Recent Development of the Top-Down Approaches in the Evaluation of Measurement Uncertainty for Testing Laboratories

by

Yeoh Guan Huah

http://consultglp.com

Overview

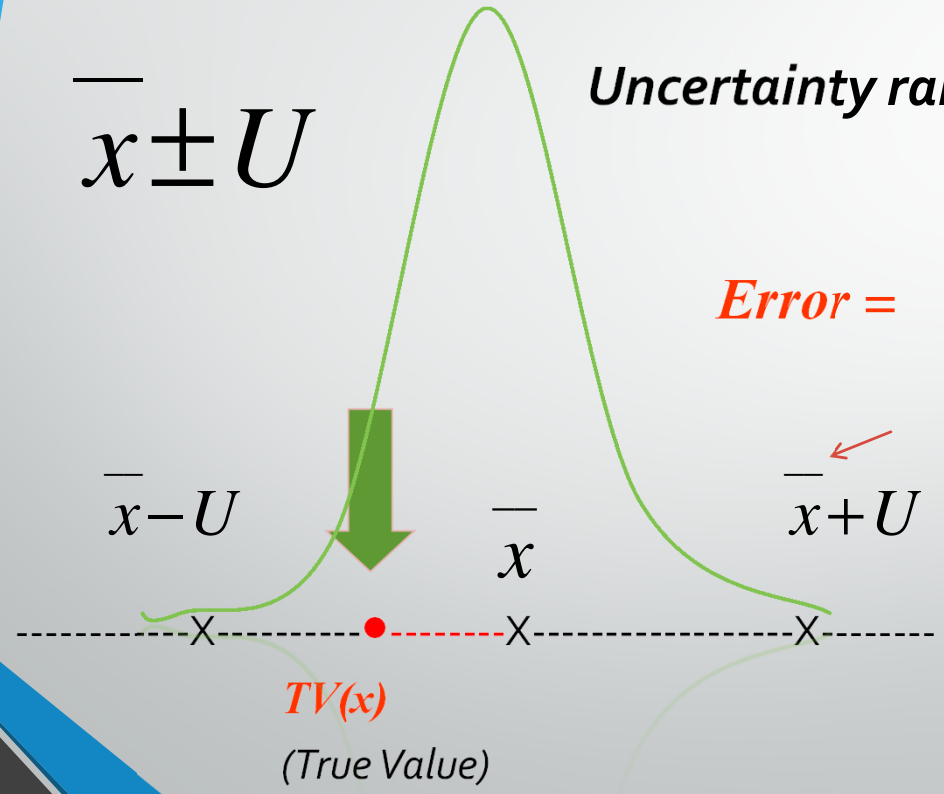
- Current MU estimation approaches
- Strengths and weaknesses of GUM (bottom-up) method
- Introducing the concept of top-down methods
- Strengths and weaknesses of top-down methods
- Some detailed discussions on top-down approaches:
  - Precision, accuracy and trueness;
  - Control chart
- Brief discussion on the other top-down approaches:
  - ISO 11352;
  - linear calibration curve of reference materials;
  - Horwitz's equation

Introduction

- There is always an element of error in all measurements
- Three types of error:
  - Gross error
  - Systematic error (bias)
  - Random error
- Systematic error is to be minimized or corrected
- Standard deviation of repeated analysis estimates random error
- Evaluation of measurement uncertainty has become an integral part of laboratory accreditation standards (ISO/IEC 17025:2005)

What is Measurement Uncertainty?

