Universität Stuttgart

Method validation

Dr.-Ing. Michael Koch
Institute for Sanitary Engineering, Water Quality and Solid Waste Management
Universität Stuttgart
Dep. Hydrochemistry
Bandtaele 2, 70569 Stuttgart
Germany
Tel.: +49 711 685 65444 / Fax: +49 711 685 67809
E-Mail: Michael.Koch@iswa.uni-stuttgart.de

Universität Stuttgart

Glossary - Validation

- Definition
  - Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled (EN ISO 9000)
- Description
  - The validation shows with the help of laboratory experiments, that the corresponding parameters of a method fulfil the requirements of the intended chemical analytical application
  - Relevant chemical analytical parameters:
    - Precision
    - Trueness
    - Limit of detection
    - Limit of determination
    - Selectivity
    - Linearity
    - Robustness

Why is Method Validation Necessary?

- Very simple

To prove that the method is fit for purpose

The Professional Duty of the Analytical Chemist

- To increase reliability of laboratory results
- To increase trust of laboratory customers
- To prove the truth

When should Methods be Validated

- New method development
- Revision of established methods
- When established methods are used in different laboratories/different analysts etc.
- QC indicates method changes
- Comparison of methods
- if the lab fails in a PT and the problem could not be found

Validation of Standard Methods?

- Standard methods can be assumed as validated to a basic degree
- I.e., one can assume that the method is suitable for the scope mentioned in the standard
- The laboratory has to verify that it can reach the precision, trueness and other parameters described in the standard

Validation of In-house Methods

- In-house methods or the use of standard methods outside the scope of the standard require a complete validation
- I.e., all method characteristics have to be determined and compared with the requirements of the intended purpose

Determination of Method Characteristics

- Some of the method characteristics of the basic method (determination of calibration standards) are determined during the basic calibration of the method
  - Working range
  - Homogeneity of variances
  - Linearity
  - Standard deviation of residues s_y
  - Slope of the calibration function / sensitivity b
  - Process standard deviation s_{x0}
- This is described e.g. in: Funk, Dammann, Donnevert (1995): Quality Assurance in Analytical chemistry. Wiley

Basic Calibration

- If the analytical procedure needs a calibration the measurement does not lead directly to the result
- The measurement result can be converted into the analytical result using the **analytical function**

$$\hat{x} = f(\hat{y}) \quad with: \begin{matrix} \hat{x} \text{ analytical result} \\ \hat{y} \text{ measured value} \end{matrix}$$

- Basing upon the **calibration function**:

$$y = f(x) \quad with: \begin{matrix} x \text{ content of substance in the standard solution} \\ y \text{ corresponding measured value} \end{matrix}$$

Basic Calibration

- During the basic calibration the analytical method is calibrated only with standard solutions
- I.e., no sample preparation (extraction digestion etc.), only standard solutions in pure solvent

Definition the Working Range

- First step: Definition of a preliminary working range on the basis of:
  - The practical need
  - The practically feasible possibilities
    - Measurement result at the lower limit of the working range must be significantly different from the blank values
    - The required analytical precision must be reached over the whole working range
    - If we want to use a simple linear regression the residues must be homogeneous and there must a linear relationship between analyte content and measured value

Preparation of Standard Samples

- Requirements:
  - Purity, free from matrix resp. defined matrix
  - Homogeneity
  - Representativeness for real samples
    - Chemically similar compounds
    - Same oxidation state
    - etc.
  - Stability, possibilities to preserve
  - No influence by sample containers and environmental conditions

Preparation of Standard Samples

- Production of Standard Samples
  - Consider precision of balances and volume measuring equipment
  - Weighing is always more precise and should therefore be favoured
  - Avoid successive dilutions
  - Prepare 6...10 standard samples with equidistant concentration over the whole working range

Linear Calibration Function

- The regression analysis delivers the calibration function y = a + bx
- Slope (Sensitivity)

$$b = \frac{\sum[(x_i - \bar{x}) \cdot (y_i - \bar{y})]}{\sum(x_i - \bar{x})^2}$$

