Uncertainty estimation in chemometrics

The example of multivariate (and multiway) prediction uncertainty

Klaas Faber

Chemometry Consultancy
www.chemometry.com

Univariate *vs.* multivariate (and multiway) practice

Personal communication with Prof. Massart:

"Univariate calibration is nice and clean,
multivariate calibration is wild west."

The chemometric domain is characterized by a general lack of sound statistical underpinning: confidence and prediction intervals, hypothesis tests, etc.

A general note on the reporting of results

P. De Bièvre, Editorial

Measurement results without statements of reliability (uncertainty) should not be taken seriously

Accreditation and Quality Assurance, **2** (1997) 269

"So, a result without reliability (uncertainty) statement cannot be published or communicated because it is not (yet) a result. I am appealing to my colleagues of all analytical journals not to accept papers anymore which do not respect this simple logic."

Historical explanation

Multivariate (and multiway) calibration methods such as PLS have often been introduced without the associated error analysis.

A notable exception:

M. Linder and R. Sundberg

Precision of prediction in second-order calibration, with focus on bilinear regression methods

Journal of Chemometrics, **16** (2002) 12-27

Journal of Chemometrics 2003 Kowalski prize for best theoretical paper

Marie Linder

Rolf Sundberg

The relevance of error analysis for PLS was recognized at an early stage, but progress turned out to be slow

S. Wold, A. Ruhe, H. Wold and W.J. Dunn III

The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses

SIAM Journal on Scientific Computing, **5** (1984) 735-743

"In a forthcoming report we will also investigate statistical and numerical aspects on the PLS method, and show how the 8 theorems given by Marquardt (1970) for ridge and principal component regression translate into this situation."

A general note on error analysis

Calibration models are approximations at best so that the following quote applies:

"There is little profit in approximations which are good but not known to be good."

B.N. Parlett

The symmetric eigenvalue problem (1980)

Let's restrict the discussion to prediction uncertainty.

Univariate model: <u>theoretical expression</u> that <u>explicitly</u> accounts for all relevant sources of uncertainty, i.e. for <u>both</u> x- and y-variables

Model estimate and
95%-prediction bands
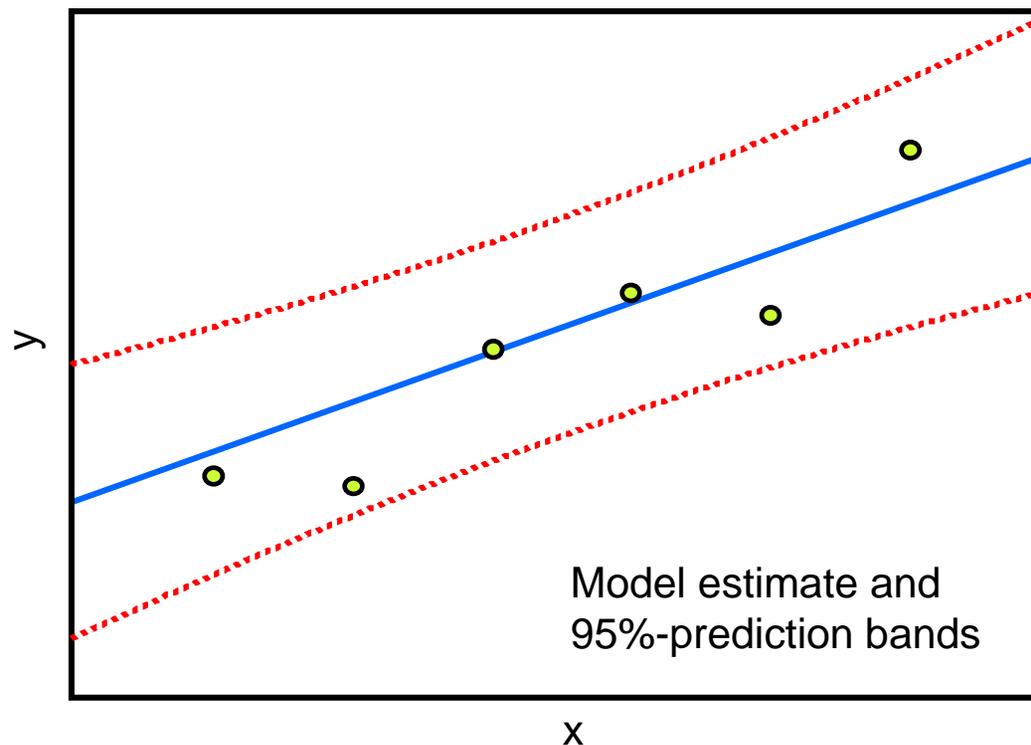
Multivariate model: empirical validation that implicitly accounts for all error sources, plus a spurious one

