the assigned value — standard deviation of the mean

$$u(C_{\text{ref}})\,\% = \frac{\overline{CV}_{\text{R}}}{\sqrt{n_{\text{Lab}}}} = \frac{8.9}{\sqrt{34}} = 1.5\ \%$$

$$u(bias) = \sqrt{RMS_{\text{bias}}{}^2 + u(C_{\text{ref}})^2} = \sqrt{2.26^2 + 1.5^2} = 2.71\%$$

The uncertainty component u(bias) for the ammonium example is calculated from RMS_{bias} and the uncertainty of the assigned values. The $\overline{CV}_{\text{R}}$ is the mean or median value of the $\overline{CV}_{\text{R}}$ for the proficiency testing rounds. The different CV_{R} values should not be significantly different over the chosen concentration interval.

TrainMiC
Training in Metrology in Chemistry

Single laboratory validation & QC approach
Step 5 Calculate combined standard uncertainty

The main equation:

$$u_c = \sqrt{u(R_w)^2 + u(\text{bias})^2}$$

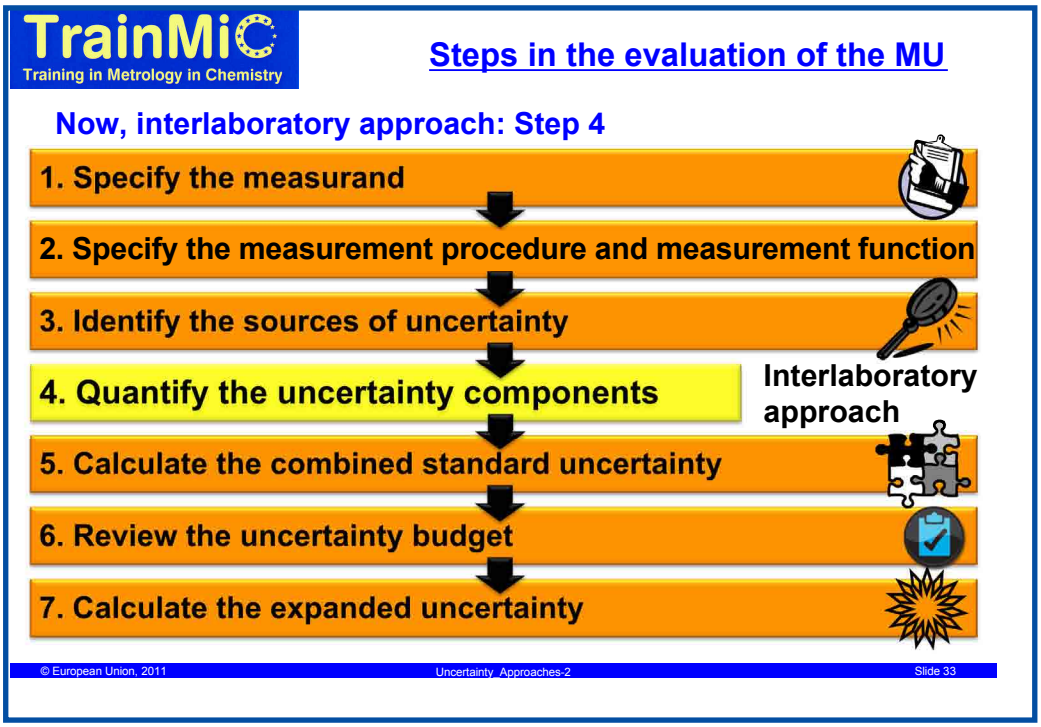
Within-laboratory reproducibility

Uncertainty of the estimate of the laboratory and the procedure bias

Ammonium in fresh water

$$u_c = \sqrt{1.67^2 + 2.71^2} = 3.18\ \%$$

The combined standard uncertainty of the concentration of ammonium in the sample is calculated by combining the two uncertainty estimates $u(R_w)$ and $u(\text{bias})$.



The first two approaches now have been presented — the modelling approach and the single laboratory validation and quality control data approach. The third approach is the interlaboratory approach.

TrainMiC
Training in Metrology in Chemistry

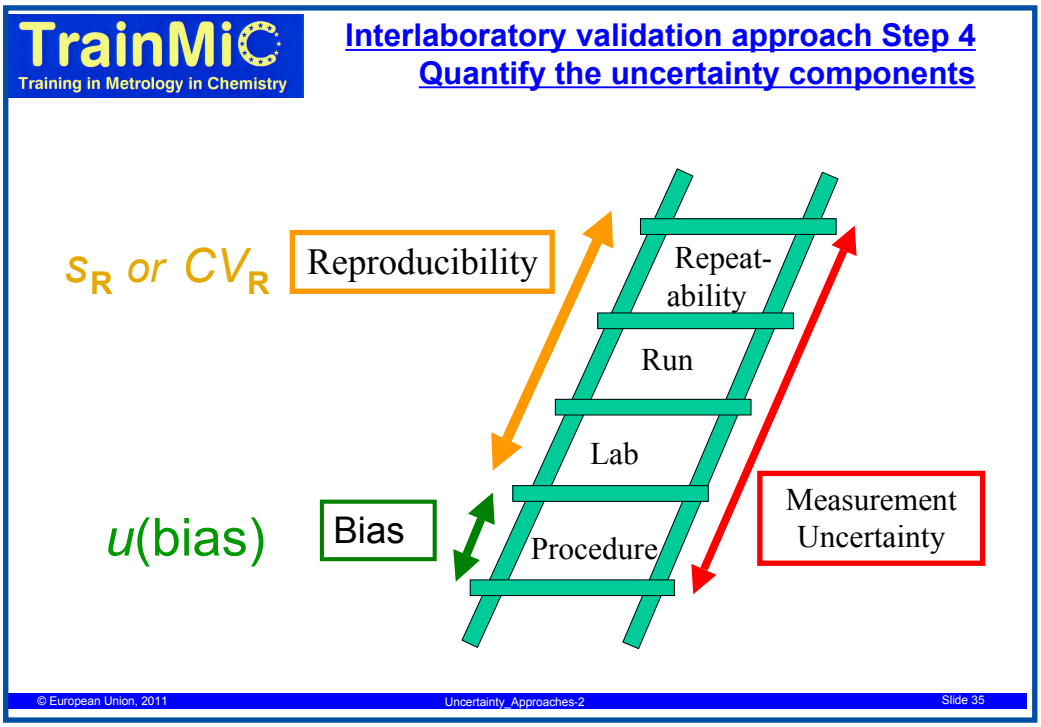
Interlaboratory validation approach Step 4 Quantify the uncertainty components

Consider results from a number of laboratories using similar test items and the same procedure as stated in ISO 5725:1994 — *Accuracy (trueness and precision) of measurement methods and results* [13]:

- the data are often found in reports of the interlaboratory validation or standardised procedures (e.g. ISO);
- procedure bias data may be available:
  - — procedure bias is usually small.

The laboratory has to investigate a possible laboratory bias.

The interlaboratory approach uses the s_R values obtained from interlaboratory validation performed according to ISO 5725:1994 [13]. The procedure has usually been refined before the interlaboratory validation so that the bias is insignificant.



In this approach, three major components (i) repeatability, (ii) day-to-day variation and (iii) laboratory bias are grouped into one.

TrainMiC
Training in Metrology in Chemistry

Interlaboratory validation approach Step 4
Quantify the uncertainty components

Table B.1 — Precision data for the determination of ammonium(N) with flow injection analysis (FIA) and spectrometric detection

Sample No	Matrix	l	n	o %	$\bar{x}$ mg/l	s_R mg/l	CV_R %	s_r mg/l	CV_r %
1	Drinking water [a]	15	56	6,67	0,284 1	0,027 9	9,81	0,011 4	4,01
2		15	54	10,0	0,901 9	0,038 2	4,24	0,017 6	1,95
3		15	55	8,33	0,531 8	0,026 1	4,91	0,016 5	3,10
4	Surface water [b]								

l is the number of laboratory sets (four values each per set);
n is the number of outlier-free individual analytical values;
o is the relative number of outliers;
$\bar{x}$ is the total mean;
s_R is the standard deviation of the reproducibility;
CV_R is the coefficient of variation of the reproducibility;
s_r is the standard deviation of the repeatability;
CV_r is the coefficient of variation of the repeatability.

EN ISO 11732:2005 — *Water quality — Determination of ammonium nitrogen* [15]

CV_R 9.81 % at 0.28 mg L^{-1}

Interlaboratory studies according to ISO 5725:1994 [13] typically provide the repeatability standard deviation s_r and reproducibility standard deviation s_R and may also provide an estimate of trueness (measured as bias with respect to a known reference value). The application of these data to the evaluation of measurement uncertainty is discussed in detail in ISO 21748:2010 [14]. In this slide, the results from an interlaboratory study as reported in EN ISO 11732:2005 [15] are shown.

The sample in the previous examples had a concentration of approximately 0.2 mg L^{-1}-1 — in this case, the nearest is 0.28 mg L^{-1}.

TrainMiC
Training in Metrology in Chemistry

Interlaboratory validation Step 5
Calculate combined standard uncertainty

If the interlaboratory study has been performed according to ISO 5725:1994 [13], then according to ISO 21748:2010 (*)

$$u_c = CV_R$$

Ammonium in fresh water

at 0.215 mg L^{-1}

u_c = 10 % or 0.021 mg L^{-1}

(*) ISO 21748:2010 — *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation* [14].

It is recommended that additional uncertainties associated with factors not adequately covered by the interlaboratory comparison (ILC) are identified and evaluated, particularly: (i) sampling (ILC rarely include a sampling step); (ii) sample pretreatment (e.g. ILC test samples are homogenised prior to circulation); (iii) variation in conditions (variation between conditions when ILC samples are measured and conditions used when test samples are measured); (iv) changes in sample type (in cases where the properties of the ILC sample differs from those of test samples, this needs to be considered). In order to use the CV_R value obtained in the ILC, the laboratory also has to establish that they can achieve a comparable repeatability standard deviation s_r.

TrainMiC
Training in Metrology in Chemistry

PT approach

In PT studies, laboratories may use different measurement procedures.

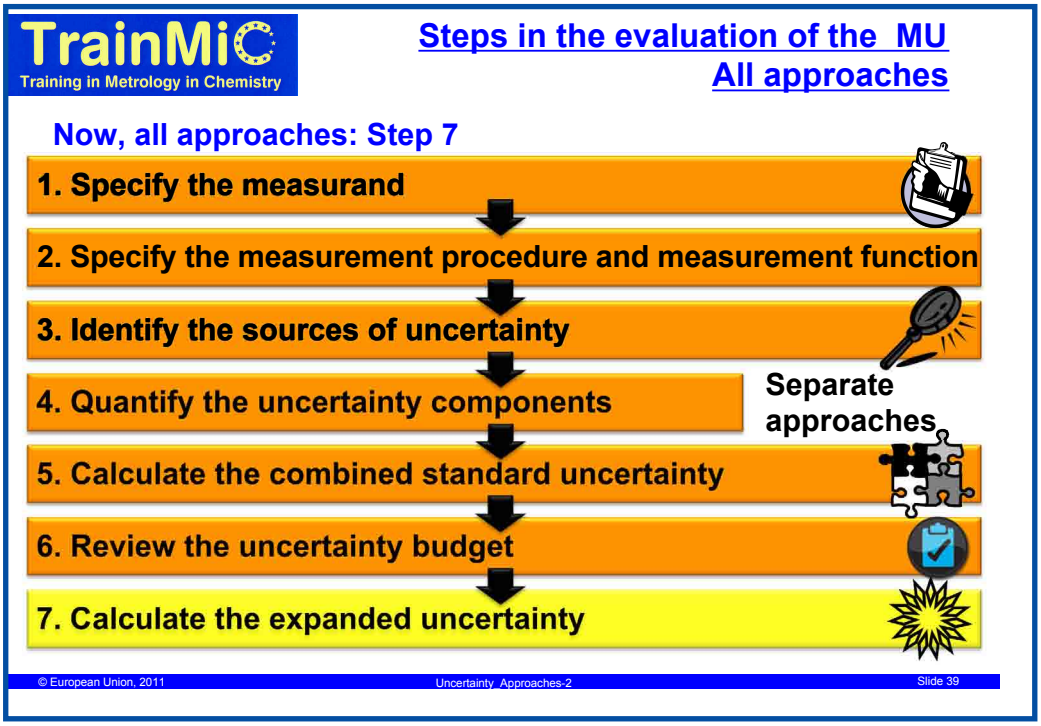
- The standard deviation of the participating laboratories, s_{PT}, can be used as a **very crude initial estimate** of uncertainty (works only if the laboratories use the same procedure).

The use of the PT approach is not generally advisable.

- In this example, CV_{PT} are generally 9–11 % in this concentration range (i.e. similar to the CV_R given in the interlaboratory comparison).

$$u_c = CV_{PT} = 10\ \%$$

This approach is also presented in the EUROLAB report [8], but is not generally recommended since, in most cases, laboratories use different procedures when analysing proficiency testing samples.



The three different approaches have now been presented for Steps 1 to 6.

TrainMiC
Training in Metrology in Chemistry

All approaches Step 6
Calculate expanded uncertainty

The expanded uncertainty U is obtained by multiplying the combined standard uncertainty $u_c(y)$ by a coverage factor k

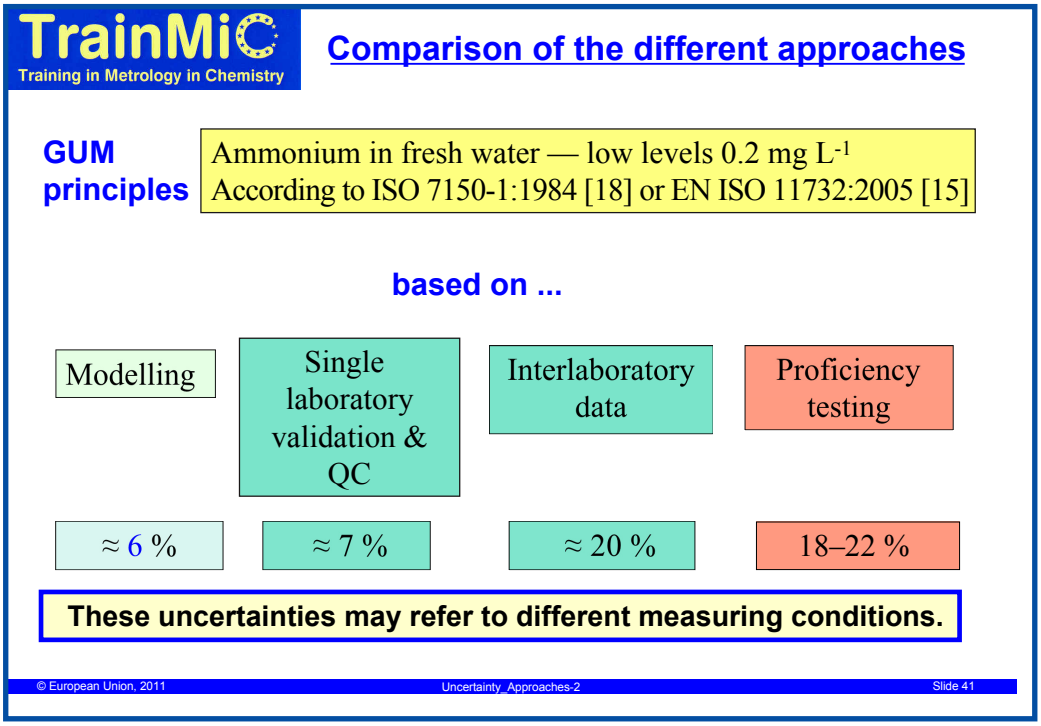
$$U = k \times u_c$$

Ammonium expressed as nitrogen: $C_{sample} = (0.215 \pm U)$ mg L^{-1}

Approach	k	U mg L^{-1}	U relative %
Modelling	2	0.012	6
Single laboratory	2	0.014	7
Interlaboratory	2	0.043	20

the interval ($y - U$, $y + U$) is the range that may be expected to encompass approximately 95 % (when k = 2) of the distribution of values that could reasonably be attributed to the measurand.

In Step 7, the expanded uncertainty is calculated. Here, a comparison with the target uncertainty, if available, is recommended. From the EU Drinking Water Directive (Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption) we estimated a target uncertainty of 15 % at a level of 0.5 mg L^{-1}.



These uncertainties refer to different measuring conditions (slide 8).

The different conditions are: (i) repeatability (modelling); (ii) within-laboratory reproducibility (single laboratory validation & quality control data; and (iii) reproducibility (interlaboratory and proficiency testing). In the example presented, all the proficiency testing participants used the same EN ISO 11732:2005 procedure [15].

TrainMiC
Training in Metrology in Chemistry

Conclusions

- If you have
  - — Competence and time
  - — Data on all important influencing quantities
    - Use the Modelling approach.
- If you have
  - — Validation data
  - — Quality control data and results from bias estimates (reference procedure, CRM, PT, spiking)
    - Use the single-laboratory validation approach.
- If you are using a highly standardised procedure within its scope
    - Use the Interlaboratory validation approach.