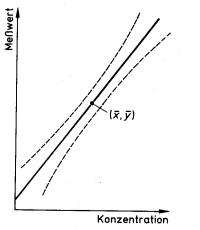
- Intercept

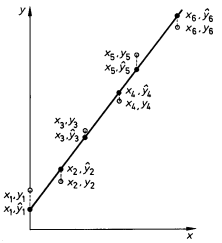
$$a = \bar{y} - b\bar{x}$$

- Standard deviation of residues (dispersion of values around the regression line)

$$s_y = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{N-2}} \quad \text{mit } \hat{y}_i = a + bx_i$$

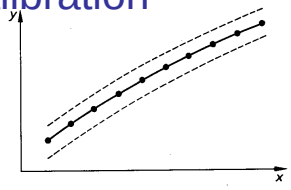
- Process standard deviation $s_{x0} = \frac{s_y}{b}$
- Process variation coefficient $V_{x0} = \frac{s_{x0}}{\bar{x}} \cdot 100\%$





Non-linear second-order calibration function y=a + bx + cx²

- Calculation is a bit more complex (see ISO 8466-2)
- Function coefficients



$$a = \left(\sum y_i - b \cdot \sum x_i - c \cdot \sum x_i^2\right) / N$$

$$b = \frac{Q_{xy} - c \cdot Q_{x^3}}{Q_{xx}}$$

$$c = \frac{Q_{xy} \cdot Q_{x^3} - Q_{x^2y} \cdot Q_{xx}}{\left(Q_{x^3}\right)^2 - Q_{xx} \cdot Q_{x^4}}$$

$$Q_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{N}$$

$$Q_{xy} = \sum \left(x_i y_i\right) - \left(\sum x_i \frac{\left(\sum y_i\right)}{N}\right)$$

$$Q_{x^3} = \sum \left(x_i^3\right) - \left(\sum x_i \frac{\left(\sum x_i^2\right)}{N}\right)$$

$$Q_{x^4} = \sum \left(x_i^4\right) - \left(\frac{\left(\sum x_i^2\right)^2}{N}\right)$$

$$Q_{x^2y} = \sum \left(x_i^2 y_i\right) - \left(\sum y_i \cdot \frac{\sum x_i^2}{N}\right)$$

Second-order calibration function

- Standard deviation of residues

$$s_y = \sqrt{\frac{\sum \left(y_i - \hat{y}_i\right)^2}{N - 3}} \quad \text{with} \quad \hat{y}_i = a + bx_i + cx_i^2$$

- Sensitivity
  - First derivation of the calibration function $E(x) = b + 2c \cdot x$
  - Resp. in the middle of the working range $E(\bar{x}) = b + 2c \cdot \bar{x}$
- Process standard deviation

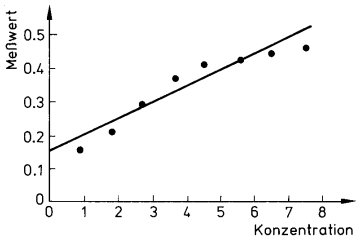
$$s_{x0} = \frac{s_y}{E(\bar{x})}$$

- Process variation coefficient

$$V_{x0} = \frac{s_{x0}}{\bar{x}} \cdot 100\%$$

Check for Linearity

- If possible always use linear calibration function, polynomial regression only in special cases
- Visual linearity test
  - Graphical display incl. Calibration line
  - If there is an obvious non-linearity refrain from a statistical test



Check for linearity

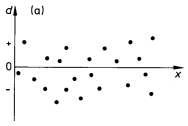
- Mandel-test
  - Calculation of the linear calibration function y=a+bx and the second-order calibration function $y=a+bx+cx^2$ including the corresponding standard deviations of residues s_{y1} (linear) and s_{y2} (non-linear)
  - Calculation of the difference of variances DS^2:

    $DS^2 = (N-2)s_{y_1}^2 - (N-3)s_{y_2}^2$ with a degree of freedom $f = 1$

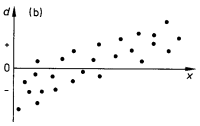
  - Check with F-test $$F_{observed} = \frac{DS^2}{s_{y_2}^2}$$
  - Compare with tabulated value $F_{critical}$ for $f_1=1$, $f_2=N-3$, P=99%
  - If $F_{observed} < F_{critical}$, then we get **no** better adjustment with the second-order calibration function
  - The calibration function then is linear

Residual Analysis

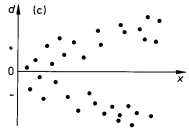
- Residues are the vertical distances of the measure values from the regression line
- Residues should be normally distributed



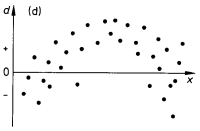
Normally distributed residues
correct model



linear trend
wrong model or calculation error



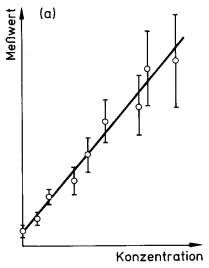
Increasing variances
inhomogeneity of variances

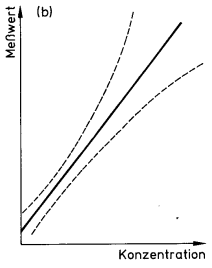


Non-linear
wrong regression function

Homogeneity of Variances

- Linear regression assumes constant (homogeneous) imprecision (variance of values) over the whole working range
- Inhomogeneous variances:





- Inhomogeneity of variances not only leads to a higher imprecision, but due to a possible change in the slope of the regression line to a higher bias

Universität Stuttgart

Check for Homogeneity of Variances

- Measure highest and lowest standard sample ten times each
- Calculate variances for both data sets

$$s_i^2 = \frac{\sum (y_{ij} - \bar{y}_i)^2}{n_i - 1}$$

- Check with the F-test

$$F_{observed} = \frac{s_N^2}{s_1^2}$$

- If $F_{observed} > F_{critical}$, then the variances are **not** homogeneous
- Possible consequences:
  - Reduce working range
  - Weighted regression
  - Multiple-Curve-Fitting

Universität Stuttgart

Outlier test

- Calibration data have to be free from outliers
- The test for outliers assumes the correctness of the chosen regression approach
- From the residual analysis potential outliers can be identified
- The residual standard deviation is calculated with all values ($s_{y,A1}$) and without the outlier-suspect value ($s_{y,A2}$)
- The check can be made using the F-test or the t-test

Outlier Test using F-test

- The residual standard deviations are checked for significant differences
- Calculate

$$F_{observed} = \frac{(N_{A1}-2)s_{y_{A1}}^2 - (N_{A2}-2)s_{y_{A2}}^2}{s_{y_{A2}}^2}$$

- And compare with the critical value from the table with $f_1=1$, $f_2=N_{A2}-2$, $P=95\%$
- If $F_{observed} < F_{critical}$, no outlier

Outlier T using t-test

- Calculate prediction interval of the regression line <u>without</u> outlier

$$PI(\hat{y}_O) = \hat{y}_O \pm t \cdot s_{y_{A2}} \cdot \sqrt{1+\frac{1}{N_A}+\frac{(x_O-\bar{x})^2}{\sum x_i^2 - \frac{1}{N_A}(\sum x_i)^2}}$$

$$= a_2 + b_2 \cdot x_O \pm t \cdot s_{y_{A2}} \cdot \sqrt{1+\frac{1}{N_A}+\frac{(x_O-\bar{x})^2}{\sum x_i^2 - \frac{1}{N_A}(\sum x_i)^2}}$$

t = tabulated value of t - distribution ($P = 95\%, f = N_A\text{-}2$)
N_A = N-1
x_O = concentration of eliminated outlier
$\bar{x}$ = mean of all x_i (without x_A)

- If the potential outlier lies within the prediction interval, it has to be included again in the data set
- **If a value is statistically proven to be an outlier, then the cause for the outlying value <u>must</u> be searched and eliminated. Then repeat the <u>complete</u> calibration.**

Limit of Detection, Limit of Quantitation

- With these values the lower limit of the working range can be characterised
- There are numerous different definitions and calculation methods