$\overline{x} \pm U$



Uncertainty range : $\overline{x} - U$ ..... $\overline{x} + U$

Error = $\overline{x} - TV$

Measurement Uncertainty ≠ Error

Current MU estimation methods

- **"Bottom-up" method** :
  - GUM
- **"Top-down" methods** :
  - Using precision (repeatability), accuracy (reproducibility) and trueness (no bias)
  - Using quality control chart
  - Linear calibration of reference QC materials of different concentrations (constant and proportional Std Deviation)
  - Horwitz's equation
- **Monte Carlo method** – an alternative to GUM

Common approach – GUM (Bottom-up) method

- Reference : "**G**uide to the Expression of **U**ncertainty of **M**easurement" - ISO/IEC Guide 98
- Use of a mathematical or statistical model to describe the linear functional relationship between the analyte and the influencing factors (uncertainty components or budgets).
- A test procedure, y involves many steps and each step can have a standard uncertainty, expressed as standard deviation, say x_i :

$$y = f(x_1, x_2, ..., x_n)$$

- The combined or total std uncertainty of ***independent*** components is:

$$u(y)^2 = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i} \right]^2 u(x_i)^2$$

"Law of propagation of uncertainty"

Propagation Law of Standard Uncertainty (Standard Deviation)

- If these uncertainties are NOT independent, there is an extra **covariance** factor to be considered:

$$u^2(y) = \sum \left(\frac{\partial f}{\partial x_i} \right)^2 u(x_i)^2 + \sum \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} Cov(x_i, x_j)$$

$$u_c(y) = \sqrt{\sum_{i=1}^{N} \left[\frac{\partial f}{\partial x_i} \right]^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} r(x_i, x_j) u(x_i) u(x_j)}$$

Strengths and weaknesses of GUM method

- Critically assesses the test method for uncertainty contributors,
- Consistent with other fields of measurements,
- Uncertainty estimated is relevant specifically to individual laboratory,
- Good for new method development
- ***but***,
- Is tedious and time consuming in MU evaluation process
- Assumes unrealistically that certain errors are random and/or independent;
- Ignores method bias by assuming all systematic errors have been corrected for;
- Does not apply well without a mathematical model.

Holistic top-down approaches

- Looking at the overall performance of the test method
- The performance of a test method is judged by its method precision, accuracy and trueness (*no bias*)
- It follows the basic principle of GUM, using law of propagation of uncertainty (standard deviation)
- Any accredited laboratory should already have a robust QA and QC system in place
- All accredited tests should have routinely been collecting many QC data for disposal
- Top-down approaches are getting more popular amongst testing laboratories in Europe and the North America.

Holistic top-down approaches

- Only applied under the condition of strict quality control and stable test methods
- Each accredited test method should have the followings established:
  - Repeatability, in terms of s_r
  - Intermediate precision (or intermediate reproducibility) within lab, in terms of $s_{R'}$
  - Reproducibility between labs, in terms of s_R

"Top-down" methods

- **ISO 21748** – "*Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation*"
- **ASTM D6299** – "*Applying statistical quality assurance and control charting techniques to evaluate analytical measurement system*"
- **ISO 11352 : 2012** – *"Water quality – Estimation of measurement uncertainty based on validation and quality control data"*
- **ISO 11095** – "*Linear calibration using reference materials*"

"Top-down" publications

- Using the **Horwitz's equation**
- **EuroLab Technical Report No. 1/2006** – "*Guide to the evaluation of measurement uncertainty for quantitative test results*"
- **Eurachem / CITAC Guide CG4 (3rd edition, 2012)** – "*Quantifying Uncertainty in Analytical Measurement*"
- **Nordtest NT TR 537 (ver 3.1, 2012)** – "*Handbook for calculation of measurement uncertainty in environmental laboratories*"

Strengths and weaknesses of Top-down approaches

- It is simpler in MU estimation process
- Assesses overall performance of test method by the laboratory concerned
- ***but:***
- Cannot by itself identify where the major errors could be occurring in the analytical process;
- That reproducibility R data of a test method may not be representative for variability of results, unless it has been standardized.

Top-down 1 – Precision, Accuracy & Trueness – ISO 21748

To confirm the lab measurement system is under control; No systematic error (bias); Having collected QC data and analyzed samples for precision evaluation over a prolonged period of time

Successful participation in PT or interlab crosschecks or having used standard or primary classical methods for comparison

Using the R outcome of PT, corss-check or standard/classical methods or experiences to estimate MU, based on precision and reproducibility data

Precision, accuracy and trueness – ISO 21748

- The lab has implemented regular QC protocols to monitor the test method performance.
- Consider the repeatability r and reproducibility R of the standard method or PT programs, and confirm the lab's ability to comply by applying analysis of variance (ANOVA).
- Pre-requisite : The lab has to confirm its analytical data are not bias. If there is a bias, the lab must find out the root cause and correct the systematic error.