Depending on the data available, different approaches for the evaluation of measurement uncertainty can be chosen. In order to use the interlaboratory approach, the laboratory must demonstrate its competence is equivalent to those involved in the interlaboratory validation.

If target uncertainty is available, the fitness for purpose regarding uncertainty can be assessed.

TrainMiC
Training in Metrology in Chemistry

Final message

One can choose different approaches for the evaluation of uncertainty depending on the purpose and available data.

NOTE
The evaluated uncertainty may refer to different measuring conditions.

Uncertainty_Approaches-2
Slide 43

Depending on the purpose and the available data, different approaches for the evaluation of measurement uncertainty evaluation can be selected.

If detailed knowledge of the different uncertainty components is needed, the modelling approach should be the first choice.

When data are available in the laboratory (validation, quality control), a single laboratory validation and quality control data is a possible approach.

In most cases, a laboratory using a standard method within its scope should use the interlaboratory approach.

The uncertainties obtained refer to different measuring conditions: (i) repeatability conditions (i.e. uncertainty for one result obtained in a laboratory); (ii) intermediate precision conditions (within-laboratory reproducibility), a typical uncertainty for results using the procedure under routine conditions in one laboratory; (iii) reproducibility conditions, a typical uncertainty for results from any competent laboratory using this procedure.

Chapter 3

Statistics for analytical chemistry — Part I

The aim of this presentation is to focus on some statistical tools that are required for the evaluation of uncertainty and the interpretation of interlaboratory comparisons (ILC). The following topics are presented: average, standard deviation, population distribution (normal, rectangular and triangular), law of propagation of uncertainty, type of uncertainties (A and B) and scoring of ILC. The proper understanding of these issues is essential to achieve a correct evaluation of the 'combined uncertainty' compliant with GUM. Several examples are discussed in detail.

TrainMiC
Training in Metrology in Chemistry

Statistics for analytical chemistry

Part I

Last updated - January 2011

The aim of this presentation is to explain that statistics is a useful tool for data treatment and provides means of reaching objective decisions.

TrainMiC
Training in Metrology in Chemistry

Content I

Part 1:

- Statistics of repeated measurements
- Statistics for the estimation of measurement uncertainty
- Significance tests
- Reporting of measurement results

Statistics is a very broad field. The essential statistics required for quality control, measurement uncertainty and validation of analytical methods are presented in two presentations dedicated to the use of statistics in analytical chemistry. The topics discussed in the first presentation are shown in this slide.

TrainMiC
Training in Metrology in Chemistry

Content II

Part 2:

- Regression and correlation
- α and β errors
- Limit of detection
- Control charts
- Analysis of variance (ANOVA)

The statistical terms covered in the second presentation are listed on this slide.

TrainMiC
Training in Metrology in Chemistry

Overview

Statistics of repeated measurements

- Normal distribution
- Calculation of the most common statistical parameters

Statistics for the estimation of measurement uncertainty

Significance testing

- Is a result statistically significantly different?

Reporting of measurement results

- Significant figures
- Rounding results

This presentation includes:

- the most important concepts and terms used;
- calculations of the most common statistical parameters;
- basic statistics for the evaluation of uncertainty;
- significance testing;
- reporting analytical results.

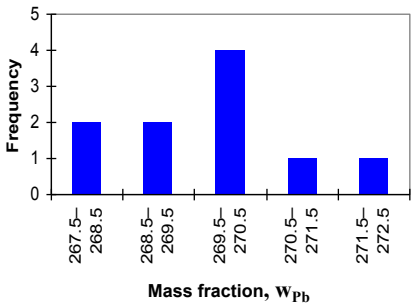
TrainMiC
Training in Metrology in Chemistry

Statistics of repeated measurements

TrainMiC
Training in Metrology in Chemistry

Frequency distribution

Mass fraction of lead in wine ($ng\ g^{-1}$)	
271.4	268.4
267.8	269.6
268.7	272.5
269.5	270.1
269.6	268.6



The table shows a typical set of analytical data — a series of repeated measurements of the lead content in a wine sample, obtained in one analytical laboratory. As expected, random variation results in a set of slightly different measured values.

The chart shows a histogram of the obtained results. Each bar represents the number of measurements falling in a given range (i.e. the frequency of occurrence). Hence, the chart shows the distribution of the obtained results.

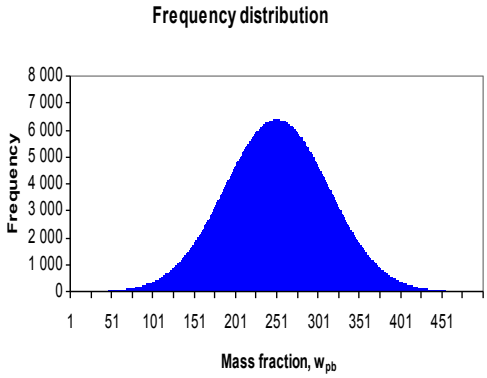
If more data had been available, the histogram would have conveyed a more definite impression of the distribution and would be more symmetric.

Two points to stress here are:

- data are concentrated in the central region of the histogram;
- the distribution is roughly symmetrical.

TrainMiC
Training in Metrology in Chemistry

Normal distribution



Several sources contribute to the measurement results.

In practice, we can assume that all measurements are normally distributed.

Statistics 4.0
Slide 7

Finally, with a very large amount of data and a large number of ranges, the shape of the underlying population becomes clear. One can think now of the population distribution as being described not by a histogram but by a smooth curve, the function of which we could, in principle, determine.

The Normal Distribution

As the name implies, the normal distribution describes the way results are commonly distributed. The very large majority of measurements subject to several different effects (environment, reagent variation, instrument 'noise', etc.) will, repeated frequently, fall into a normal distribution, with most results clustered around a central value and a decreasing number at greater distance. The distribution has potentially an infinite range — values may turn up at great distances from the centre of the distribution.

TrainMiC
Training in Metrology in Chemistry

Normal distribution

The **normal distribution**, also called **Gaussian distribution,** is a **continuous probability distribution** and can be closely approximated by a curve called the 'normal distribution curve'

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The function can be described by μ (arithmetic mean of a population) and σ (standard deviation of a population).

σ characterises the dispersion of the values.

 Statistics 4.0 Slide 8

Normal distribution occupies a special place among all statistical distributions.

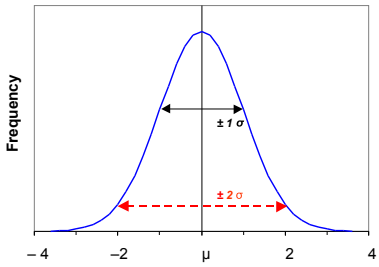
All quantitative parameters derived from measurements have probability distribution functions (i.e. they are not known exactly). Usually, the analyst would like to obtain the true value from the measurement, but it is never possible. If the measurements are repeated sufficiently, the expectation is that the mean value will be close to the true value, with the actual results spread around it.

A normal distribution implies that if a large number of measurements of the same system is made, the values will be distributed around the mean value, and the frequency of a result will become lower the further away the result is from the mean. A normal distribution is a probability curve where there is a high probability of an event occurring near the mean value, with a decreasing chance of an event occurring as one moves away from the mean.

The curve of the normal distribution is bell shaped and is completely determined by only two parameters: the central value μ and the standard deviation σ.

TrainMiC
Training in Metrology in Chemistry

Sample and population



Population: all possible data are available, so μ and σ can be calculated

Sample: only a subset of the population is known, so *x* and *s* can be determined as estimates of μ and σ

In general, we do not have access to the entire population of measurement data. When we are asked to measure the concentration of an analyte, we usually make a limited number of measurements on test portions and use the results as our best estimate of the true analyte concentration.

The limited number of measurements on test portions represent a sample of the total population of results.

If the whole population is known, the true mean value μ and standard deviation σ can be calculated. If only a sample is known, these parameters have to be estimated.

TrainMiC
Training in Metrology in Chemistry

Distribution of repeated measurements

For a set of n values x_i

Standard deviation

$$s(x_i) = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Mean value

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} (x_i)$$

Standard deviation of the mean

$$s(\bar{x}_i) = \frac{s(x_i)}{\sqrt{n}}$$

Variance

$$V(x_i) = s^2(x_i)$$

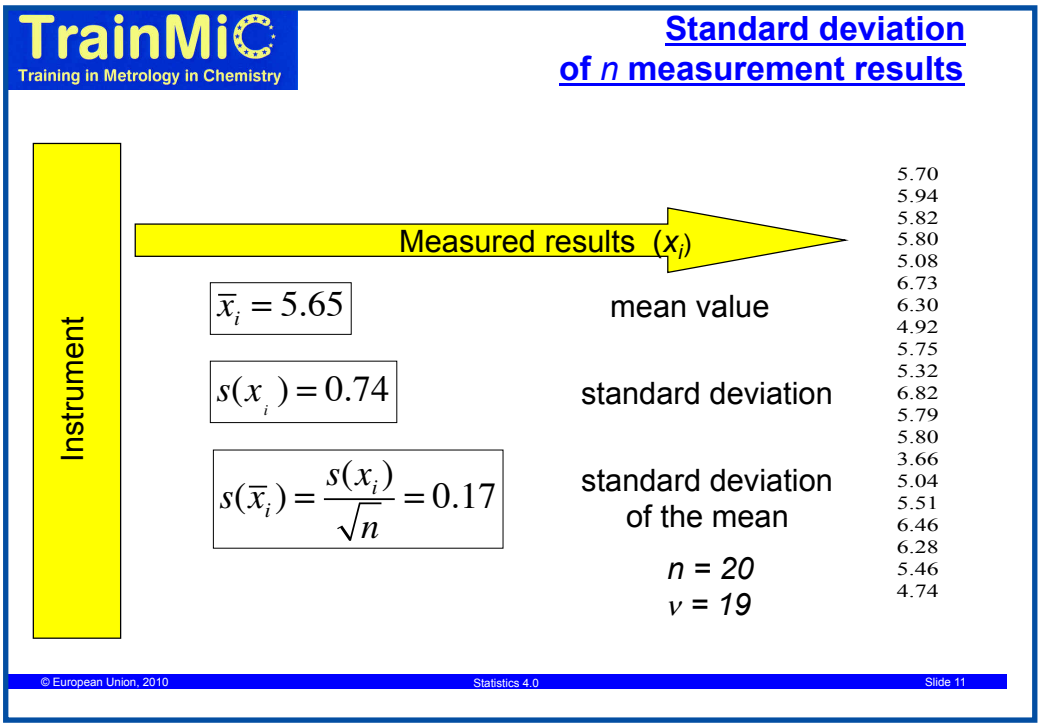
Relative standard deviation or coefficient of variation

$$RSD = \frac{s(x_i)}{\bar{x}} \text{ or } RSD\ (\%) = CV\ \% = \frac{s(x_i)}{\bar{x}} \times 100$$

The distinction between sample and population is important because it affects how some statistical parameters are calculated. In this presentation, only statistical parameters related to samples from a population are discussed.

For a set of n values x_i, the following statistical parameters can be defined.

- The mean value (arithmetic average) of all measurement results. If the sample is randomly taken then the average is the best estimate of the population mean.
- Standard deviation is the positive square root of the variance.
- Standard deviation of the mean is an estimate of the standard deviation of the mean values that would arise if repeated samples were taken from the population. Standard deviation of the mean is smaller than the standard deviation of a sample.
- Variance: the variance measures the extent to which the results differ from each other; the larger the variance, the greater the spread of data is.
- The relative standard deviation is a measure of the spread of data in comparison to the mean of the data.



This slide shows the statistical parameters for a set of 20 randomly distributed repeated measurements.

Standard deviation of the individual result is given by the following equation:

$$s(x_i) = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}$$

$n - 1$ is the degrees of freedom of the standard deviation (represented by ν).

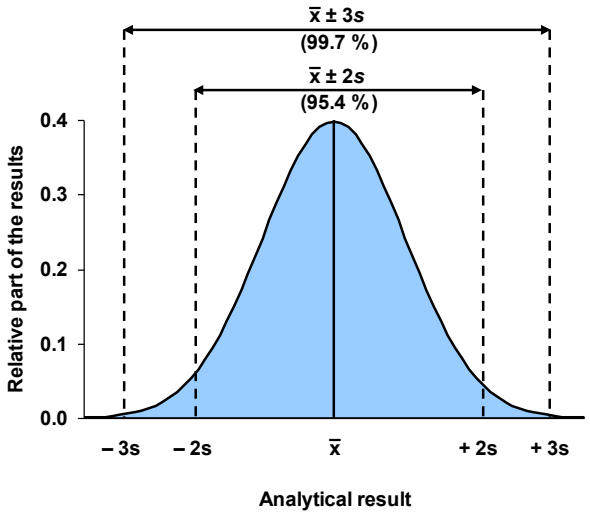
$s(\overline{x})$ is the standard deviation of the mean of *n* repeated measurements, given by the equation:

$$s(\overline{x}) = \frac{s(x_i)}{\sqrt{n}}$$

TrainMiC
Training in Metrology in Chemistry

Confidence interval

$\bar{x} \pm 1s$	68.0 %
$\bar{x} \pm 2s$	95.4 %
$\bar{x} \pm 3s$	99.7 %



The normal distribution is characterised by the parameter μ, which describes the centre, or location of the distribution. However, μ is not sufficient to completely characterise the distribution, since several different distributions could be located at the same point. Therefore, a second parameter σ to measure the spread, or dispersion, of the distribution is needed. When dealing with a sample of the population, we only have estimates of μ and σ (i.e. $\bar{x}$ and s respectively).

For a normal distribution with sample mean $\bar{x}$ and standard deviation s, approximately 68.3 % of the population values lie within ± s of the mean, approximately 95.4 % of the population values lie within ± 2s of the mean value and approximately 99.7 % of the population values lie within ± 3s of the mean.

TrainMiC
Training in Metrology in Chemistry

Statistics for the estimation of measurement uncertainty

 Statistics 4.0 Slide 13

The next few slides show the statistics required for the estimation of measurement uncertainty.

TrainMiC
Training in Metrology in Chemistry

Combined and expanded uncertainty according to GUM

When there is no correlation between input quantities, the combined standard uncertainty is evaluated as the square root of the combined variance according to the law of propagation of uncertainty:

$$u_c^2(y) = \sum \left(\frac{\delta f}{\delta x_i} \right)^2 \times \left(u(x_i) \right)^2$$

Expanded uncertainty, *U*, is obtained by multiplying the combined standard uncertainty by a coverage factor *k*:

$$U(y) = k \times u_c(y)$$

(often k = 2)

When there is no correlation between input quantities, the combined standard uncertainty is evaluated as the square root of the combined variance according to the law of propagation of uncertainty. All standard uncertainties can be combined with the use of the law of propagation of uncertainty.

In order to cover a larger fraction of likely values than those covered in the range of one standard uncertainty, the expanded uncertainty is used. Expanded uncertainty, U, is obtained by multiplying the combined standard uncertainty by a coverage factor, k. In the majority of analytical applications, a factor $k = 2$ is used, meaning that expanded uncertainty covers approximately 95.4 % of the likely values. To convert an expanded uncertainty to a standard uncertainty, the expanded uncertainty value is divided by k.

TrainMiC
Training in Metrology in Chemistry

Law of propagation of uncertainty *without correlation*

$$Y = f(x_1, x_2, \ldots, x_n)$$

$$u_c^2(Y) = \sum \left(\frac{\partial f}{\partial x_i} \right)^2 \times \left(u(x_i) \right)^2$$

$$u_c(Y) = \sqrt{\left[\frac{\partial Y}{\partial x_1} u(x_1) \right]^2 + \left[\frac{\partial Y}{\partial x_2} u(x_2) \right]^2 + \cdots + \left[\frac{\partial Y}{\partial x_n} u(x_n) \right]^2}$$

In many cases, a measurand Y is not measured directly, but is determined from n other quantities $X_1, X_2, \ldots X_n$ through a functional relation f.

Among the quantities X_n, there are usually also a number of corrections as well as quantities that take into account other sources of variability, such as different observers, instruments, samples, laboratories, and times at which observations are made (e.g. different days). Thus, in the equation, the function f should express not simply a physical law but a measurement process and, in particular, it should contain all quantities that can significantly contribute to uncertainty of a measurement result. The rules of combination depend on the form of the differential in the measurement function.