Glossary – Limit of Detection (lod)

- The limit of detection is the lowest amount (of substance) of the analyte in a sample, that can be detected, but not necessarily quantified as an exact value
- Statistically
  - If this value is exceeded, we recognise with an error probability of α, that the amount of the analyte is higher than that in a blank sample

Estimate for the Limit of Detection

- Coarse Estimate:
  LoD = B+3S_0 or 0+3S_0
  (for fortified samples; typically, three times the noise level)
  - B=Blank
  - S_0=standard deviation of 10 measurements
- Alternative method (mostly used in chromatography): Signal-Noise-Ratio=3

Expression of the LoD

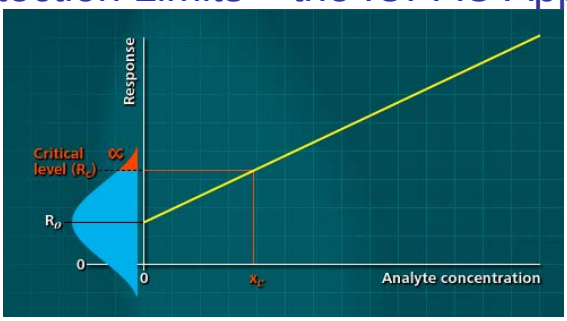
- Analyze
  - 10 independent sample blanks and get the mean sample blank value (B) **or**
  - 10 independent sample blanks fortified at lowest acceptable concentration.
- Express LoD as the analyte concentration corresponding to
  - B+3s or
  - 0+3s

  (s being the sample standard deviation).

Detection Limits – the IUPAC Approach

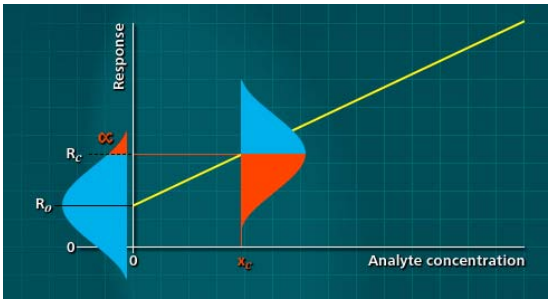
- Concepts related to the detection limit are based on two theoretical probabilities:
  - α - the probability of obtaining an analytical response above a certain limit
  - β - the probability of obtaining an analytical response below the critical limit when the analyte is present at some higher concentration

Detection Limits – the IUPAC Approach



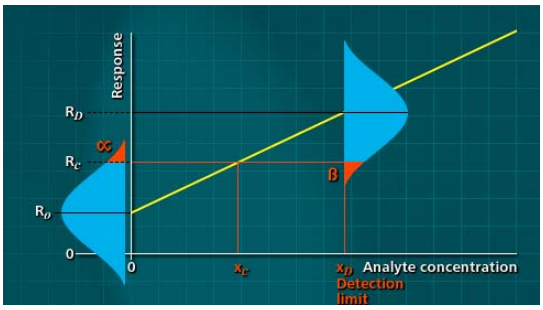
- Distribution of results at a real concentration of zero
- The results are centred around R_0 with a standard deviation of σ
- There is a probability of α, that a result would exceed the *critical value* R_c for the signal, or the corresponding value x_c, if the analyte were really absent (false positive result)
- This level therefore is a decision limit, at which we can say the analyte is present with a level of confidence (1-α)
- For 95% confidence (α=0.05): $R_c=R_0+1.65\sigma$)

Detection Limits – the IUPAC Approach



- If the analyte is really present at concentration x_c, half the results would be detected below R_c: i.e. they would be not detected (false negative)
- There must be some higher concentration where the possibility of „not detected“ is low

Detection Limits – the IUPAC Approach



- At at true concentration of x_D there is a probability of β of seeing a response below R_c
- By letting β become sufficiently small we reduce the „not detected“ to an acceptable level
- The corresponding concentration is the detection limit
- Usually both α and β are set to give 95% confidence, leading to $R_D=R_0+3.3\sigma$
- By using the calibration curve to convert to concentrations, we see that the detection limit is $x_D=3\sigma_x$