Precision, accuracy and trueness – ISO 21748

- If $y = f(x_1, x_2, \ldots, x_n)$
- The combined std uncertainty $u(y)$ of observed y value is :

$$u(y) = \sqrt{u^2(\hat{\delta}) + s_R^2 + \sum c_i^2 u^2(x_i)}$$

δ（Method fixed bias, if any, as variance）,obtained from analysis of standard materials

Intermediate precision variance（$s^2{}_R$）using

$$s_R^2 = s_L^2 + s_r^2$$

Inter-lab variance

Lab random error variance

Any other variances considered

Precision, accuracy and trueness – ISO 21748

- This ISO method states that if the following consideration is satisfied, there is no data bias:

$$|\Delta| < 2 s_D$$

Deviation estimated from actual analysis of standard material, PT program, or primary methods

Standard deviation established from standard material, PT program or primary method

Precision, accuracy and trueness – ISO 21748

- Example: using repeated analysis over a period of time on a standard reference material to check for bias:

$$|\Delta_l| = |\overline{y} - RQV|$$

Mean value of repeated analysis yi (i=1…n)

Reference Quantitative Value

$$s_D = \sqrt{s_L^2 + \frac{s_W^2}{n_l}}$$

Lab own std deviation, by n repeated analysis

Inter-lab std deviation

Precision, accuracy and trueness – ISO 21748

- Example: Using PT results to check for bias:

$$\left|\overline{\Delta}_y\right| = \frac{\sum_{i=1}^{q} (\hat{y}_i - y_i)}{q}$$

where:

$\overline{\Delta}_y$ —— Lab's average std deviation from PT program

$\hat{y}_i$ ——Lab's own results , $\hat{y}_1$, $\hat{y}_2$, ……, $\hat{y}_q$;

y_i —— Consensus values from PT programs, y_1, y_2, ……, y_q ($q \geq 1$) ;

q ——Number of PT programs participated

Precision, accuracy and trueness – ISO 21748

- The s_D of PT program for bias is :

$$s_D = \sqrt{s_L^2 + \frac{s^2(\mathrm{V}_y)}{q}}$$

After taking *q* number of PT programs, the variance of average results $\hat{y}_i$ and the consensus value y_i

Precision, accuracy and trueness – ISO 21748

- Under controlled testing environment with the laboratory intermediate precision sR' proven to be not significantly larger than the sr provided by the PT program(s), the standard uncertainty of the test method as performed by the laboratory is therefore:

$$u \approx s_R$$

- where

  s_R is the standard uncertainty of reproducibility R provided by the PT program(s)

Expanded Uncertainty U = 2 x s_R

Important notes for ISO 21748

- Although this top-down approach appears much simpler in calculation, the following rules have to be complied:
- 1) Test method must have been used over an extended period of time, involving several operators and equipment, ensuring intermediate reproducibility
- 2) The associated QA/QC data, capturing all significant contributing factors are up-to-date and the data collection is continuing with 'moving average'
- 3) To participate in recognized relevant proficiency programs with good number of participating laboratories and achieve satisfactory Z-scores, before using the reported reproducibility R data.