TrainMiC
Training in Metrology in Chemistry

Law of propagation of uncertainty examples

$$Y = (x_1 + x_2) \qquad Y = (x_1 - x_2) \Rightarrow u(y) = \sqrt{u(x_1)^2 + u(x_2)^2}$$

$$Y = (x_1 \times x_2) \qquad Y = (x_1 / x_2) \Rightarrow \frac{u(y)}{Y} = \sqrt{\left(\frac{u(x_1)}{x_1}\right)^2 + \left(\frac{u(x_2}{x_2}\right)^2}$$

$$Y = (x_1 \times x_2) \pm x_3 \qquad Y = (x_1 / x_2) \pm x_3 \Rightarrow u(y) = \sqrt{\left(x_1 \times x_2 \times \sqrt{\left(\frac{u(x_1)}{x_1}\right)^2 + \left(\frac{u(x_2)}{x_2}\right)^2}\right)^2 + u(x_3)^2}$$

 Statistics 4.0 Slide 16

Usually two main types of functional relationships are used for the measurement model (measurement function):

- addition/subtraction — combined uncertainty is obtained as a square root of the sum of squared absolute standard uncertainties (root sum of squares);
- multiplication/division — combined uncertainty is obtained as a square root of the sum of squared relative standard uncertainties.

TrainMiC
Training in Metrology in Chemistry

Different ways of estimating uncertainty

Type A evaluation of uncertainty:

statistical analysis of a series of measurements.

Type A evaluation of uncertainty is based on experiments and is quantified in terms of *the standard deviation* of the measured values

Standard uncertainty = *Standard deviation*

Type B evaluation of uncertainty: (GUM [5])

by *other means* than statistical analysis

(*previous experiments, literature data, manufacturer's information, expert's estimate*)

The uncertainty should be quantified in a way that is common to all types of measurements in chemistry, since it should be possible to compare different results.

The measurement uncertainty can be determined using statistical or non-statistical methods. Therefore, the uncertainty estimate can be one of two categories:

- Type A — obtained by statistical analysis of the data from repeated measurements.
- Type B — obtained from those sources where the value cannot be defined by repeated measurements (other means than statistical analysis of results).

Standard uncertainty from Type A and Type B evaluations are treated in the same way.

TrainMiC
Training in Metrology in Chemistry

Expression of data

Before combining different uncertainty contributions, all uncertainty contributions must be expressed/converted to standard uncertainty

when available as:

— standard deviation:	use as is
— stated range and distribution:	convert
— confidence intervals:	convert
— expanded uncertainties:	convert

 Statistics 4.0 Slide 18

In order to carry out uncertainty estimations using the ISO-GUM [5] approach, all the uncertainty contributions need to be converted to 'standard uncertainty' format. This means they all have to be expressed as standard deviations.

If the random variation is evaluated from replicate measurements, the result will be presented as a standard deviation (Type A evaluation).

Other uncertainties (e.g. specifications for glassware) may be originally expressed as a range; in other instances, a confidence interval is provided at a given confidence level.

TrainMiC
Training in Metrology in Chemistry

Type B Rectangular distribution

The value is between the limits:

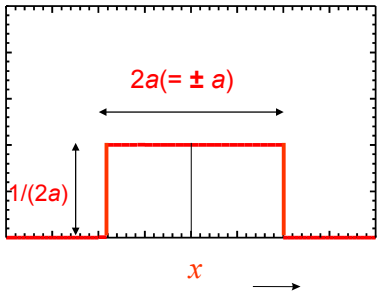
$$a_- \ldots a_+$$

The expectation:

$$y = x \pm a$$

Estimated standard deviation:

$$s = u(x) = a / \sqrt{3}$$



One can only assume that it is equally probable for the value to lie anywhere within the interval.

Very often we have uncertainty data presented in the form of '± a' and the information about the distribution is not given. In such a case, it is very appropriate and safe to assume a rectangular distribution. The rectangular distribution describes the situation when the values could, with equal probability, be anywhere in the given range. Rectangular distributions are usually described in terms of the average value and the range (2a, in the figure above). A standard deviation can be calculated for this distribution as indicated on the slide.

It is important to note that the area of the rectangle equals 1 = 2*a*1/(2a).

TrainMiC
Training in Metrology in Chemistry

Example of rectangular distribution

Certificates or other specification give limits where the value could be, without specifying a level of confidence.

'It is likely that the value is somewhere in that range.'

Rectangular distribution is usually described in terms of the average value and the range (± a).

Example:
The concentration of a calibration standard is quoted as (1 000 ± 2) mg L^{-1}. Assuming rectangular distribution, the standard uncertainty is:

$$u(x) = a / \sqrt{3} = 2 / \sqrt{3} = 1.16 \text{ mg } L^{-1}$$

Rectangular distribution is usually described in terms of the mean value and the range (± a). Certificates or other specifications usually give limits where the value could be, without specifying a level of confidence. A truly rectangular distributed uncertainty is the uncertainty due to rounding.

TrainMiC
Training in Metrology in Chemistry

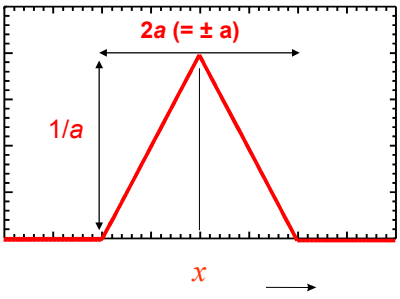
Type B Triangular distribution

Distribution used when it is suggested that values near the centre of range are more likely than near to the extremes

$$y = x \pm a$$

Estimated standard deviation:

$$s = u(x) = a / \sqrt{6}$$



The triangular distribution describes the distribution of values when it is expected that values near the centre of the range are more likely than those near to the extremes. Triangular distributions are usually described in terms of the average value and the range (2a, in the figure above). A standard deviation can be calculated for this distribution. In the case of triangular distribution, it is reasonable to expect that the value is in the centre of a given range rather than near to the extremes. It is important to note that the triangular area is equal to 1.

TrainMiC
Training in Metrology in Chemistry

Example of triangular distribution

Values close to the nominal value are more likely than those near the boundaries.

Example (*volumetric glassware*)

The manufacture quotes a volume for the flask of (100 ± 0.1) mL at T = 20 °C.

Nominal value most probable!
Assuming triangular distribution the standard uncertainty is:

$$u(x) = a/\sqrt{6} = 0.1/\sqrt{6} = 0.04 \text{ mL}$$

If in doubt, use the rectangular distribution.

In the example, the manufacture quotes a volume for the flask of (100 ± 0.1) mL at T = 20° C.

Assuming triangular distribution, the standard uncertainty is 0.04; assuming rectangular distribution $u(x)$ = 0.06 mL.

TrainMiC
Training in Metrology in Chemistry

Type B Confidence interval

The individual measurement results are distributed around the mean value.

The estimate of the mean value (μ) lies within the **C**onfidence **I**nterval (*CI*), with a probability of ($1 - \alpha$), *having 'n – 1'* degrees of freedom:
(*where n = number of replicates*)

$$95\,\%\ \ CI = t_{(0.05,n\text{-}1)} \times s / \sqrt{n}$$

To convert a *CI* to a standard uncertainty, divide by $t_{(0.05,\, n-1)}$

Knowing the sampling distribution of the mean value, one can see if a range assumed to include the true value can be defined (excluding any systematic effects). Such a range is the confidence interval (*CI*). This slide shows that this interval depends on the number of replicates used to estimate the standard deviation and the level of confidence required. The t value in the formula depends both on the level of confidence required and the degrees of freedom ($n - 1$) and can be found in tables for the t distribution. Information about the uncertainty of a value may be given as a confidence interval.

A point to note here is that uncertainty and confidence interval should not be confused. The confidence interval may not reflect the true variability.

If the data are given as 'A concentration is given as a confidence interval', then this should be converted to a standard uncertainty using the formula presented in the slide.

TrainMiC
Training in Metrology in Chemistry

t-distribution (Student's t-distribution)

This is the probability distribution used when the population is normally distributed, but the sample size is small.

When the sample mean is $\bar{x}$ and $s(x_i)$ is the sample standard deviation, then the quantity

$$t = \frac{\bar{x} - \mu}{s(x_i)/\sqrt{n}}$$

has a t-distribution with $\nu = n - 1$ degrees of freedom.

The probability distribution that arises when the sample size is small and there is a problem estimating the mean value of a normally distributed population with mean μ and standard deviation σ is called a t-distribution.

Note that there is a different t-distribution for each sample size. When one speaks about a specific t-distribution, we have to specify the degrees of freedom. The t-distribution curves are symmetric and bell-shaped like the normal distribution. However, the spread is different from that of standard normal distribution. t-distribution is the basis of Student's t-tests for the statistical significance.

TrainMiC
Training in Metrology in Chemistry

Significance testing

Statistics 4.0
Slide 25

A significance test involves testing the truth of a null hypothesis (e.g. the analytical procedure has no bias). The 'null' hypothesis implies that there is no difference between the measured value and the known value other than that accounted for by random variability.

Statistics can be used to calculate the probability of observing a given value taking into account the random variability. The lower the probability that the observed difference occurs by chance, the less likely it is that the null hypothesis is true.

Significance testing is an important tool in procedure validation. Most significance tests are named after the particular statistic used: t-test uses *t* statistics, the F-test uses *F* statistics, etc.

TrainMiC
Training in Metrology in Chemistry

Significance testing — overview

A decision at a given level of confidence about a population is based on observations from a sample of the population.

Tests covered:

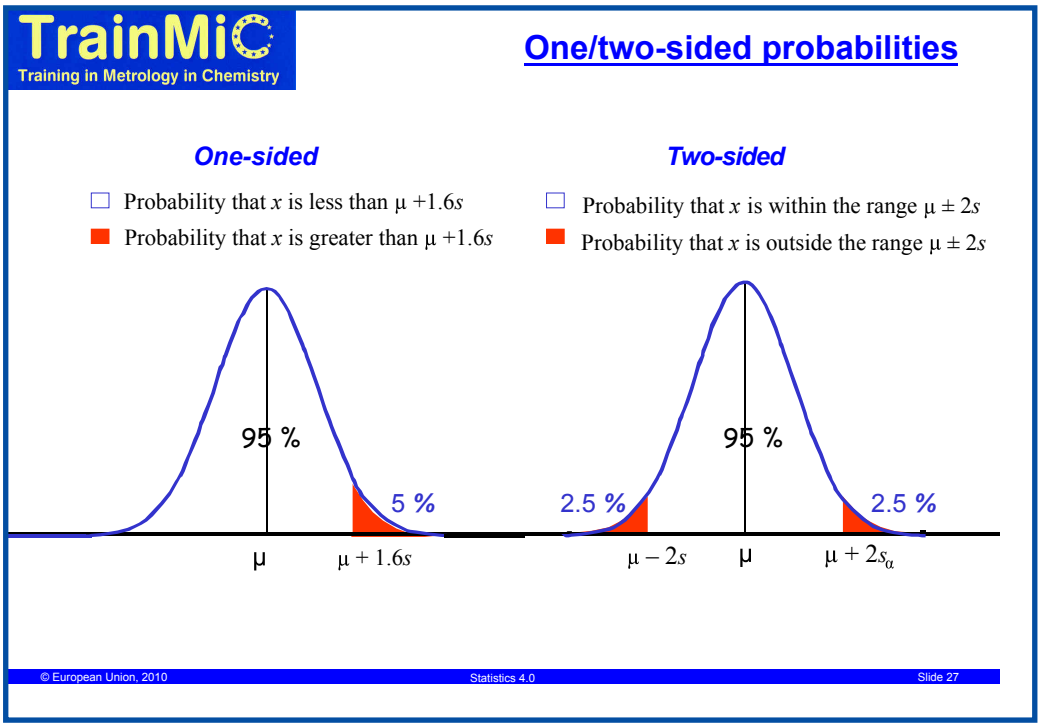
- t-test
  - — Testing for a significant difference between the (i) means and a reference value; (ii) two data sets (difference of means); or (iii) difference between pairs of measurement.
- F-test
  - — Testing for a significant difference between the spreads of two data sets (difference of *s*).

A significant test is used for:

- comparison of an experimental mean value with a known value;
- comparison of two experimental mean values;
- comparison of the standard deviations of two sets of data.

We may wish to test whether procedure A is more precise than procedure B (i.e. one-sided test) or we may wish to test whether procedure A and procedure B differ in their precision (i.e. two-sided test).

A significant difference between the spreads of two data sets (difference of *s*) can also be investigated.



The slide gives the background information about one-sided/two-sided probability.

The area under the distribution curve gives the probability of a result being in a particular region.

There are two ways of assigning 95.4 % of a distribution.

- One-sided (tailed)
  A one-sided test is referred to as a one-tailed test of significance (e.g. a limit for the specification of a product). It is only of interest whether a certain limit is exceeded or not. The critical region for a one-sided test is the set of values less than the critical value of the test, or the set of values greater than the critical value of the test.
- Two-sided (tailed)
  The critical region for a two-sided test is the set of values less than a first critical value of the test and the set of values greater than a second critical value of the test. A two-sided test is referred to as a two-tailed test of significance.

The choice between a one-sided and a two-sided test is determined by the purpose of the investigation or prior reasons for using a one-sided test, for example when analysing a reference material (Does a measured value lie within or outside the certified value ± a certain range?).
In both cases, at 95.4 % confidence level, there is a probability of approximately 5 % that the decision is wrong.

If the alternate hypothesis contains the phrase ‘different from’, the test is two-tailed.

TrainMiC
Training in Metrology in Chemistry

Significance testing: The eight steps

1. Formulate the question
2. Select the test
3. Decide on a one- or two-sided test
4. Choose the level of significance
5. Define null and alternative hypothesis
6. Determine the critical value
7. Evaluation of the test statistic using the appropriate equations
8. Decisions and conclusions

 Statistics 4.0 Slide 28

The slide is a summary of significance testing steps.

The t-test, and any statistical test, consists of the following steps.

- Formulate the question
- Select the test
- Decide on one- or two-sided test
- Choose the level of significance
- Define the null and alternative hypotheses
- Calculate the *t*-statistic for the data, t_{calc}
- Compare t_{calc} with the tabulated *t*-value, t_{crit} for the appropriate significance level and degree of freedom.
- If $t_{calc} > t_{crit}$, we reject the null hypothesis and accept the alternate hypothesis. Otherwise, we accept the null hypothesis.

TrainMiC
Training in Metrology in Chemistry

Significance testing: Steps 1–4

1. Formulate the question
2. Select the test
3. Decide on a one- or two-sided test
4. Choose the level of significance

   The level of significance is related to a probability:

   for most purposes, 95 % level of confidence is appropriate, corresponding to a of significance level of 0.05;

   at a 95 % level of confidence, there is a 5 % probability that a wrong decision will be made, rejecting the null hypothesis when it is true.

The next few slides go through the significance testing steps.

We have analysed a reference material and we want to know if the bias between the mean value of the measurements and the certified value of a reference material is 'significant' or just due to random variability.

Step 3: If the question is, 'Is the measured mean value significantly different form the certified value?', the test is two-sided.

Step 4: Is the selection of the level of significance. For most purposes, a 95.4 % level of confidence is an appropriate level. At a 95.4 % level of confidence, there is approximately a 5 % probability that a wrong decision will be made.

In significance testing, a test will indicate significance more often at a lower level of confidence (higher level of significance).

TrainMiC
Training in Metrology in Chemistry

Significance testing: Step 5

5. Define null and alternative hypotheses

Null hypothesis H_0

- The term 'null' is used to imply that there is no difference between the observed and known value, other than that which can be attributed to random variation ($\mu = x$).

Alternative hypothesis H_1

- The opposite of the null hypothesis: there is a difference,

($\mu \neq x$), where x is a sample mean; μ is a true value.

Null hypothesis

The null hypothesis, H_0, represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proven. We give special consideration to the null hypothesis. This is due to the fact that the null hypothesis relates to the statement being tested, whereas the alternative hypothesis relates to the statement to be accepted when the null hypothesis is rejected.

Alternative hypothesis

The alternative hypothesis, H_1, is a statement of what a statistical hypothesis test is set up to establish.

The final conclusion once the test has been carried out is always given in terms of the null hypothesis. We either 'Reject H_0 in favour of H_1' or 'Accept H_0': we never conclude 'Reject H_1', or even 'Accept H_1'.If we conclude 'Accept H_0', this does not necessarily mean that the null hypothesis is true, it only suggests that there is not sufficient evidence against H_0 in favour of H_1. Rejecting the null hypothesis then, suggests that the alternative hypothesis may be true.

TrainMiC
Training in Metrology in Chemistry

Significance testing: Step 5
Hypotheses and sides

H_0 *'The mean is equal to the value'* $\mu = x_0$

H_1 *'The mean is is less than the given value'* $\mu < x_0$

one-sided test

H_0 *'The mean is equal to the value'* $\mu = x_0$

H_1 *'The mean is greater than the given value'* $\mu > x_0$

one-sided test

H_0 *'The mean is equal to the value'* $\mu = x_0$

H_1 *'The mean is not equal to the given value'* $\mu \neq x_0$

two-sided test

Three different cases are presented in this slide.

H_0, the null hypothesis, in all cases is: 'The mean value is equal to the true value' $\mu = x_0$

If the alternate hypothesis contains greater than or less than, the test is one-sided.