Glossary – Limit of Quantitation (loq)

- The limit of determination is the lowest amount (of substance) of the analyte in a sample, that can be quantified with a sufficient accuracy

Estimate for the Limit of Quantitation

- Coarse Estimate:
  LoQ = $B+10S_0$ or $0+10S_0$
  - B=Blank
  - S_0=standard deviation of 10 measurements
- Alternative method (mostly used in chromatography):
  Signal-Noise-Ratio=10

Glossary - Selectivity

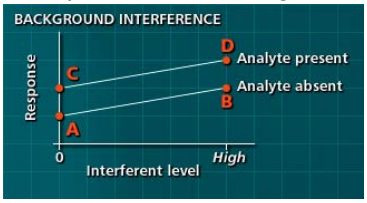
- Selectivity is a measure that shows to what extend a method can be used to determine certain analytes in mixtures and matrices without interferences caused by other components which have a similar behaviour
- IUPAC recommends, to use the similar term “specificity” not any more
- During a validation it is necessary to check if the method has a sufficient selectivity for the intended purpose

Lack of Selectivity - Interferences

- Interferences can be detected by adding a potential interfering substance to a „normal“ matrix and analysing with and without the interfering substance
- A suitable procedure uses 4 solutions:

Analyte	Content of interfering substance zero	high
without	A	B
with	C	D

- One possible interfering influence is the background interference:



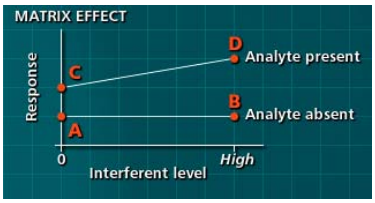
A, B, C, D are the measurement results

If A≠B, then there is a background interference (shift of the measurement results).

An interfering substance, different from the analyte, produces a measurement value. This happens also in the absence of the analyte.

Lack of Selectivity - Interferences

- Another interference is he matrix effect



A, B, C, D again are measurement results

If (D-C)≠(B-A), then there is a matrix effect.

This type of interference changes the slope of the calibration function (i.e. change of sensitivity)

To check, if the differences are significant, repeat the measurements of A, B, C and D and test with 2-sample-t-test

Matrix effects can also be uncovered with the standard addition procedure

Glossary – Robustness/Ruggedness

- Robustness/Ruggedness is the insensitiveness of a method to small deviation in the experimental procedure described in the method
- ‚Soft' methods, sensitive to slight deviations, are unlikely to perform well in interlaboratory comparisons
- Possible
- Potential topics to be checked in a robustness study:
  - Volumes
  - Concentrations of reagents
  - Duration of heating and extraction procedures
  - Temperatures
  - pH
  - ...

Ruggedness Test

- There is an economical design (a fractional factorial design) for testing for ruggedness.
- Up to n factors can be tested simultaneously in an experiment requiring $2^k > n > 2^{k-1}$ runs of the experiment (k is an integer)
- Thus, for an experiment involving seven factors, eight runs are required.
- Each run contains a combination of factors at perturbed levels
- The perturbed levels should be either higher (+) or lower (-) than the levels specified in the method procedure.
- The degree of perturbation should represent the maximum excursion from the specified level likely to be encountered in normal practice, e.g. if the method requires heating to 100°C for 1 hour, reasonable perturbed times might be 50 and 70 minutes

Layout of a Ruggedness Test - Example

Factor	Run number							
	1	2	3	4	5	6	7	8
Sample weigth	+	+	+	+	-	-	-	-
Conc. reagent 1	+	+	-	-	+	+	-	-
Conc. reagent 2	+	-	+	-	+	-	+	-
Total volume	-	-	+	+	+	+	-	-
Time of heating	-	+	-	+	+	-	+	-
Reaction temperature	-	+	+	-	-	+	+	-
pH of solution	+	-	-	+	-	+	+	-
Analyte found	68	59	67	64	64	66	60	70
+ = positive perturbation - = negative perturbation								