Top-down 2 : Control chart method ASTM D6299

- **Control chart** : Using the control chart data to estimate measurement uncertainty

AD Test
- To test the normal distribution of data
- To check the independence of data

Set up QC chart
- Observing data trend
- To check if QC data are out of control

Estimating MU
- Intermediate precision data is the standard uncertainty, u

Top-down : Control chart method ASTM D6299

- On the routine QC data collected from stable QC/LCS samples, conduct an Anderson-Darling statistic test (AD*) to check their data randomness (normality) and independence。
- When the data normality and independence are confirmed, set up their QC data chart and moving range MR chart to visually check their data trend
- Once the data trueness and analysis protocols are under control, the lab can use the standard deviation of moving range $S_{R'}$ as the standard uncertainty u 。

Using Anderson-Darling AD statistic test for QC data normality and independence

Anderson Darling statistic test

- Calculation formulae:

$$AD = -\frac{\sum_{i=1}^{n}(2i-1)\left[\ln(p_i)+\ln(1-p_{n+1-i})\right]}{n} - n$$

$$AD^{*} = AD^{2}\left(1+\frac{0.75}{n}+\frac{2.25}{n^{2}}\right)$$

where:	
n = No. of data points	
pi = Probability	

AD^{*}— is the corrected value of AD. It can be estimated from standard deviation s to give AD_{s} * or moving range MR to give AD_{MR} *

Interpretations of AD_s^* and AD_{MR}^*

- a) $AD_s^* < 1.0$ and $AD_{MR}^* < 1.0$: Accept the fact that the QC data are normal and independent, and use s or MR to set up the QC chart

- B) $AD_s^* > 1.0$ and $AD_{MR}^* > 1.0$, indicating QC data have lost control;

- c) $AD_s^* < 1.0$ and $AD_{MR}^* > 1.0$, indicating the QC data are randomly distributed (normal) but not fully independent.

Preparing the quality control chart

- To confirm there is no outlier in the set of data, x_i
- To arrange data x_i $(i=1\ldots n)$ ascending: $x_1 \le x_2 \le \ldots\ldots x_n$
- To use the following normalization formula to (w_i) :

$$w_i = \frac{x_i - \bar{x}}{s_i}$$

Convert w_i value to its normal probability pi value

- Moving Range (MR) : $MR_i = \left| x_{i+1} - x_i \right|$
- Std deviation equation (s) :

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

Calculate s_i and s_{MR}

Plotting QC chart

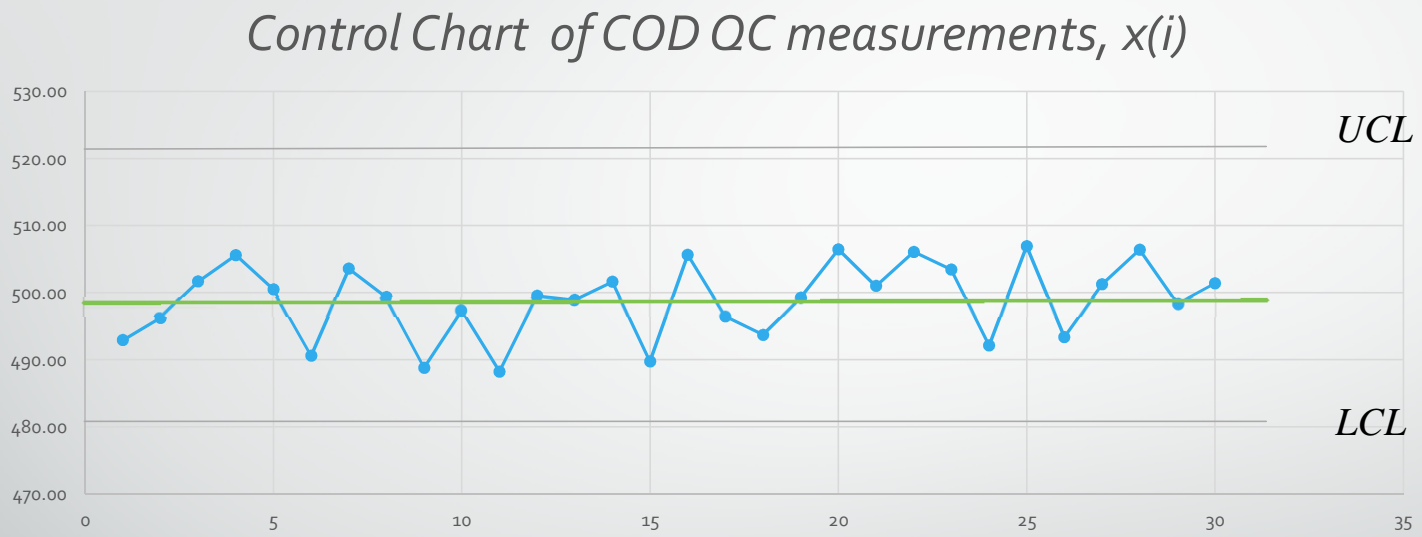
- Consider no less than 20 QC data which have been statistically tested by Anderson-Darling test， and use standard deviation $s_{R'}$ to set up a control chart :