If the alternate hypothesis contains different from, the test is two-sided.

TrainMiC
Training in Metrology in Chemistry

Significance testing: Step 6

6. Determine the critical value

- The **critical value** for a hypothesis test is a threshold to which the value of the test statistic in a sample is compared to determine whether or not the null hypothesis is rejected.
- the critical value for any hypothesis is set by:

➢ *the level of significance required*

➢ *degrees of freedom*

➢ *whether the test is one-sided or two-sided*

- Critical values are found in tables (also in Excel)

Significance testing involves comparing a calculated value with a critical value. The relevant statistics (*t*, F, etc.) are calculated for the data set in question and compared with the appropriate critical value. Each significant test has its own set of critical data (from statistical tables).

TrainMiC
Training in Metrology in Chemistry

Significance testing: Steps 7,8

7. Evaluation of the test statistic using the appropriate equations
8. Decisions and conclusions

calculated test statistic < critical value

- no significant difference at selected confidence level (under the given experimental conditions)

calculated test statistic > critical value

- there is a significant difference at selected confidence level (under the given experimental conditions)

*A significance test shows whether there is sufficient evidence to reject the **null hypothesis** at the selected level of confidence.*

Significance tests show whether there is sufficient evidence to reject the null hypothesis at a particular level of confidence. The procedure to determine if the result of the test is significant or not is the same for all tests. The calculated test statistic value is compared with the appropriate critical value. If the calculated value is greater than the critical value, the result of the test indicates a significant difference. This indicates that the observation (e.g. the difference between two mean values) is unlikely to have happened by chance, and the null hypothesis is rejected. If the calculated value is less than the critical value, the result of the test is not significant. The observed difference, in this case, could have happened by chance. Each significance test has its own critical values and they can be found from statistical tables.

TrainMiC
Training in Metrology in Chemistry

Critical value

- Critical values for t-tests and F-tests can be found in statistical tables and in Excel

  Each significant test has a set of critical values!
- Type of significant test: one and two samples t-test, paired t-test or F-test
- One- or two-sided test
- Degrees of freedom
- Level of confidence

 Statistics 4.0 Slide 34

This slide is about critical values, used in significance testing.

Significance testing involves comparison of a calculated value with a critical value. The critical value depends on the:

- type of significance test
- number of tails
- degrees of freedom
- level of confidence.

The critical value of F depends on the level of significance required and the degrees of freedom $\nu_A = n_A - 1$ and $\nu_B = n_B - 1$ and can be found in statistical tables.

TrainMiC
Training in Metrology in Chemistry

One sample t-test

In a comparison of experimental mean with a reference value or a nominal value

$$t_{calc} = \left(\bar{x} - x_0\right) / \frac{s}{\sqrt{n}}$$

s is a sample standard deviation, *n* sample size,
$\bar{x}$ sample mean, x_0 *stated value*

t_{crit} value for $\alpha = 0.05$ and degree of freedom $\nu = n - 1$

This slide gives the equation used to calculate the *t* statistic for comparing the mean value of a set of observations with a stated value (e.g. legal limit).

The t-test can be used to compare a sample mean value with an accepted value (a population mean), or it can be used to compare the means of two sample sets. *s* is the sample standard deviation or is known from previous measurements, but it could also be used for measurements without previously known *s*.

The term 'null' is used to imply that there is no difference between the observed and known values other than that which can be attributed to random variation. If the *t* value exceeds a certain critical value, then the null hypothesis is rejected.

The null hypothesis is $\bar{x} = x_0$

The alternatives are:

$\bar{x} > x_0$ one-sided test
$\bar{x} < x_0$ one-sided test
$\bar{x} \neq x_0$ two-sided test

If the alternate hypothesis contains 'greater than' or 'less than', the test is one-sided. If the alternate hypothesis contains 'different from', the test is two-sided.

TrainMiC
Training in Metrology in Chemistry

Two samples t-test

In a comparison of two experimental means, when two samples are drawn from a population with not significantly different standard deviations

$$t_{calc} = \frac{\left(\bar{x}_1 - \bar{x}_2\right)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

where: s_p is a pooled standard deviation
n is the sample size
$\bar{x}_1$ and $\bar{x}_2$ sample means
degrees of freedom $\nu = n_1 + n_2 - 2$

Another way in which the results of a new analytical procedure may be tested is by comparing them with those obtained by using a second (reference) procedure. In this case, we have two sample mean values, $\bar{x}_1$ and $\bar{x}_2$. One has to decide whether the difference between the two sample means is significant, that is to test the null hypothesis.

Considering the null hypothesis, the two procedures give the same result: H_0: $\bar{x}_1 = \bar{x}_2$.

A pooled estimate of the standard deviation can be calculated from the two individual standard deviations s_1 and s_2. This procedure assumes that the samples are drawn from the population where the standard deviations are not significantly different.

TrainMiC
Training in Metrology in Chemistry

Two samples t-test

In a comparison of two experimental means, when two samples are drawn from a population with different standard deviations

$$t_{calc} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

The degrees of freedom ν for the tabulated value $t_{critical}$ is:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\left(\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}\right)}$$

If the population standard deviations are unlikely to be statistically equal, then it is no longer appropriate to pool sample standard deviations in order to give one overall estimate of standard deviation.

Two-tailed: Are the results of two procedures significantly different?

One-tailed: Is the result from procedure 1 significantly lower/higher than the result from procedure 2?

TrainMiC
Training in Metrology in Chemistry

Paired t-test: Principle

- Two procedures of analysis are applied to measure several different samples and results are compared
- The *t*-value is calculated according to:

$$t_{calc} = \frac{\bar{d}_{difference}}{s_{difference} / \sqrt{n}}$$

where $\bar{d}_{difference}$ and $s_{difference}$ are the mean and standard deviation of d_i and d_i is the differences between paired values

- The critical *t*-value is taken from the *t*-table for the selected confidence level and $n - 1$ degrees of freedom

 Statistics 4.0 Slide 38

The t-test for comparing two means is not appropriate in this case because it does not separate the variation due to procedures from that due to variation between samples.

This difficulty is overcome by looking at the difference d_i between each pair of results given by the two procedures. In order to test the null hypothesis, we test whether d_i differs significantly from 0 using *t* statistics.

To test the null hypothesis, the procedure is as follows.

1. Calculate the difference ($d_i = y_i - x_i$) between the two observations on each pair, making sure you distinguish between positive and negative differences.
2. Calculate the mean difference $\bar{d}_{difference}$
3. Calculate the standard deviation of the differences, $s_{difference}$
4. Calculate the *t*-statistic.

Under the null hypothesis, this statistic follows a t-distribution with $(n - 1)$ degrees of freedom, *n* is a number of pairs of results.

5. Compare the value $t_{calculated}$ with $t_{critical}$.

TrainMiC
Training in Metrology in Chemistry

F-test

The F-test establishes if there is a significant difference between variances .

- The F-test considers the ratio of two sample variancess (i.e. ratio of the squares of the standard deviations, s_1^2/s_2^2).
- Answer the question: Are the spreads different (i.e. do the two sets of data come from two separate populations)?

This comparison can take two forms.

1. Is the precision of Procedure A higher than the precision of Procedure B (a one-sided test)?
2. Is the precision of Procedure A significantly different from the precision of Procedure B (a two-sided test)?

 Statistics 4.0 Slide 39

The significance test described so far is used for comparing means. In many cases, it is also important to compare the standard deviations of two sets of data.

An F-test could answer the question: Are the variances different or do the two sets of data come from two different population?

As with tests on mean values, this comparison can take two forms:

- one may wish to test whether procedure A is more precise than procedure B (a one-sided test); or
- one may wish to test whether procedure A and procedure B differ in their precision (a two-sided test).

For example, when one wishes to test whether a new analytical procedure is more precise than a standard procedure, a one-sided test should be used. When the test is whether two standard deviations differ significantly (e.g. before applying a t-test) a two-sided test is appropriate.

TrainMiC
Training in Metrology in Chemistry

F-test

- The *F*-value is calculated according to the equation:

$$F_{calc} = s_1^2/s_2^2, \text{ where } s_1^2 > s_2^2$$

The ratio is compared with $F_{critical}$ values from tables:

- $F_{critical}$ for α and appropriate ν_1, ν_2 (one-sided test)
- $F_{critical}$ for α/2 and appropriate ν_1, ν_2 (two-sided test)
- If $F_{calculated} < F_{critical}$ then the variances s_1^2 and s_2^2 are not significantly different for the given confidence level
- If $F_{calculated} > F_{critical}$ then the variances s_1^2 and s_2^2 are significantly different for the given confidence level
- $F_{critical}$ *is based on two values of degrees of freedom:*

$$\nu_1 = n_1 - 1 \text{ and } \nu_2 = n_2 - 1$$

F-test sequence:

1. One- or two-tailed test? The alternatives are:

$s_1^2 > s_2^2$ one-sided test
$s_1^2 < s_2^2$ one-sided test
$s_1^2 \neq s_2^2$ two-sided test

2. Formulate the level of confidence and significance.
Level of significance required (probability α = 0.05 for 95.4 % level of confidence)

3. The ratio calculated can be compared with values from tables:

one-tailed critical value $F_{critical}$ for α and appropriate ν_A, ν_B (one-sided test)
one-tailed critical value $F_{critical}$ for α/2 and appropriate ν_A, ν_B (two-sided test)

If $F_{observed} < F_{critical}$ then the variances s_A^2 and s_B^2 are not significantly different for the chosen confidence level.

If the null hypothesis is accepted, then the ratio should be close to 1.

The critical value of *F* depends on the level of significance required and the degrees of freedom $\nu_A = n_A - 1$ and $\nu_B = n_B - 1$ and can be found in statistical tables.

TrainMiC
Training in Metrology in Chemistry

Reporting of measurement results

TrainMiC
Training in Metrology in Chemistry

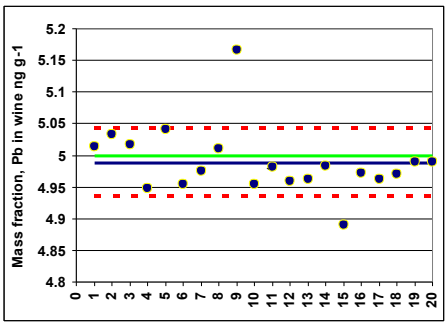
Reporting results

- Outlier rejection
- Significant figures
- Rounding results
- Reporting with expanded uncertainty

The remaining part of the presentation deals with outliers and the reporting of measurement results.

TrainMiC
Training in Metrology in Chemistry

Grubbs' test for outliers



	Data
1	5.01
2	5.03
3	5.02
4	4.95
5	5.04
6	4.96
7	4.97
8	5.01
9	4.97
10	4.96
11	4.98
12	4.95
13	4.96
14	4.98
15	4.89
16	5.16
17	4.96
18	4.97
19	4.98
20	4.99

In statistics, an outlier is an observation that is numerically distant from the rest of the data. Statistics derived from data sets that include outliers may be misleading. However, results should not be removed without a thorough examination of the data.

This slide is about Grubbs' test for outliers. Grubbs' test is used to detect outliers in a univariate data set. It is based on the assumption of normality. That is, we should first verify that our data can be reasonably approximated by a normal distribution before applying the Grubbs' test. Grubbs' test detects one outlier at a time. This outlier is from the data set and the test is iterated until no outliers are detected. However, multiple iterations change the probabilities of detection, and the test should not be used for sample sizes of fewer than six. Grubbs' test is also known as the maximum normed residual test.

TrainMiC
Training in Metrology in Chemistry

Grubbs' test for outliers

- A result appears to differ unreasonably from the others in the set
- In order to use Grubbs' test for an outlier, that is to test H_0: all measurements come from the same population, the statistic G is calculated

$$G_{calc} = |x_i - \bar{x}| / s$$

where x_i is a suspect value and s and $\bar{x}$ are calculated with the suspect value included

- If $G_{calc} > G_{critical}$, the suspect value may be rejected
- *Rejection only on statistical grounds is questionable!!!*

It should be mentioned that this presentation of Grubbs' test is a simplified presentation. The Grubbs' test statistic is the largest absolute deviation from the sample mean in units of the sample standard deviation. The test assumes that the population is normal. It applies to a single outlier.

Grubbs' test is defined for the hypothesis:

- H_0: there are no outliers in the data set;
- H_1: there is at least one outlier in the data set.

The critical values for G are given in tables. The values given are for a two-sided test, which is appropriate when is not known in advance at which extreme an outlier may occur.

TrainMiC
Training in Metrology in Chemistry

Rules for the number of significant figures

- All non-zero digits are significant — **1.234 g has four significant figures**
- Zeros between non-zero digits are significant — **1 002 kg has four significant figures**
- Leading zeros to the left of the first non-zero digit are not significant — **0.01° C has only one significant figure**
- Trailing zeros that are also to the right of a decimal point in a number are significant — **0.0230 mL has three significant figures**

TrainMiC
Training in Metrology in Chemistry

Significant figures

General presentation of the measurement result: (201 ± 26) ng g^{-1}

The number of significant figures in a result is the number of figures that are known with some degree of reliability.

Keep all the figures during calculations

 Statistics 4.0 Slide 46

All measurements are approximations — no measurement result could be without uncertainty. In carrying out calculations, the general rule is that the accuracy of a calculated result is limited by the least accurate measurement involved in the calculation.

Examples: the number 13.2 has three significant figures
the number 13.20 has four significant figures
the number 0.001 has one significant figure
the number 1.000 has four significant figures

Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs states: 'The results shall be expressed in the same units and with the same number of significant figures as the maximum level laid down in Regulation'.

TrainMiC
Training in Metrology in Chemistry

Rounding of results

- 'It usually suffices to quote $u_c(y)$ and U to at most two significant digits, although in some cases it may be necessary to retain additiona digits to avoid round-off errors in subsequent calculations'.

(GUM [5], para. 7.2.6)

U and x ($y = x \pm U$) shall have the same number of decimal places

Round the final result when the measurement uncertainty has been calculated.

 Statistics 4.0 Slide 47

Rules for rounding off numbers

When the answer to a calculation contains too many significant figures, it must be rounded off. There are 10 digits that can occur in the last decimal place in a calculation. One way of rounding off involves underestimating the answer for five of these digits (0, 1, 2, 3 and 4) and overestimating the answer for the other five (5, 6, 7, 8 and 9). This approach to rounding off is summarised as follows. If the digit is smaller than 5, drop this digit and leave the remaining number unchanged. If the digit is 5 or larger, drop this digit and add 1 to the preceding digit.

- If the digit to be dropped is greater than 5, the last retained digit is increased by one. For example, 18.6 is rounded to 19.
- If the digit to be dropped is less than 5, the last remaining digit is left as it is. For example, 18.4 is rounded to 18.
- If the digit to be dropped is 5, and if any digit following it is not zero, the last remaining digit is increased by one. For example, 12.51 is rounded to 13. If the digit to be dropped is 5 and is followed only by zeros, the last remaining digit is increased by one if it is odd, but left as it is if even. For example, 13.50 is rounded to 14, 12.50 is rounded to 12. This rule means that if the digit to be dropped is 5 followed only by zeros, the result is always rounded to the even digit.

In addition, subtraction, multiplication and division, the result is rounded off to the last common digit occurring furthest to the right in all components. Another way to state this rule is: in addition and subtraction, the result is rounded off so that it has the same number of decimal places as the measurement having the fewest decimal places.

TrainMiC
Training in Metrology in Chemistry

Reporting results with uncertainty?

A result may be presented as follows:

$$w_{Cd} = (21.4 \pm 4.2)\ mg\ kg^{-1}$$

But what is 4.2?

- Standard deviation?
- Rectangular interval?
- Triangular interval?
- Confidence interval without specified numbers of degrees of freedom?
- Confidence interval with specified numbers of degrees of freedom?
- Combined uncertainty (u_c)?
- Expanded uncertainty (U)? Is 'k' specified?

If $w_{Cd} = (21.4 \pm 4.2)\ mg\ kg^{-1}$ ($k = 2$), then 4.2 is the expanded uncertainty.

 Statistics 4.0 Slide 48

The value 4.2 might be a standard deviation, rectangular interval, triangular interval or confidence interval without specified numbers of degrees of freedom.

If the result is given as a value ± uncertainty, the k factor should be stated and the level of confidence this provides.

For a laboratory, the uncertainty determines the significant figures to be used in the presentation of measurement result. The European co-operation for Accreditation (EA) statement and GUM agree on this [5]. The expanded uncertainty should have no more than two significant figures.

Statistics is a very useful tool used to answer a number of questions. Nevertheless, statistics should always be applied with a critical view of the results and whether they make scientific sense.
No blind use of statistics!