- The effect of a factor is given by:
  (mean of runs with +-perturbed)-(mean of runs with --perturbed)
- In the example the effect of the time of heating is
  (59+64+64+60)/4 – (68+67+66+70)/4=-6
- The method is easy to interpret so long as only one or two of the factors are sensitive

Trueness

- Trueness of the analytical method can be examined by several methods
  - Analysis of a reference material
  - Analysis of a certified reference material
  - Analysis of an in-house reference material
  - Interlaboratory comparison
  - Reference methods
  - Recovery experiments

Precision

- Precision of the method can be examined by repeated measurement
- Repetitions can be made under
  - Repeatability conditions
  - Between-batch-conditions
- Precision check is often done in combination with control charts

Standard Addition Procedure - What's that?

- Standard addition procedure is a calibration in the real sample by stepwise addition of a defined amount of analyte

Standard Addition Procedure – in which Cases?

- If differences in the composition of the matrix have a strong influence on the trueness of the result (matrix effects)
- If no matrix-matched calibration standards are available
- If only a few samples have to be analysed

Standard Addition Procedure - Preconditions

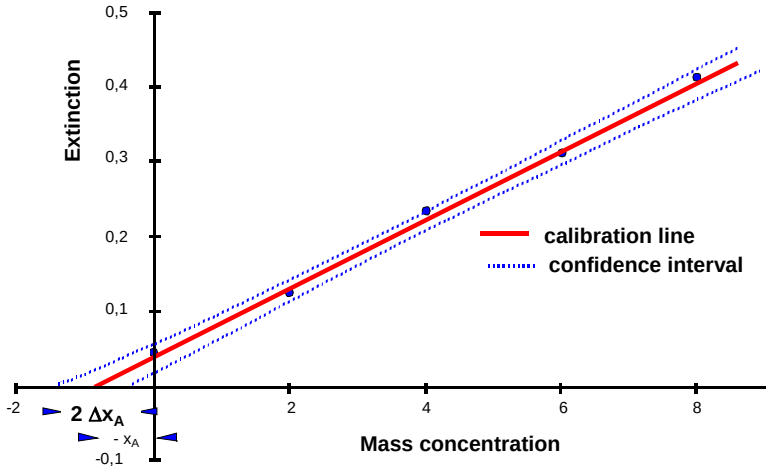
- Blank and background corrected measurement values y_1
- Linear correlation between measurand y and content x
- Standard deviation of residues $s_{y,x}$ independent from y (Homogeneity of variances)
- Homogeneous sub-sampling must be possible
- Precise addition of analyte must be possible

Standard Addition Procedure - Procedure

- Division in n equal sub-samples
- Addition of known increasing portions z_i of the analyte to n-1 sub-samples in equidistant steps and normalisation of all sub-samples
  $\rightarrow$ pairs o variates $(x_{z_1,j};y_{1,j})(x_{z_2,j};y_{2,j})\ldots(x_{z_n,n_a};y_{n,n_a})$
- Linea Regression
  $\rightarrow\ y = a + b \cdot x$
- Extrapolation to the intercept point with the abscissa delivers the sought content
  $\rightarrow\ x_A = -x_{(y=0)} = -a/b$

x = content quantity
y_i = blank and background corrected measurement
n_a = number of measurements per sub-sample
j = index for repeated measurements
x_A = content of the analysed sample

Standard Addition Procedure - Graphical Display



Standard Addition Procedure - Uncertainty

- The uncertainty of the calculated value x_A can be quantified with the half width of the confidence interval of the result

$$\Delta x_A = \frac{t_{f,\alpha} \cdot s_{y,x}}{b} \cdot \sqrt{\frac{1}{n \cdot n_A} + \frac{\left[\bar{x}_z - (-x_A)\right]^2}{\sum (x_{z_i} - \bar{x}_z)^2}}$$

x_A = half width of the confidence interval of x_A
$t_{f,\alpha}$ = two - sided quantile of the t - distribution with probability α
$\bar{x}_z$ = arithmetic mean of x_{z_i}
$\sum (x_{z_i} - \bar{x}_z)^2$ = sum of all squares of deviations of all x_{z_i}