Root mean squared error of prediction (RMSEP) for independent test set of *N* objects:

$$\text{RMSEP} = \sqrt{N^{-1}\sum_{n=1}^{N}\left(\hat{y}_n - y_{n,\text{ref}}\right)^2}$$

$\hat{y}_n = $ prediction for object *n*

$y_{n,\text{ref}} = $ associated reference value

Two obvious problems with test set validation (or cross-validation for that matter)

1. A crucial assumption is that the reference values are sufficiently precise; this is certainly not always true (octane rating, classical Kjeldahl) - often the prediction is even better than the reference value. The reference noise is a spurious contribution to RMSEP.

2. The result (RMSEP) is a constant measure for prediction uncertainty that cannot lead to prediction intervals with correct coverage probabilities (say 95%). Therefore, object-specific prediction uncertainties are required.

Concerning the first problem: the model is blamed for the uncertainty in the reference values

A calibration model (the 'secondary' method) is often better than the 'primary' reference method, but how to actually prove this?
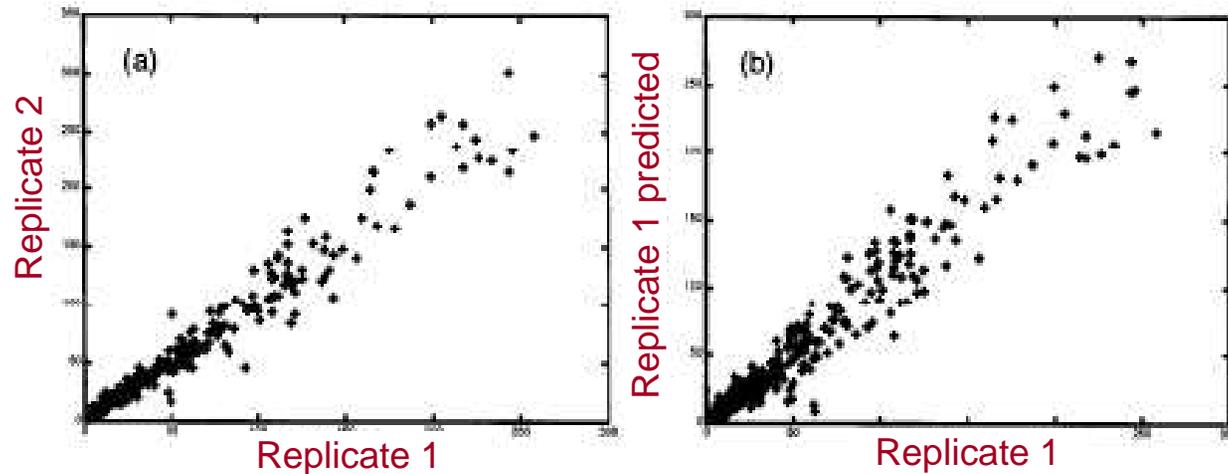
For a good discussion of this controversy, see:

R. DiFoggio

Examination of some misconceptions about near-infrared analysis

Applied Spectroscopy, **49** (1995) 67-75

Illustrative (multiway) example

- The uncertainty in the reference values is level-dependent.

- The prediction residual is mainly due to the reference error: replicate 2 is not much closer to replicate 1 than the model prediction.