Summary

- Statistical parameters
- Various distributions
- Statistics for the evaluation of uncertainty of results according to ISO-GUM [5]
- Statistics for procedure performance studies
- Significance testing
- Presentation of results

Chapter 4

Statistics for analytical chemistry — Part II

In this presentation, statistical concepts that provide the necessary foundations for more specialised expertise in any area of chemical analysis are briefly discussed. The selected topics (regression and correlation, linear regression, calibration, residuals and residual analysis) illustrate the basic assumptions of most analytical methods and are necessary components of our general understanding of the 'quantitative analysis'. Further information is included and mostly deals with the functional aspects on the concepts widely used for validation of analytical methods as α and β errors, limit of detection and control charts. The simplest form of the analysis of variance (ANOVA) — one-way ANOVA is also discussed.

The aim of this presentation is to familiarise users with the basics of applied statistics and to help them to design and conduct their experiments properly and extract as much information from the results as they legitimately can.

TrainMiC
Training in Metrology in Chemistry

Statistics for analytical chemistry

Part II

Last updated - January 2011

Statistics is a tool providing a means of reaching objective decisions and also a useful tool for summarising data.

TrainMiC
Training in Metrology in Chemistry

Aim

To familiarise users with applied statistics, help them to design and conduct their experiments properly and extract as much information from the results as they legitimately can.

The aim of this presentation is to familiarise users with applied statistics, and provide some help to design and conduct experiments properly and to extract as much information from the results as is legitimately possible.

TrainMiC
Training in Metrology in Chemistry

Summary

- Regression and correlation
- α and β errors
- Limit of detection
- Control charts
- Analysis of variance (ANOVA)

The topics covered by this presentation are shown in the slide.

TrainMiC
Training in Metrology in Chemistry

Regression and correlation

TrainMiC
Training in Metrology in Chemistry

Regression

- A statistical measure that attempts to determine the strength of the relationship between one dependent variable (usually denoted by *Y*) and a series of other changing variables *X* (known as independent variables).
- A **regression equation** indicates the nature of the relationship between two (or more) variables and the extent to which one can predict some variables by knowing others, or the extent to which some are associated with others.
- The relationship, typically in the form of a straight line (linear regression), that best approximates all the individual data points (**regression line**).

Simple regression is used to examine the relationship between one dependent and one independent variable. After performing an analysis, the regression statistics can be used to predict the dependent variable when the independent variable is known.

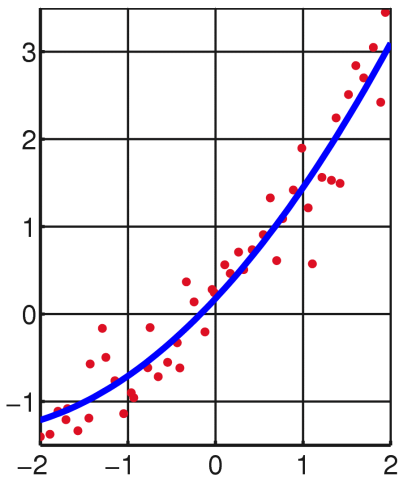
A regression equation allows us to express the relationship between two (or more) variables algebraically and indicates the nature of the relationship between them. In particular, it indicates the extent to which one can predict some variables by knowing others, or the extent to which some are related to others.

A regression line is a line drawn through the points on a scatter plot to summarise the relationship between the variables being studied.

TrainMiC
Training in Metrology in Chemistry

Regression analysis

- **Regression analysis** is a statistical tool for the investigation of relationships between variables.
- **Regression analysis** is the process of constructing a mathematical model or function that can be used to predict or determine one variable by another variable or other variables.



 Statistics 2 - 2.2 Slide 6

Regression analysis provides a 'best-fit' mathematical equation for the relationship between the dependent variable (response) and independent variable(s) (covariates). Regression analysis helps us to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held constant. In regression analysis, it is also of interest to characterise the variation of the dependent variable around the regression function, which can be described by a probability distribution. Regression analysis refers to techniques for the modelling and analysis of numerical data consisting of values of a dependent variable and of one or more independent variable.

Example: signal related with the concentration.

TrainMiC
Training in Metrology in Chemistry

Correlation

- Correlation quantifies the degree to which two variables are related. The variables are not designated as dependent or independent.
- Correlation does not find a best-fit line.
- The correlation coefficient (*r*) that shows how much one variable tends to change when the other one does.
- Regression goes beyond correlation by adding prediction capabilities.

What is correlation? When two variables vary together, statisticians say that there is covariation or correlation. The correlation coefficient, *r*, quantifies the direction and magnitude of correlation. The correlation analysis reports the value of the correlation coefficient.

TrainMiC
Training in Metrology in Chemistry

Linear regression

TrainMiC
Training in Metrology in Chemistry

Linear regression (1)

- Simple linear regression aims to find a linear relationship between a response variable and a possible predictor variable by the method of least squares.
- A linear regression equation is usually written

$$Y = bx + a + e$$

Y is the dependent variable

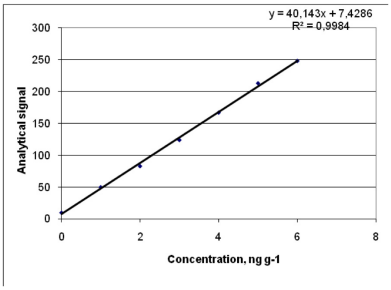
a is the intercept

b is the slope or regression coefficient

x is the independent variable

r is a correlation coefficient

e is the error term



For the given concentration of *x*, calculating the predicted values of y indicates how close the actual values are to the estimated one.

In statistics, linear regression includes any approach to modelling the relationship between a variable *y* and one or more variables denoted by *x*, such that the model depends linearly on the unknown parameters to be estimated from the data. Such a model is called a 'linear model'.

A regression equation allows us to express the relationship between two (or more) variables. It indicates the nature of the relationship between two (or more) variables. In particular, it indicates the extent to which you can predict some variables by knowing others, or the extent to which some are related to others.

TrainMiC
Training in Metrology in Chemistry

Linear regression (2)

- Regression parameters for a straight line model ($Y = a + bx$) are calculated by the least squares method (minimisation of the sum of squares of deviations from a straight line).

Regression assumptions:

- Y is **linearly related** to x or a transformation of x:
- deviations from the regression line (**residuals**) follow a **normal distribution**.
- deviations from the regression line (**residuals**) have **uniform variance**.

Linear regression is based on a number of assumptions. In particular, one of the variables must be 'fixed' experimentally and/or precisely measureable. So, the simple linear regression methods can be used only when we define some experimental variable (temperature, pH, dosage, etc.) and test the response of another variable to it.

The most common form is a linear regression of y on x (i.e. the x values are deemed to be known exactly and the only error occurs in the determination of y). The position of the line is determined by two factors: the slope and the intercept. For the given concentration of x, calculating the predicted values of y indicates how close the actual values are to the estimated one.

It should be clear that there are other linear regression models than the least squares regression model.

TrainMiC
Training in Metrology in Chemistry

Least squares (1)

- The least squares method is a technique for fitting a straight line through a set of points in such a way that the sum of the squared vertical distances from the observed points to the fitted line is minimised.
- The best fit in the least-squares sense minimises the sum of squared residuals, a **residual** being the difference between an observed value and the value provided by a model.

The goal of linear regression is to adjust the values of slope and intercept to find the line that best predicts y from x. The line of regression of y on x assumes that all errors are in y-direction (between the experimental points and the calculated line). Since some of these deviations are positive and some negative it is sensible to seek to minimise the sum of the squares of the residuals. The goal of regression is to minimise the sum of the squares of the vertical distances of the points from the line. This explains the frequent use of the term 'method of least squares for the procedure'. Parameters are estimated to give a 'best fit' of the data.

Most commonly, the best fit is evaluated by using the least squares method. In a narrow sense, the least squares method is a technique for fitting a straight line through a set of points in such a way that the sum of the squared vertical distances from the observed points to the fitted line is minimised. 'Least squares' means that the overall solution minimises the sum of the squares of the errors made in solving every single equation.

TrainMiC
Training in Metrology in Chemistry

Least squares (2)

Least squares can be interpreted as a method of fitting data.

The best fit in the least-squares sense minimises the sum of squared residuals.

The line of regression Y on x calculated on this principle must pass through the centroid of all points (x, y)

where $\bar{x}$ is the mean of the x values, $\bar{y}$ is the mean of y values.

The most important application is in data fitting. The best fit in the least-squares sense minimises the sum of squared residuals, a residual being the difference between an observed value and the value provided by a model. The method of least squares, used to obtain this best line, minimises the sum of squares of the differences between the actual value of y and the predicted value (y residuals): the line obtained is the best line that can be fitted to the data. The line of regression of x on y assumes that all the errors occur in the x direction and also passes through the centroid of the points. If we maintain rigidly the concentration that the analytical signal is always plotted on the y-axis and the concentration on the x-axis, it is always the line of regression of y on x that we must use in calibration experiments.

TrainMiC
Training in Metrology in Chemistry

Correlation coefficient (1)

The correlation coefficient (r) is given as a measure of the degree of association between two variables (a change in x produces a predictable change in Y):

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}$$

$$-1 \le r \le 1$$

r^2 is the proportion of the total variance (s^2) of Y that can be explained by the linear regression of Y on x.

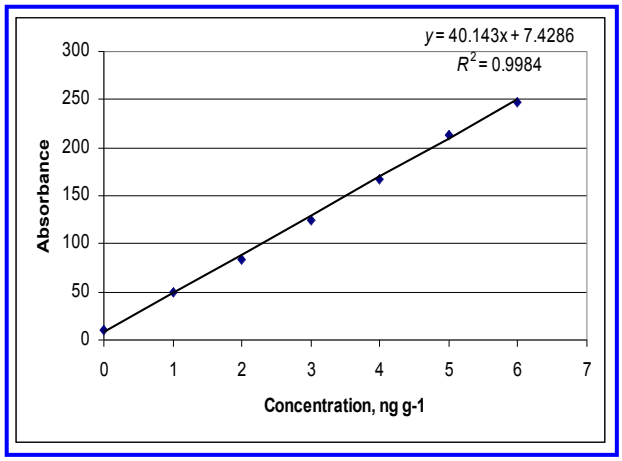
Covariance of two variables x and y is:

$$\sum\left\{\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right\} / n$$

TrainMiC
Training in Metrology in Chemistry

Correlation coefficient

Excel

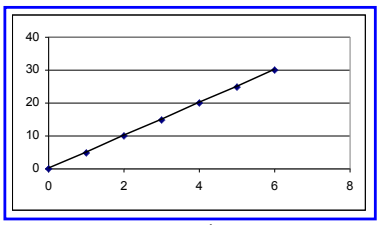


Although correlation coefficients are easy to calculate, they are often misinterpreted. The calibration curve must always be plotted — otherwise a straight line relationship might wrongly be deduced from the calculation of r.

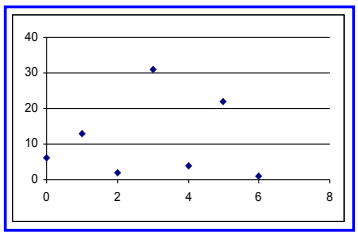
With linear regression, it is conventional to use the abbreviation r^2. With non-linear regression, the convention is to use R^2.

TrainMiC
Training in Metrology in Chemistry

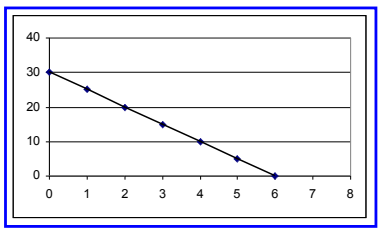
Correlation coefficient (2)



$r = 1$



$r = 0$



$r = -1$

The calibration curve must always be plotted — a straight line relationship might wrongly be deduced from the calculation of *r*.

It can be shown that r can take values in the range $-1 \leq r \leq 1$.

An r value of -1 describes perfect negative correlation (i.e. all the experimental points lie on a straight line of negative slope).

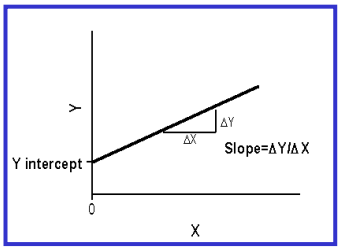
When $r = +1$, there is perfect positive correlation (i.e. all the points lying exactly on a straight line with a positive slope).

When there is no linear correlation between x and y, the value r is close to zero.

Experience shows that even quite poor looking calibration plots give high r values. In such a case, the numerator and denominator in the r equation are nearly equal. It is very important to do the calculation with an adequate number of significant figures. Zero correlation coefficient does not mean that x and y are entirely unrelated — it only means that they are not linearly related.

TrainMiC
Training in Metrology in Chemistry

Slope and intercept



The slope of the least squares line is:

$$b = \frac{\sum_i \{(x_i - \bar{x}) \times (y_i - \bar{y})\}}{\sum_i (x_i - \bar{x})^2}$$

The intercept of the least squares line is:

$$a = \bar{y} - b\bar{x}$$

The regression determined by *a* and *b* is known as the line of regression of *y* on *x*. The line of regression of *x* on *y* is not the same line (*except when R = 1 exactly*).

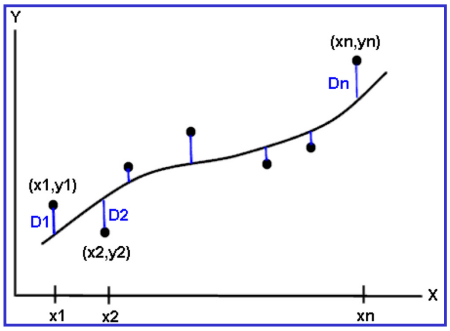
The slope equals the change in *y* for each unit change in *x*. It is expressed in the units of the *y*-axis divided by the units of the *x*-axis. If the slope is positive, *y* increases as *x* increases. If the slope is negative, *y* decreases as *x* increases.

It is important to emphasise that equations given in this slide must not be misused — they will only give useful results when prior study (calculation of *r* and visual inspection of the points) has indicated that a straight line relationship is realistic for the experiment in question.

TrainMiC
Training in Metrology in Chemistry

Residuals

The residual represents unexplained (or residual) variation after fitting a regression model. It is the difference (or leftover) between the observed value of the variable and the value suggested by the regression model.



The sum of squares of residuals =

$$= D_1^2 + D_2^2 + \cdots + D_N^2$$

$$D_i = y_i - \overline{y_i}$$

$$\overline{y_i} = a + bx_i$$

A regression line is a line drawn through the points on a scatter plot to summarise the relationship between the variables being studied. For a given value of x, say x_1, there will be a difference between the value y_1 and the corresponding value as determined by the 'best fitting' curve. This distance, D_1, is referred to as a residual.

When the regression line slopes down (from top left to bottom right), this indicates a negative or inverse relationship between the variables; when it slopes up (from bottom right to top left), a positive or direct relationship is indicated.

A residual is the difference from the actual y value and the value obtained by plugging the x value (that goes with the y value) into the regression equation. Using these residuals, the following definition has been developed: of all curves approximating a given set of data points, the curve having the property that $(D_1 + D_2 + D_3 + \ldots + D_n)$ is a minimum is called 'the best-fitting curve'. A curve having this property is said to fit the data in the least-squares sense and is called 'a least-squares curve'.

TrainMiC
Training in Metrology in Chemistry

Residual analysis

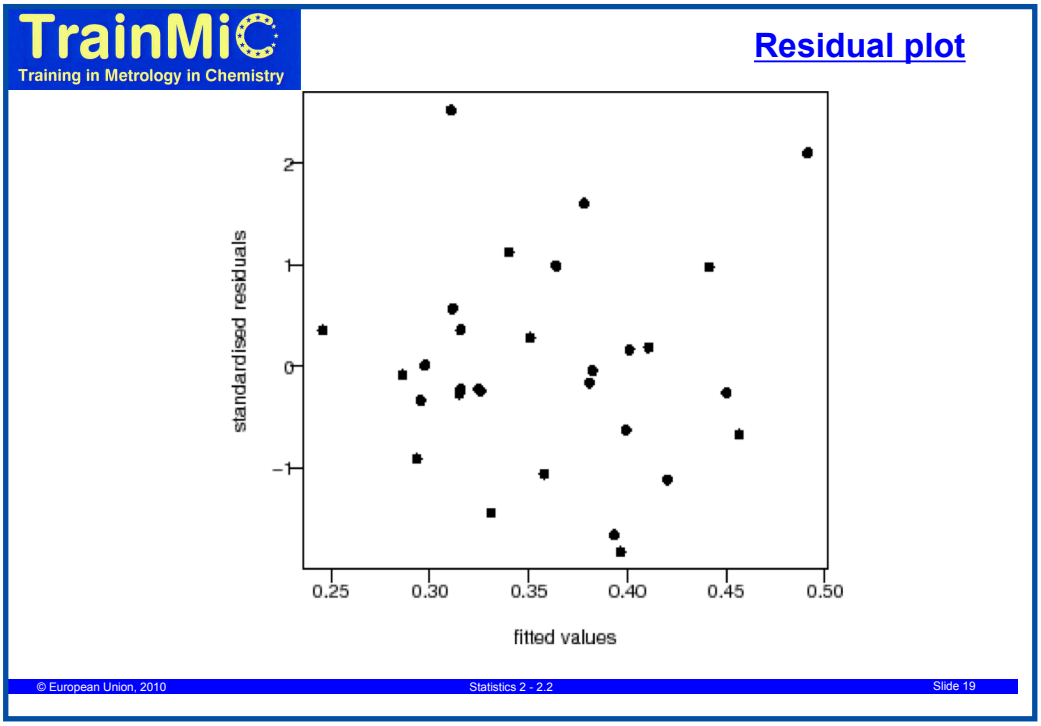
It is the sum of squares of residuals that is minimised to find the least squares line.