$$UCL = \bar{x} + 3s_{R'}$$

$$LCL = \bar{x} - 3s_{R'}$$

- where：
- UCL and LCL are upper and lower control limits, respectively
- $$s_{R'} = S_{MR} = \frac{|\overline{MR}|}{1.128}$$

Plotting the QC chart



One may either visually examine the data trend or use Student's *t*-test to evaluate deviations 。

Student's *t*-test for data deviation trend

- Under controlled intermediate precision condition， use no less than 20 data (x_i) with reference value RQV to plot the control chart and calculate 2-tailed *t*-tests.

$$t = \frac{\sqrt{n}\left|\bar{x} - RQV_i\right|}{s_{x_i}} \qquad t_{MR} = \frac{\sqrt{n}\left|\bar{x} - RQV_i\right|}{\overline{MR}/1.128}$$

- Set α = 0.05, compare the *t* value against $t_{critical\ value}$， and also t_{MR} value with a $t_{critical\ value}$ at $(n-1)/2$ degrees of freedom

- If $t_{MR} \leq t_{critical\ value}$， it indicates that the collated data in the analysis process have no significant deviation from the reference value

- If not, one has to examine the root cause of data deviation.

Under the quality control conditions, the intermediate precision = the standard uncertainty of the test method in the laboratory

$$s_{R'} = \overline{MR} \,/\, d2 = \overline{MR} / 1.128$$

Expanded Uncertainty U = 2 x $s_{R'}$

The constant factor $d2$ = 1.128 is given by ISO 8258

Top-down 3 – Use of validation and internal / external quality control data (ISO 11352)

- It is similar to ASTM D6299 using the control chart method to estimate measurement uncertainty, *but ….*
- examines any laboratory and method bias ub
- considers different forms of quality control sample for random errors :
  - Stable control sample - use sR' from control chart
  - Stable synthetic control sample – use sR' from control chart + matrix uncertainty
  - Unstable synthetic control sample – use sR' from control chart + uncertainty between batches

Top-down 4 – Linear calibration of reference standards (constant and proportional std deviation) (ISO 11095)

Std working calibration

- Using different concentration *RQV's*
- Perform **ANOVA**

ANOVA to check:

- Pure regression error<repeatability error
- Validity of linear working curve

MU evaluation

- Based on low and high concentration stds.
- Using ANOVA technique to estimate MU

Top-down 5 method – Experience model

- Mainly based on Horwitz's equation

$$\sigma_H = 0.02c^{0.8495}$$

- where, σ_H : largest uncertainty of measurements with 95% confidence
- c : test data
- This equation is derived from RSD_R (reproducibility) values of numerous PT programs in water and metal analysis in the past many years
- It is more for academic interest and is used as a benchmark for the performance of an analytical method
- One may derive his own equation upon research into various PT program results

Conclusions

- It is obvious that the top-down estimation methods are less tedious and cover the overall performance of the laboratory and method uncertainty
- With some training, the top-down approaches are not that difficult to be adopted because all accredited laboratories should already have a robust laboratory quality system in place. It is a matter of practice to collate and study the QC data (excluding outliers) at regular intervals.
- The GUM method has its useful roles during the new method development process.
- Monte Carlo simulation is an alternative to GUM using computer to generate data under specified distribution for each uncertainty component.

Thank you for your kind attention!

guanhuah.yeoh@alsglobal.com

Welcome to visit my blog site and comment: *http://consultglp.com*