R. Bro, *Chemometrics and Intelligent Laboratory Systems*, **38** (1997**)** 149-171

Concerning the second problem: benefits of an object-specific prediction uncertainty

- Construction of realistic prediction intervals, e.g. for monitoring the performance of an analysis using control samples, see ASTM standard E1655, "Standard practices for infrared, multivariate, quantitative analysis"

- RMSEP - a constant value - poorly describes the prediction uncertainty for extreme objects (extrapolation):
  - detect lower concentrations in analytical chemistry (than actually present in the training set);
  - analyte determination using standard additions
  - drug with higher activity in quantitative structure activity relationship (QSAR) modelling;
  - better product in sensory science.

Remarks concerning extrapolating an empirical model

Calibration models are data-driven, not based on first principles. Extrapolation should therefore be regarded with caution. The common advice that one should never extrapolate is, however, often too cautious because it may simply contradict the goal of modeling.

> "In some instances it may not be necessary to predict outside the region of the data: however, in other situations, predictions outside the region of the available data are of great interest."

R.D. Snee

Validation of regression models: methods and examples

Technometrics, 1**9** (1977) 415-428

Example 1: PLS model of the Fearn NIR data set

extrapolation: high uncertainty

Apparent RMSEP contains a large spurious component.
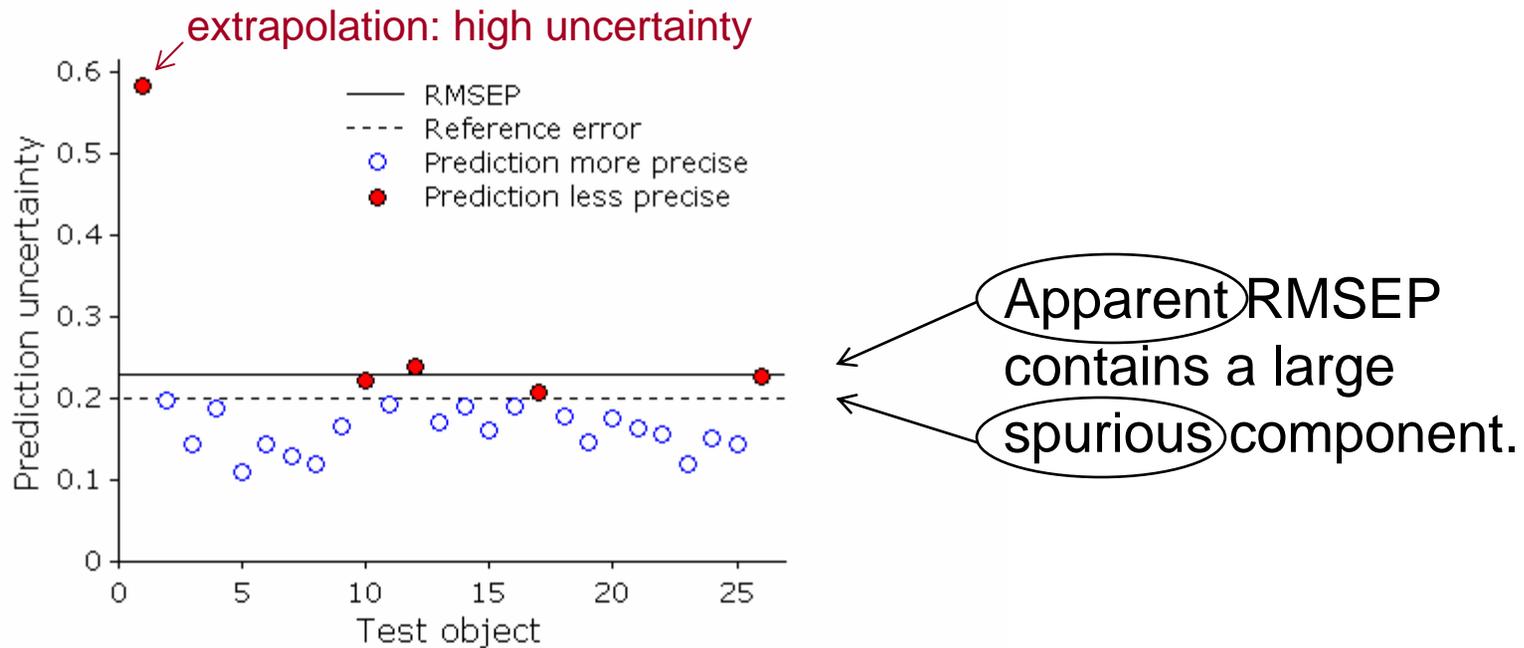
## Observations:

❑ 21 out of 26 predictions are more precise than the reference values;

❑ the apparent RMSEP grossly overestimates the actual uncertainty.