The assumption of constant standard deviation of the *y* value (an uniform dispersion of data points about the regression line) — homoscedasticity model.

If the standard deviation of the *y* values are not constant — heteroscedasticity model.

Each difference between the actual *y* values and the predicted *y* values is the error of the regression line at a given point and is referred to as a residual. One of the major uses of residual analysis is to test some of the assumptions underlying regression. The following are the assumptions made in simple regression analysis.

1. The model is linear.
2. The variables have constant variances.
3. The variables are independent.
4. The variables are normally distributed.



A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. A residual is positive when the point is above the curve, and is negative when the point is below the curve. If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data, otherwise, a non-linear model is more appropriate. The plot in the slide shows a random pattern, indicating a good fit for a linear model. Mild deviations of data from a model are often easier to spot on a residual plot.

TrainMiC
Training in Metrology in Chemistry

Residual, slope and intercept standard deviations

For a *y* on *x* regression, the following definitions apply:

Residual standard deviation:

$$s_{y/x} = \sqrt{\frac{\sum_i (y_i - \bar{y}_i)^2}{n-2}}$$

$y_i - \bar{y}_i$ are residuals

$n - 2$ degrees of freedom

Standard deviation of slope:

$$s_b = \frac{s_{y/x}}{\sqrt{\sum_i (x_i - \bar{x})^2}}$$

Standard deviation of intercept:

$$s_a = s_{y/x} \sqrt{\frac{\sum_i x_i^2}{\sum_i (x_i - \bar{x})^2}}$$

The calculated regression line will, in practice, be used to estimate the measurand in test materials by interpolation. The random errors in the values for the slope and intercept are thus of importance. First, we must calculate $s_{y/x}$ which estimates the random error in the *y* direction.

The $\hat{y}_i$ values for a given value of *x* is calculated from the regression equation. The formula for the residual standard deviation is very similar to the equation for the s_d of a set of repeated measurements, with $(n - 2)$ is degrees of freedom.

With $s_{y/x}$ we can now calculate s_a and s_b and then estimate the confidence limits for the slope and intercept.

TrainMiC
Training in Metrology in Chemistry

Confidence intervals for the slope, intercept and regression line

- Using the residual standard deviation, we can obtain estimates of the standard deviations of the slope (s_b), intercept (s_a) and the regression line ($s_{y/x}$).
- The confidence interval for these are:

$a \pm t \times s_b$ $\qquad$ $b \pm t \times s_a$

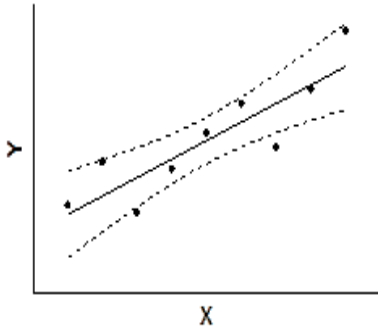
$$y \pm t \times s_{y/x}$$

t — *95 %* confidence level, two-tailed test with $n - 2$ degrees of freedom

TrainMiC
Training in Metrology in Chemistry

Confidence intervals

- Two confidence bands surrounding the best-fit line define the confidence interval of the best-fit line.



The confidence limits for *x* range by using unweighted regression line

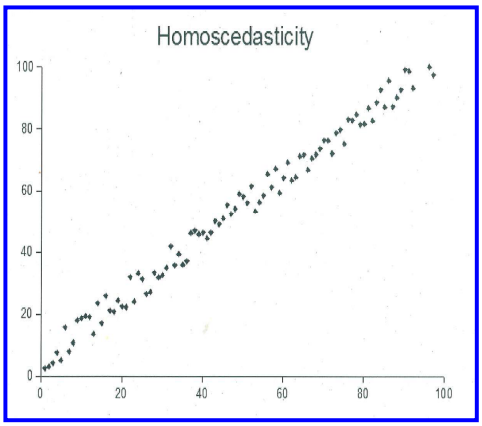
Statistics 2 - 2.2
Slide 22

The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter. A very wide interval may indicate that more data should be collected before anything more definite can be said about the parameter.

TrainMiC
Training in Metrology in Chemistry

Homoscedasticity

Homoscedasticity refers to the fact that the variance σ^2 of the response (or 'dependent') variable *y* is constant across the range of the predictor(s) *x*.



The property of having equal statistical variances

Homoscedasticity requires that the standard deviation and variance of the error terms are constant for all *x*, and that the error terms are drawn from the same population. This indicates that there is a uniform scatter or dispersion of data points about the regression line.

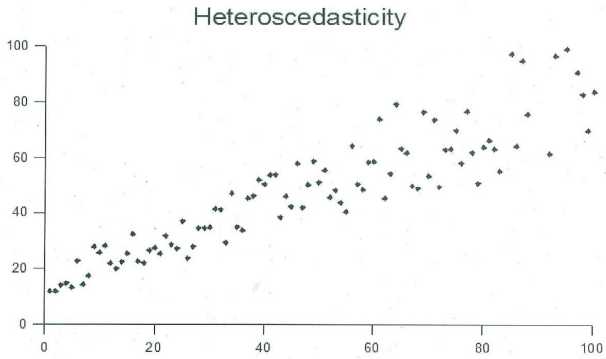
The assumption of homoscedasticity is that the residuals are approximately equal for all predicted values. Data are homoscedastic if the residuals plot is the same width for all values.

In regression analysis, homoscedasticity means a situation in which the variance of the dependent variable is the same for all the data.

Homoscedasticity facilitates analysis because most methods are based on the assumption of equal variance. In this case, the *y* direction errors in the calibration curve to be approximately equal for all the points and unweighted regression calculation is legitimate.

TrainMiC
Training in Metrology in Chemistry

Heteroscedastisity



Weighted least squares as a solution to heteroscedastisity

As previously noted, in many cases, the data are heteroscedastic (i.e. standard deviation of the *y* values increases/changes with the concentration of the analyte, rather than having the same value at all concentrations). In such a case, weighted regression calculations should be used instead.

TrainMiC
Training in Metrology in Chemistry

Ways to detect heteroscedasticity

- Scatter plot of *X* against *Y* (prior to analysis)
- Scatter plot of predictions against residuals, either
  - Scatter plot of *X* against relative value of residuals
  - Scatter plot of *X* against the absolute value of the residuals
  - Scatter plot of *X* against the squared residuals
  - F-test

TrainMiC
Training in Metrology in Chemistry

Correlation/linear regression

Correlation

- Quantifies the degree to which two variables are related but doesn't aim to find a linear relationship
- It doesn't matter which of the two variables we call '*X*' or '*Y*'
- Almost always used when both variables are measured

Linear regression

- The line that best predicts *Y* from *X* is not the same as the line that predicts *X* from *Y*
- The *X* variable is often something we experimentally manipulate (time, concentration, etc.) and the *Y* variable is something we measure

TrainMiC
Training in Metrology in Chemistry

Calibration

- **Calibration function**

  Functional relationship between the expected value of the response variable and the value of the net state variable, *X*

  The calibration function is conceptual and cannot be determined empirically. It is estimated through calibration.

- **Calibration**

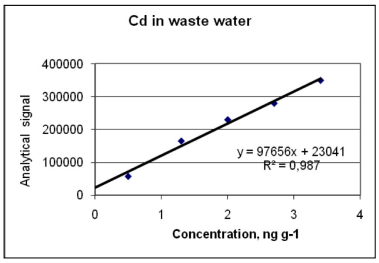
  Complete set of operations which estimates under specified conditions the calibration function from observations of the response variable, *Y*, obtained on reference states
  (see also VM3 [1]: Entry 2.39)

Calibration is a functional relationship between the expected value of the response variable and the value of the net state variable, x.

TrainMiC
Training in Metrology in Chemistry

Experimental data

x_i	y_i
0.5	57000
1.3	165000
2	230000
2.7	280000
3.4	350000



$$y_i = 97656x_i + 23041$$

$$R^2 = 0.9870$$

A calibration curve is a plot of how the instrumental response, the so-called analytical signal, changes with the concentration ([6]) of the analyte (measurand). An analyst prepares a series of measurement standards across a range of concentrations near the expected concentration of analyte in the unknown sample. Concentrations of the standards must lie within the working range of the technique (instrumentation) they are using. Analysing each of these standards using the chosen technique will produce a series of measurements. For most analyses a plot of instrument response v analyte concentration will show a linear relationship. The operator can then measure the response of the unknown sample and, using the calibration curve, can *interpolate* to find the concentration of analyte.

Most analytical techniques use a calibration curve. There are a number of advantages to this approach, such as that a calibration curve provides a reliable way to calculate the uncertainty of the concentration calculated from the calibration curve (using the statistics of the least squares line fit to the data).

([6]) Measurand can also be expressed in terms other than 'concentration' (eg. mass fraction).

TrainMiC
Training in Metrology in Chemistry

Type I (α) and Type II (β) errors

TrainMiC
Training in Metrology in Chemistry

α and β errors

- **Type I error**, also known as an **α error** or a '**false positive**': the error of rejecting a null hypothesis when it is actually true.
- **Type II error**, also known as a **β error**, or a '**false negative**': the error of failing to reject a null hypothesis when it is in fact not true.
- The ***probability*** of committing a Type I error is an **α error** or **level of significance.**
- The ***probability*** of committing a Type II error is a **β error.**

 Statistics 2 - 2.2 Slide 30

Because the hypothesis testing process uses sample statistics calculated from random data to reach conclusions about population parameters, it is possible to make an incorrect decision about the null hypothesis. In particular, two types of errors can be made in testing hypotheses: Type I errors and Type II errors.

- A Type I error (α) is committed by *rejecting a true null hypothesis*. With a Type I error, the null hypothesis is true, but the researcher decides that it is not.
- A Type II error (β) is committed when *failing to reject a false null hypothesis*. In this case, the null hypothesis is false, but a decision is made not to reject it.

Actually, because β occurs only when the null hypothesis is not true, the computation of β varies with the many possible alternative parameters that might occur. Unlike α, β is not usually stated at the beginning of the hypothesis testing procedure.

TrainMiC
Training in Metrology in Chemistry

False positives and false negatives (1)

False Positive, or Type I (α) error, means concluding that a substance is present when it is not.

False Negative, or Type II (β) error, means concluding that a substance is not present when it is.

	Null false	Null true
Fail to reject null	Correct Decision	Type II error (β)
Reject null	Type I error (α)	Correct Decision (power)

Let us imagine using a given analytical procedure in the concentration domain, knowing its precision along the different concentration levels and the results having a normal distribution. If we analyse many blank samples, we would obtain a distribution of values resembling that of a normal distribution.

The concentration values (in absence of bias in the procedure) would be distributed around zero with a given standard deviation, σ_0. This means that, as a result of the measurement of several blank samples, we could obtain a non-zero concentration, associated with σ_0. Being responsible for the results provided by the laboratory, we would like to limit the distribution at some point. This point is the critical level, L_C, and allows us, once the sample has been measured, to make a decision whether the analyte is present or not. If the concentration obtained is higher than L_C, then it probably does not correspond to a blank and we could state that the analyte is present in the sample. We, however, are running a risk when limiting the distribution at L_C. There is a certain probability that the analysis of a blank sample would give as a result a concentration value higher than L_C. In this case, we would falsely conclude that the component is present. This probability, α, is a Type I error, or, more commonly, the probability of committing a false positive.

Choosing the value of α is our decision, depending on the risk of being wrong we are willing to accept. We could, for example, fix L_C at a concentration level of zero. The risk of committing a false positive in this case would be of 50 % (any concentration value above zero found in a sample would be taken as a positive detection). Defining L_C in such a way that the risk is limited to, for instance, 5 % ($\alpha = 0.05$) seems a more appropriate decision in most situations.

TrainMiC
Training in Metrology in Chemistry

False positives and false negatives (2)

Experiment	Analyte not present	Analyte present
	Decisions	
Not detected		
($x < x_c$)	True positive ($P = 1 - \alpha$)	False negative ($P = \beta$)
Detected		
($x > x_c$)	False positive ($P = \alpha$)	True positive ($P = 1 - \beta$)

How are α and β related? First of all, because α can only be committed when the null hypothesis is rejected and β can only be committed when the null hypothesis is not rejected, a researcher cannot commit both a Type I error and a Type II error at the same time on the same hypothesis test. Generally, α and β are inversely related. If α is reduced, then β is increased, and vice versa.

Recall that:

x — concentration of the analyte
x_c — concentration of the analyte at the limit of detection

TrainMiC
Training in Metrology in Chemistry

Limit of detection

TrainMiC
Training in Metrology in Chemistry

Limit of detection (1)

- The signal at the limit of detection or the quantity, X_L, is derived from the smallest measure, that can be detected with reasonable certainty for a given analytical procedure. The value of X_L is given by the equation:

$$X_L = x_{bl} + ks_{bl}$$

where:

x_{bl} is the mean of blank measures,
s_{bl} is the standard deviation of the blank measures, and
k is a numerical factor chosen according to the confidence level desired.

(IUPAC Recommendations, 1995) [19]

There is always some uncertainty associated with any instrumental measurement. This also applies to the baseline (or background or blank) measurement (i.e. the signal obtained when no analyte is present). Various criteria have been applied to this determination; however, the generally accepted rule in analytical chemistry is that the signal must be at least three times greater than the background noise.

Formally, the limit of detection (LOD) is defined as the concentration of analyte required to give a signal equal to blank plus three times the standard deviation of the blank. So, before any calibration or sample measurement is performed, we need to evaluate the blank. This gives the minimum signal that can be interpreted as a meaningful measurement. To find the associated concentration, the calibration curve should be used to convert the signal to a concentration.

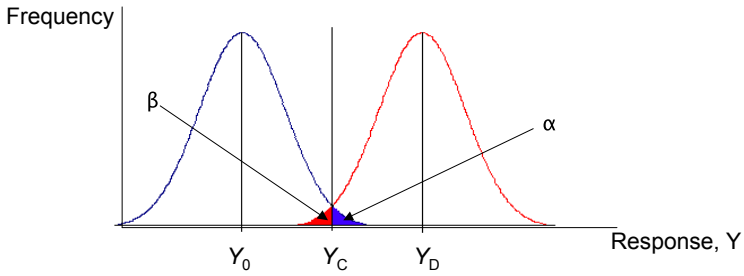
Where no blank has been measured, we can use the calibration data and regression statistics instead.

The LOD represents the level below which we cannot be confident whether or not the analyte is actually present. It follows from this that no analytical method can ever conclusively prove that a particular chemical substance is not present in a sample, only that it cannot be detected. In other words, there is no such thing as zero concentration!

TrainMiC
Training in Metrology in Chemistry

Limit of detection (2)

The Limit of Detection (LOD) is the smallest quantity of analyte, of which it can be said, with a given level of confidence, is present in the sample.



Y_0 is the response variable corresponding to blank
Y_C is the critical value of the response variable
Y_D is the response variable corresponding to LOD

This graph is in the signal domain. The limit of detection (LOD) is the smallest quantity of analyte, of which it can be said, with a given level of confidence, that it is present in the sample. As shown in the figure in the slide, the LOD depends on the variation of the method at the blank level, σ_0, and on two risk values α and β (α corresponds to the risk of detecting the analyte although it is not present).

The limit identified as the critical value is usually obtained by multiplying the standard deviation of observation from a blank variable, σ_0, by one-tailed Student's *t* value for infinite degrees of freedom and the appropriate value of α and adding this to the mean blank response if the blank response is significantly different from a false positive rate 5 % ($\alpha = 0.05$ is typically used). This gives a critical value of 1.65 σ_0 if the response variable corresponding to the blank is zero.

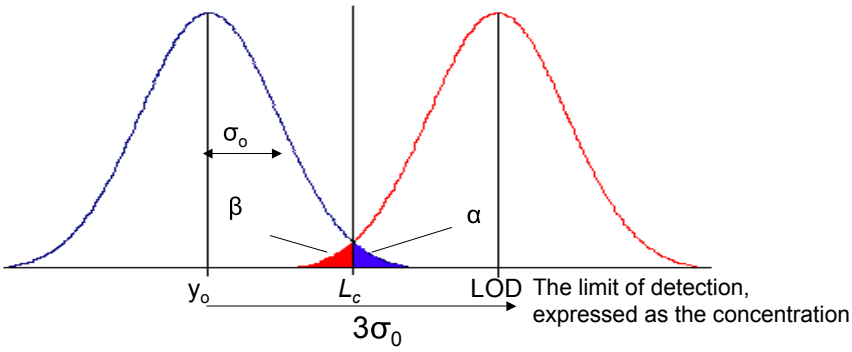
The critical value Y_C is determined by three parameters: the blank value, α value, and the σ_0. With Y_C fixed, the LOD depends solely on β, the value of the risk of not detecting the analyte although it is present.

Typically, β is set equal to α, that is 0.05 % to represent a 5 % false negative rate and *t* is taken for the greatest degrees of freedom, that is $t = 1.65$.

The limit of detection is then approximately: LOD = Y_C + (1 σ_0 1.65) + (1 σ_0 1.65) = $Y_C + 3.3\sigma_0$.

TrainMiC
Training in Metrology in Chemistry

Graphical representation of LOD



y_o is the mean of the blank measures
L_c is the decision level
LOD is the limit of detection

On this slide, the minimum single reply, with a stated probability which can be distinguished from a suitable blank value, is given. The limit of detection defines the point at which the analysis becomes possible and this may be different from the lower limit of the determinable analytical range.

By default, α and β are set to 5 %. If the distribution of the values is presumed to be Gaussian, and if the dispersion is presumed to be constant in the blank-LOD range, then LOD values are given by $LOD = y_0 + 3\sigma_0$.

TrainMiC
Training in Metrology in Chemistry

Control charts

- Control charts are used to track regular measurements of an ongoing process, and to signal when such a process has reached the point of going 'out of control' (i.e. may no longer be governed by the same properties, such as mean or standard deviation).
- The control chart has upper and lower control limits for sample statistics provided by successive samples taken over time.
- If control value meets certain criteria for being 'extreme' (e.g. too far away from the mean), this is a signal to investigate the process and determine whether anything is wrong.

 Statistics 2 - 2.2 Slide 37

The control chart is a graph used to study how a measurement process changes over time. Data are plotted in time order. A control chart always has a central line for the average, an upper line for the upper control limit and a lower line for the lower control limit.

TrainMiC
Training in Metrology in Chemistry

Chart details

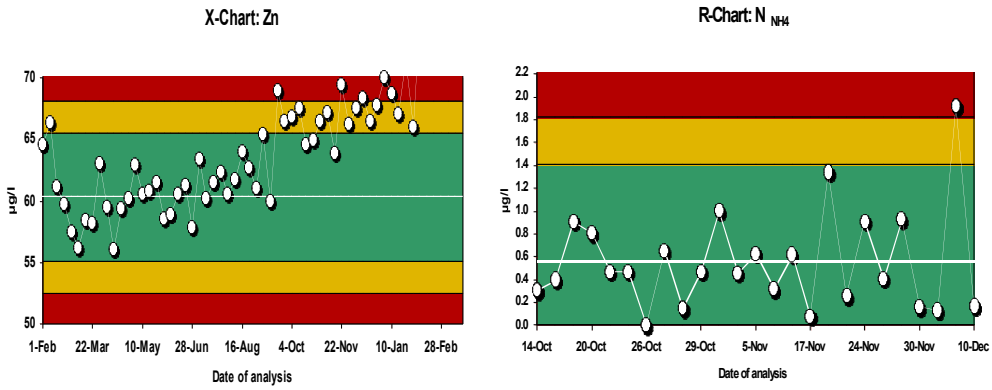
A control chart consists of:

- points representing a statistic (e.g. a mean, range, proportion) of measurements of a quality characteristic in samples taken from the process at different times;
- the mean of this statistic using all the samples is calculated;
- a centre line is drawn at the value of the mean of the statistic;
- the standard deviation (e.g. standard deviation/sqrt(*n*) for the mean) of the statistic is also calculated using all the samples;
- upper and lower control limits that indicate the threshold at which the process output is considered statistically 'unlikely' are drawn typically at 3σ from the centre line.

This slide explains how statistics is used to build a control chart.

TrainMiC
Training in Metrology in Chemistry

Examples



For more details, see Internal Quality presentation

This slide shows two examples of control charts: X-chart and R-chart.

- X-chart: in this chart, the sample means are plotted in order to control the mean value of a variable
- R-chart: in this chart, the sample ranges are plotted in order to control the variability of a variable.

TrainMiC
Training in Metrology in Chemistry

Analysis of variance

(ANOVA)

TrainMiC
Training in Metrology in Chemistry

Why use ANOVA

- ANOVA separates different sources of variation
- *Variation within batch*
- *Variation between batches*
- Within-batch and between batches estimates of variance can be compared
- ANOVA can be applied to any data, that can be grouped by a particular factor(s)
- ANOVA is used to compare sets of data

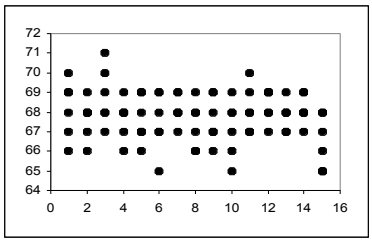
ANOVA (analysis of variance) is a powerful statistical technique which can be used to separate and estimate different causes of variation and to compare sets of data. Furthermore, the different sources of variation can be compared to determine if they are significantly different, under the assumption that the sampled populations are normally distributed.

There is 'a between-group variation' and 'a within-group variation'. The idea behind the analysis of variance is to compare the ratio of 'between-group variance' to 'within-group variance'.

ANOVA applies a statistical F-test to test the statistical significance of the differences among the obtained means of two or more random samples from a given population. It is assumed that the variances of the individual groups are similar (i.e. not statistically significantly different).

TrainMiC
Training in Metrology in Chemistry

Between and within-group variations



The variance is the mean of the squared deviations about the mean (*MS*) or the sum of the squared deviations about the mean (*SS*) divided by the degrees of freedom (*df*)

$$s^2 = V = MS = \frac{SS}{df} = \frac{\sum_{i=1}^{N}(x-\bar{x})^2}{N-1}$$

The null hypothesis is: there is no difference in the population means of the different levels of factor A.

The alternative hypothesis is: the means are statistically not the same.

Student's t-test can be used to compare the means of two sets of data. The t-test tells us if the variation between two groups is 'significant'.

ANOVA allows the comparison of multiple data sets. Multiple t-tests are not the answer because as the number of groups grows, the number of needed pair comparisons grows quickly. Also, doing multiple two-sample t-tests would result in a greatly increased chance of committing a Type I error.

Therefore, ANOVA has an advantage over a two-sample t-test.

Example: For seven groups, there are 21 pairs. If we test 21 pairs, we should not be surprised to observe things that happen only 5 % of the time. Thus, in 21 pairings, a $P = 0.05$ for one pair cannot be considered significant. ANOVA puts all the data into one number (F) and gives us one P for the null hypothesis.

TrainMiC
Training in Metrology in Chemistry

One-way ANOVA

- This is simplest type of analysis of variance used when there are equal numbers of observations (e.g. replicates, samples)
- When data can be grouped by a single factor
- Consider *p* different levels of a single factor (laboratory, sample, days) and suppose that *n* observations have been made at each level giving *N* total results (*N* = *pn*).
- The aim of the experiment is to determine if there are differences between the *p* levels

TrainMiC
Training in Metrology in Chemistry

One-way ANOVA

- The hypotheses are as follows:

  H_0: $M_1 = M_0$ there is no difference between the p levels

  H_1: $M_1 > M_0$ there is difference between the p levels

- To determine whether there is a significant difference among the means, the mean squares are compared using an F-test.
- $F_{critical}$ is obtained from tables of one-tailed F values at the appropriate level of significance (α) and (p – 1) and p(n – 1) degrees of freedom.

Recall that:

n is the number of observations at each level
p is the number of levels
$N = pn$ is the total number of observations
M is mean square values

TrainMiC
Training in Metrology in Chemistry

ANOVA results table

Source of Variation	Sum of squares	*df*	Mean square	*F*	*P*-value	$F_{critical}$
Between-group	S_1 = (i) − (iii)	$p - 1$	$M_1 = S_1/(p - 1)$	M_1/M_o		
Within-group	S_0 = (ii) − (i)	$N - p$	$M_o = S_o/(N - p)$			
Total	$S_1 + S_0$ = (ii) − (iii)	$N - 1$				

df is degrees of freedom
p is the number of groups of data (levels)
N is the total number of observations
P-value is the probability
$F_{critical}$ is the critical value for *F*

ANOVA calculations are usually done using software. A results table is shown in the slide and the next slide shows a results table in Excel.

The table shows that the variation in the data is divided into within-group and between-group components.

The mean square terms are variances which are calculated by dividing the sum of square terms by their associated degrees of freedom. The degrees of freedom for between groups is $(p - 1)$, whereas the total number of degrees of freedom is $(N - 1)$. The degrees of freedom for the within-group term is $(N - p)$. If each group contains n values the number of data points is $N=pn$. The degrees of freedom (df) for the within-group terms can be written as $p(n - 1)$ and the number of data points is $N = pn$.

TrainMiC
Training in Metrology in Chemistry

One-way ANOVA

$$(i) = \left(\sum_{i=1}^{p} \left(\sum_{k=1}^{n} x_{ik} \right)^2 \Big/ n \right) \qquad (ii) = \sum_{i=1}^{p} \sum_{k=1}^{n} x_{ik}^2$$

$$(iii) = \left(\sum_{i=1}^{p} \sum_{k=1}^{n} x_{ik} \right)^2 \Big/ N$$

The mean square values provide the components of variance attributable to the different level.
For the one-way analysis of variance these are:

$$M_o = \sigma^2_{wi}$$

$$M_1 = n\sigma^2_{be} + \sigma^2_{wi}$$

The mean square values M_0 and M_1 provide the components of variance attributable to the different levels.

TrainMiC
Training in Metrology in Chemistry

ANOVA results

- A one-tailed F-test compares mean square values
  H_0: $M_1 = M_0$

  H_1: $M_1 > M_0$

  The null hypothesis will apply if there is no variation, other than random.

- If $F > F_{critical}$ null hypothesis (H_0) is rejected.

  The between-group variance is significantly greater than the within-group variance and there are significant differences between the means of the data set.

- If the *P*-value is less than the significance level for the test

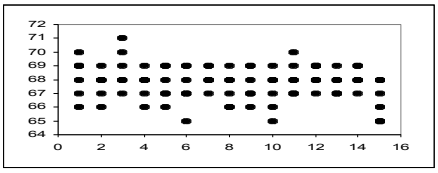
  (usually $\alpha = 0.05$), then the null hypothesis is rejected

In addition to F-tests, ANOVA can be also interrelated using *P*-values (probability).

TrainMiC
Training in Metrology in Chemistry

One-way ANOVA in Excel

	Replicates	1	2	3	4	5	6
Vials	1	66	68	67	69	70	69
	2	66	67	68	68	68	69
	3	71	67	68	69	68	70
	4	66	68	67	68	68	69
	5	67	67	66	69	69	68
	6	65	67	67	69	68	69
	7	67	68	68	68	69	69
	8	67	66	66	68	68	69
	9	67	67	66	69	68	69
	10	66	65	67	68	69	68
	11	67	67	69	68	68	70
	12	67	68	69	69	68	69
	13	67	67	68	69	68	68
	14	67	68	68	69	68	69
	15	65	66	65	68	68	67



SUMMARY

Groups	Count	Sum	Average	Variance
1	6	409	68.2	2.2
2	6	406	67.7	1.1
3	6	413	68.8	2.2
4	6	406	67.7	1.1
5	6	406	67.7	1.5
6	6	405	67.5	2.3
7	6	409	68.2	0.6
8	6	404	67.3	1.5
9	6	406	67.7	1.5
10	6	403	67.2	2.2
11	6	409	68.2	1.4
12	6	410	68.3	0.7
13	6	407	67.8	0.6
14	6	409	68.2	0.6
15	6	399	66.5	1.9

Sum of Squares SS

Mean Squares MS

ANOVA

Source of Variation	SS	df	MS	F	P-value	$F_{critical}$
Between-group	26.2	14	1.87	1.34	0.207	1.83
Within-group	104.8	75	1.40			
Total	131.0	89				

Repeatability stdev	s_r	1.18	**=sqrt(MSW)**
Between-group stdev	s_L	1.21	**=sqrt((MSB-MSW)/N)**
			(N replicates)

The simplest ANOVA method is one-way ANOVA, when there is one factor (e.g. analyst, temperature) either controlled or random, in addition to the random error in measurement.

ANOVA allows the sources of variation to be separated. By applying ANOVA, we can obtain s_r and s_R, repeatability and reproducibility standard deviations, respectively, used in interlaboratory comparison and homogeneity studies.

TrainMiC
Training in Metrology in Chemistry

Example — Sampling

Batch	S1A1	S1A2	S2A1	S2A2
B1	402	325	361	351
B2	382	319	349	362
B3	332	291	397	348
B4	280	278	358	321
B5	370	409	378	460
B6	344	318	381	392
B7	297	333	341	315
B8	336	320	292	306
B9	372	353	332	337
B10	407	361	322	382

S1 and S2: Primary samples from sampling location 1 and 2 of one production batch
A1 and A2: Analyses of duplicate test samples of a primary sample S
Analysed mean value (test sample 40 g): 348 µg g^{-1}

(Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide, *Measurement uncertainty arising from sampling: a guide to methods and approaches* (2007) [16].)

This slides shows the application of ANOVA for sampling. More details are available in the Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide [16].

TrainMiC
Training in Metrology in Chemistry

Example — Sampling

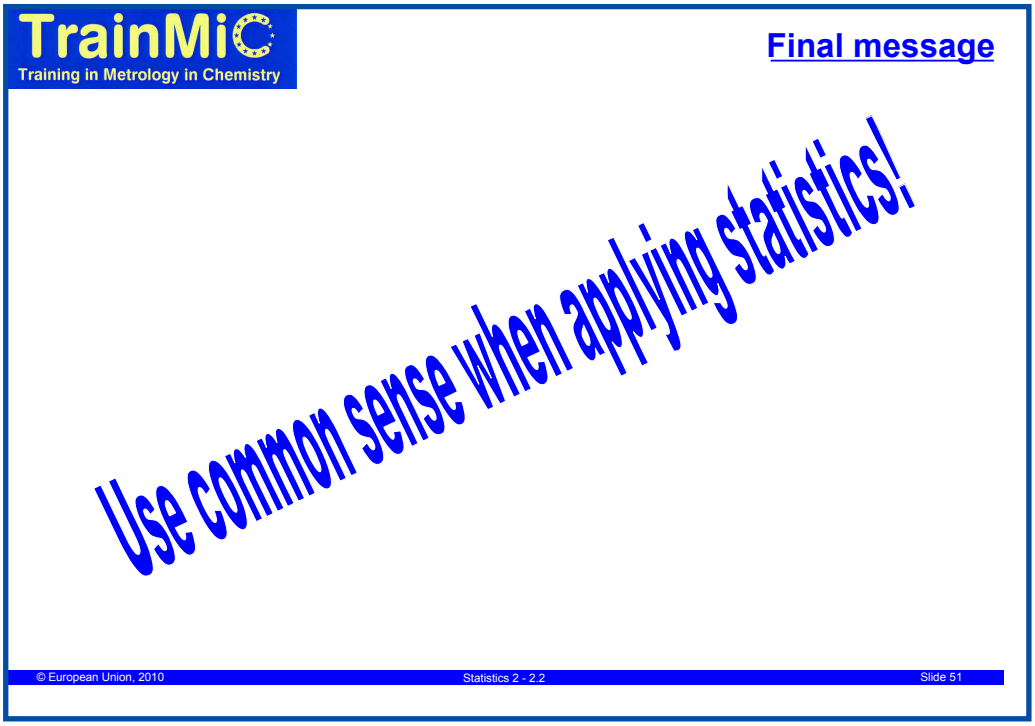
SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	10	3522	352.2	1786.4
Column 2	10	3307	330.7	1372.233
Column 3	10	3511	351.1	964.5444
Column 4	10	3574	357.4	2073.378

ANOVA

Source of variation	*SS*	*df*	*MS*	*F*	*P-value*	$F_{critical}$
Between-group	4148.1	3	1382.7	0.89256	0.454338	2.866266
Within-groups	55769	36	1549.139			
Total	59917.1	39				

The one-way output from Excel in this example shows that between-sample mean square is smaller than within-sample mean square and the result of the F-test shows that this difference is not significant.



Statistics is a very useful tool in helping to answer a number of questions. Nevertheless, statistics should always be applied with a critical view of the results, whether they make scientific sense.

References

[1] JCGM 200:2012 — *International vocabulary of metrology — Basic and general concepts and associated terms* (VIM), Joint Committee for Guides in Metrology (JCGM), 2008 (http://www.bipm.org), also published as ISO/IEC Guide 99: 2007, International Organisation for Standardisation (ISO)/International Electrotechnical Commission (IEC), Geneva, 2007.

[2] Paul De Bievre, Helmut Gunzler (eds), *Measurement Uncertainty in Chemical Analysis*, Springer, 2010, ISBN 3642078842.

[3] DIN 3841— *German standard methods for the examination of water, waste water and sludge — Sludge and sediments (group S)*.

[4] EN ISO/IEC 17025:2005 — *General requirements for the competence of testing and calibration laboratories*, International Organisation for Standardisation (ISO)/ International Electrotechnical Commission (IEC), Geneva, 2005.

[5] JCGM 100:2008 — Corrigendum: 2010 — *Evaluation of measurement data — Guide to the expression of uncertainty in measurement* (GUM), Joint Committee for Guides in Metrology (JCGM), 2008 (http://www.bipm.org), also published as ISO/IEC Guide 98-3:2008, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement* (GUM:1995), International Organisation for Standardisation (ISO)/International Electrotechnical Commission (IEC), Geneva, 2008.

[6] Eurachem/CITAC Guide CG4, *Quantifying Uncertainty in Analytical Measurement*, Second edition, 2000 (http://www.eurachem.org).

[7] Bertil Magnusson, Teemu Näykki, Håvard Hovind, Mikael Krysell, Nordtest TR537 guide, *Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories*, Edition 2, Nordisk InnovationsCenter, 2004 (http://www.nordicinnovation.net/nordtest.cfm).

[8] EUROLAB Technical Report No 1/2007, *Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation*, European Federation of National Associations of Measurement, Testing and Analytical Laboratories (EUROLAB), 2007 (http://www.eurolab.org).

[9] JCGM 101:2008 — *Evaluation of measurement data — Supplement 1 to the 'Guide to the expression of uncertainty in measurement' — Propagation of distributions using a Monte Carlo method*, First edition 2008, Joint Committee for Guides in Metrology (JCGM), 2008 (http://www.bipm.org).

[10] Eurachem/CITAC Guide, *Use of uncertainty information in compliance assessment*, First edition, 2007 (http://www.eurachem.org).

[11] Eurachem/CITAC Information leaflet, *Use of uncertainty information in compliance assessment*, 2009 (http://www.eurachem.org).

[12] Bettencourt da Silva, R., Camões, M. Filomena, 'The Quality of Standards in Least Squares Calibrations', *Analytical Letters*, 1532-236X, Volume 43, Issue 7 and 8, 2010, pp. 1257–1266.

[13] ISO 5725:1994 (Parts 1–4 and 6) — *Accuracy (trueness and precision) of measurement methods and results*, International Organisation for Standardisation (ISO), Geneva, 1994 (see also ISO 5725-5:1998 for alternative methods of estimating precision).

[14] ISO 21748:2010 — *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation, International Organisation for Standardisation* (ISO), Geneva, 2010.

[15] EN ISO 11732:2005, 5 — *Water quality — Determination of ammonium nitrogen — Method by flow analysis* (CFA and FIA) and spectrometric detection, International Organisation for Standardisation (ISO), Geneva, 2005.

[16] Ramsey, M. H., Ellison, S. L. T. (eds), Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide, *Measurement uncertainty arising from sampling: a guide to methods and approaches*, Eurachem (2007), ISBN 978-0-948926-26-6 (available from the Eurachem secretariat).

[17] Kragten, J., 'Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique', *Analyst*, 119, 1994, pp. 2161–2166.

[18] ISO 7150-1:1984 — *Water quality — Determination of ammonium — Part 1: Manual spectrometric method*, International Organisation for Standardisation (ISO), Geneva, 1984.

[19] International Union of Pure and Applied Chemistry, 'Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995)', *Pure and Applied Chemistry*, Vol. 67, No 10, 1995, pp. 1699–1723.

Further reading

APLAC TC 005, 'Interpretation and guidance on the estimation of uncertainty of measurement', *Testing*, Issue No 4, Asia Pacific Laboratory Accreditation Cooperation, 2010 (http://www.aplac.org).

EA-4/02 M: 1999 — *Expression of the Uncertainty in Measurement in Calibration*, European co-operation for Accreditation, 1999 (http://www.european-accreditation.org).

EA-4/16 G:2003 — EA *guidelines on the expression of measurement uncertainty in quantitative testing*, European co-operation for Accreditation, 2003 (http://www.european-accreditation.org).

EUROLAB Technical Report No 1/2002, *Measurement Uncertainty in Testing — A short introduction on how to characterise accuracy and reliability of results including a list of useful references*, European Federation of National Associations of Measurement, Testing and Analytical Laboratories (EUROLAB), 2002 (http://www.eurolab.org).

EUROLAB Technical Report No 1/2006, *Guide to the Evaluation of Measurement Uncertainty for Quantitative Tests Results*, European Federation of National Associations of Measurement, Testing and Analytical Laboratories (EUROLAB), 2006 (http://www.eurolab.org).

ILAC G17:2002, *Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard* ISO/IEC 17025, International Laboratory Accreditation Cooperation (ILAC) (http://www.ilac.org).

ISO 3534-1:2006 — *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*, International Organisation for Standardisation (ISO), Geneva.

ISO 3534-2:2006 — *Statistics — Vocabulary and symbols — Part 2: Applied statistics*, International Organisation for Standardisation (ISO), Geneva.

ISO 3534-3:1999 — *Statistics — Vocabulary and symbols — Part 3: Design of experiments*, International Organisation for Standardisation (ISO), Geneva.

ISO 11843-5:2008 — *Capability of detection — Part 5: Methodology in the linear and non-linear calibration cases*, International Organisation for Standardisation(ISO), Geneva.

ISO 80000-1:2009 — *Quantities and units — Part 1: General*, International Organisation for Standardisation (ISO), Geneva.

ISO/TR 22971:2005 — *Accuracy (trueness and precision) of measurement methods and results — Practical guidance for the use of* ISO 5725-2:1994 *in designing, implementing and statistically analysing interlaboratory repeatability and reproducibility results*, International Organisation for Standardisation (ISO), Geneva.

ISO/TS 21749:2005 — *Measurement uncertainty for metrological applications — Repeated measurements and nested experiments*, International Organisation for Standardisation (ISO), Geneva.

ISO 13528:2005 — *Statistical methods for use in proficiency testing by interlaboratory comparisons*, International Organisation for Standardisation (ISO), Geneva.

IUPAC, Periodic Table of the Elements, 2011: for updates to this table, see online (http://www.iupac.org/highlights/periodic-table-of-the-elements.html).

IUPAC, *Compendium of Chemical Terminology* (Gold Book), 2007 (http://goldbook.iupac.org).

IUPAC, *Quantities, Units and Symbols in Physical Chemistry* (Green Book), 2007 (http://www.bipm.org).

Majcen, N., Taylor, P. (eds), *Practical examples on traceability, measurement uncertainty and validation in chemistry*, Volume 1, Second edition, Publications Office of the European Union, 2010, ISBN 978-92-79-12021-3.

Miller, J. N., Miller, J. C., *Statistics and Chemometrics for Analytical Chemistry*, Fifth edition, Pearson Education Limited, 2005, ISBN 0-13-129192-0.

NIST/SEMATECH, e-Handbook of Statistical Methods, 8.12/2008 (http://www.itl.nist.gov/div898/handbook/).

Olivieri, A. C., Faber, N. M., Ferré, J., Boqué, R., Kalivas, J. H., Mark, H., 'Uncertainty estimation and figures of merit for multivariate calibration', IUPAC Technical Report, *Pure and Applied Chemistry*, 2006, 78(3), pp. 633–661 (http://www.iupac.org).

Vetter, T. W., National Institute of Standards and Technology (NIST), 'Quantifying measurement uncertainty in analytical chemistry — A simplified practical approach (Kragten)', *Proceedings of Measurement Science Conference* 2001, Session V-B, Anaheim, 2001 (http://www.p2pays.org/ref/18/17628.pdf).

Index

A

alternative hypothesis, *159*, 221
analyte, 26, 31, 51, 138, 203, 207, 210, 211, 213, 214
analytical procedure, 154, 165, 168, 210
analytical sciences, *22*
ANOVA, 179, 182, 219, 220, 221, 224, 226, 227, 228
applied statistics, 179, 181

B

bias, 11, 32, 65, 70, 72, 73, 109, 112, 113, 116, 117, 119, 120, 121, 154, 158, 210
blank sample, 99, 210

C

calibration, 74, 76, 81, 98, 100, 179, 191, 193, 194, 202, 207, 213
calibration standard, *75*
calibration standards, *74*, *80*
certified reference material, 70, 113, 115
certified value, 70, 77, 156, 158
chemistry, 40, 131, 213
combined standard uncertainty, *59*, *104*, *105*, *117*, *143*
combined uncertainty, 113, 129, 145
concentration, 26, 74, 75, 81, 87, 88, 90, 91, 97, 98, 100, 105, 108, 111, 114, 116, 117, 121, 138, 152, 185, 189, 191, 203, 207, 210, 211, 213
confidence, 27, 28, 48, 58, 62, 147, 149, 152, 156, 158, 162, 163, 169, 177, 201
confidence level, *59*, *79*
confidence limit, *199*
contribution, 55, 80, 105
control chart, 216, 217
control limit, *216*
correction, 35, 70, 73
correlated variables, 51, 52
correlation, 103, 143, 179, 182, 186, 193, 194
covariance, 192

D

definition, 22, 26, 29, 31, 32, 196
definition of a measurand, *30*, *31*, *42*
degrees of freedom, 59, 140, 152, 153, 163, 167, 169, 177, 199, 214, 224
dependent variable, 184, 185, 202
designate, 35
directive, 125
distribution, 129, 135, 136, 137, 141, 148, 149, 150, 151, 152, 153, 156, 167, 172, 185, 210, 215

E

empirical approach, *22*
Eurachem, 40, 74, 90
EUROLAB, 40, 86, 93, 123
expanded uncertainty, 58, 60, 61, 63, 81, 125, 143, 177

F

F-test, 154, 168, 169, 220, 226, 229

G

Grubbs' test, 172, 173
guides, 40
GUM, 19, 39, 40, 47, 61, 65, 85, 129, 147, 177, 178

H

histogram, 135, 136
homoscedasticity, 202

I

independent variable, 184, 185, 198
influence quantities, *51*
influence variables, *45*
input quantity, *48*, *51*, *52*, *53*, *54*, *56*, *57*, *74*, *100*, *143*
intended use, 29, 33, 34, 39, 43, 71
interlaboratory data, *22*
interlaboratory validation approach, 65, 66
intermediate precision, 65, 71, 77, 92, 128

internal quality control, 110
interpretation of results, 35
intra-laboratory data, *22*
ion chromatography, 67
ISO/IEC 17025, *19*, *35*
ISO 17043, *93*
ISO 13528, *93*
ISO 5725, *93*, *119*, *121*, *122*
ISO 21748, *93*, *121*, *122*
ISO 11732, *95*, *97*, *121*, *126*
ISO 7150, *126*

L

least squares regression, 74, 189
least squares, 190, 191, 207
legal limit, *164*
legislation, 35
limit of detection, *179*, 182, *211*, *213*, 214, *215*
limit of quantification, *71*, *91*, *95*, *98*

M

maximum limit, *35*
mean difference, *167*
mean value, *137*, *139*, *141*, *149*, *152*, *153*, *155*, *158*, *160*, *162*, *164*, *165*, *168*, *218*
measurand, 27, 28, 29, 30, 31, 32, 42, 43, 45, 69, 72, 96, 144, 199, 207
Measurement function, 41
measurement procedure, 19, 32, *33*, 34, *43*, *66*, *71*, *72*, 85
measurement quality, *44*, *59*
measurement uncertainty, 19, *20*, *21*, *22*, *23*, *26*, *27*, *28*, *29*, *30*, *32*, 33, 34, *35*, 36, *38*, *39*, *40*, *41*, *42*, *44*, *45*, *49*, *62*, *65*, *66*, *69*, *71*, *72*, *80*, 85, *86*, *87*, *89*, *92*, *94*, *99*, *106*, *114*, *121*, *127*, *128*, *131*, *142*, *146*
modelling approach, *19*, *22*, *65*, *66*, *81*, *85*, *92*, *94*, *106*, *118*, *128*

N

Nordtest, 40, 109
normal distribution, 136, *137*, *141*, *153*
null hypothesis, 154, 157, 159, 160, 162, 164, 165, 167, 169, 209, 211, 221

O

one-sided test, 156, 168
optimisation, 66, 80
outlier, 172, 173
output quantity, *54*, *57*

P

partial derivative, 55
population, 129, 136, 138, 139, 141, 153, 164, 165, 166, 168, 173, 202, 209, 220, 221
precision, 45, 52, 77, 92, 108, 155, 168, 210
principles, 23, 38, 39, 40, 65, 67
probability, 27, 28, 137, 148, 154, 156, 158, 169, 185, 210, 215
proficiency testing, 93, 115, 116, 123, 126
propagation, 48, 51, 52, 53, 55, 56, 57, 78, 129, 143

Q

quality, 29, 77, 88, 92, 109, 110, 111, 118, 126, 128, 131
quantity, 26, 28, 29, 42, 45, 53, 214

R

recovery, 35, 73, 77
regression, 74, 76, 98, 179, 182, 184, 185, 188, 189, 190, 191, 193, 196, 197, 198, 199, 202, 203, 207, 213
regression line, 184, 196, 197, 199, 202
regulation, 35, 175
relative standard deviation, *139*
relative standard uncertainty, *57*, *65*, *75*, *77*
repeatability, 71, 80, 92, 109, 120, 121, 122, 126, 128, 227

reproducibility, 65, 66, 92, 109, 110, 111, 121, 126, 128, 227
residuals, 179, 190, 191, 196, 198, 202
residual test, 172

S

sample, 26, 30, 31, 33, 35, 45, 62, 64, 65, 73, 76, 81, 109, 117, 121, 122, 135, 138, 139, 141, 153, 166, 172, 207, 209, 210, 213, 214, 218, 221, 229
sample mean, *173*
sampling, 30, 35, 122, 152, 228
significant figures, 175, 176, 177, 194
single laboratory validation approach, 65
slope, 55, 103, 189, 190, 194, 195, 199
sources of uncertainty, *43*, *44*, *45*, *46*, *55*, *65*, *74*, *100*, *101*, *108*
standard, 33, *35*, 140
standard deviation, 48, 77, *92*, *99*, *109*, 121, *122*, *129*, *137*, *138*, 139, *141*, *147*, *148*, *150*, *152*, *153*, *155*, *164*, *165*, 166, *167*, *168*, *173*, *177*, *199*, *202*, *203*, *210*, *213*, *214*, *227*
standard deviation of the mean, *139*, *140*
standard measurement uncertainty, *42*
standard method, *93*, *128*
standard procedure, *168*
standard uncertainty, *48*, *53*, *55*, *58*, *59*, *63*, *79*, *99*, *143*, *145*, *146*, *147*, *151*, *152*
statistical parameters, 133, 139, 140

T

target measurement uncertainty, *43*
target uncertainty, 81, 94, 96, 125, 127
traceability, 32, 33, 70, 72
triangular distribution, 150
true mean value, *138*
true value, 26, 27, 28, 29, 48, 115, 137, 152, 160
trueness, 71, 92, 108, 112, 121
t-test, 154, 157, 164, 167, 168, 221
two-sided test, 155, 156, 168, 173
Type, 146, 163, 209, 211
Type A evaluations, *46*
Type B evaluations, *47*

U

uncertainty budget, *30*, *34*, *49*, *60*, *80*, *81*, *105*
uncertainty components, *26*, *27*, *29*, *43*, *45*, *46*, *48*, *49*, *50*, *59*, *65*, *66*, *74*, *75*, *78*, *80*, *81*, *92*, *99*, *102*, *128*
uncertainty, evaluation of, *19*
uncertainty, law of propagation of, *50*
upper limit, *43*

V

validation, 19, 32, *33*, 65, 66, *71*, 85, 88, 92, 93, 107, 109, 110, 118, 119, 126, 127, 128, 131, 154, 179
variance, 139, 219

W

warning limit, *111*

X

X-chart, 111, 218

Summary

TrainMiC® is a European programme for lifelong learning on how to interpret the metrological requirements in chemistry. It is operational across many parts of Europe via national teams. These teams use shareware pedagogic tools which have been harmonised at European level through the joint effort of many experts across Europe working as an editorial board. The material has been translated into 14 different languages.

This report includes four TrainMiC® presentations:

1. Uncertainty of measurement — Part I Principles;
2. Uncertainty of measurement — Part II Approaches to evaluation;
3. Statistics for analytical chemistry — Part I; and
4. Statistics for analytical chemistry — Part II.

LA-NA-25207-EN-N

The mission of the Joint Research Centre (JRC) is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of European Union (EU) policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policymaking process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

Price (excluding VAT)
in Luxembourg: EUR 25

Publications Office

ISBN 978-92-79-23071-4

doi:10.2787/58527