K. Faber and B.R. Kowalski, *Journal of Chemometrics*, **11** (1997) 181-238

N.M. Faber, X.-H. Song and P.K. Hopke, *Trends in Analytical Chemistry*, **22** (2003) 330-334

Example 2: comparison of <u>theory</u> with resampling and an approach implemented in the Unscrambler (PLS)

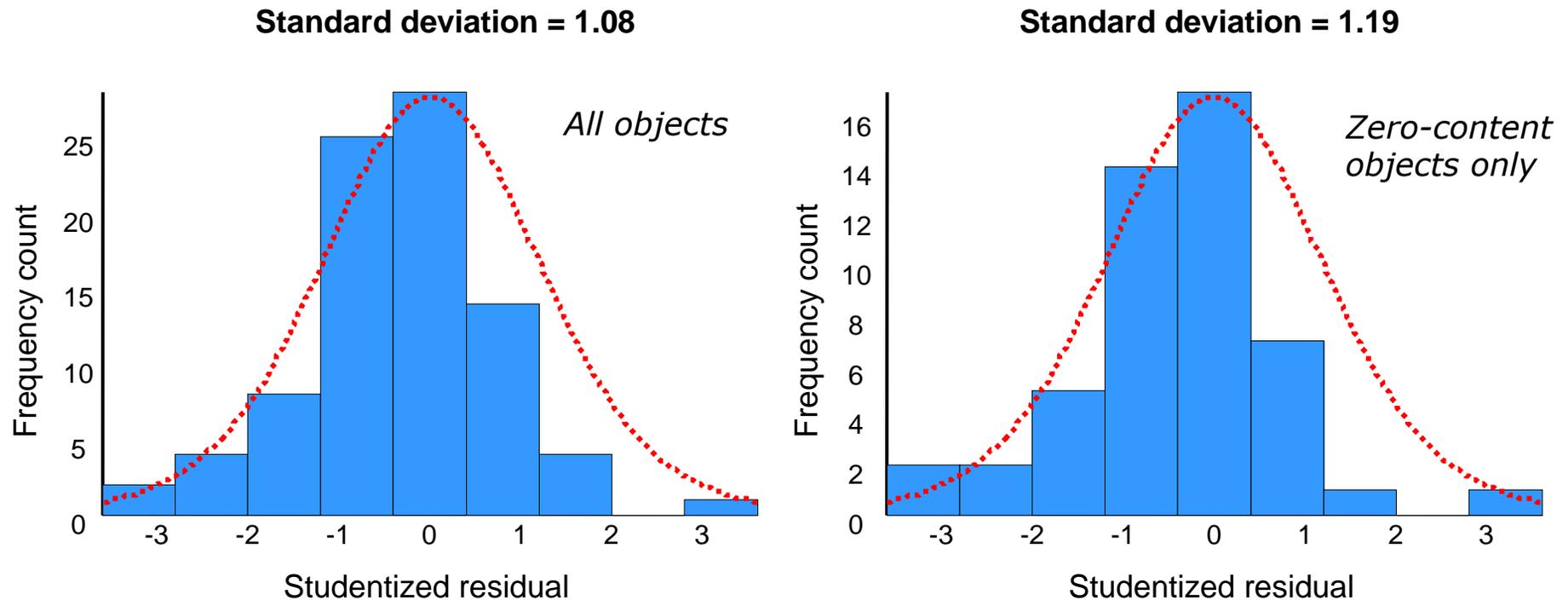extrapolation: high uncertainty is well predicted by theory
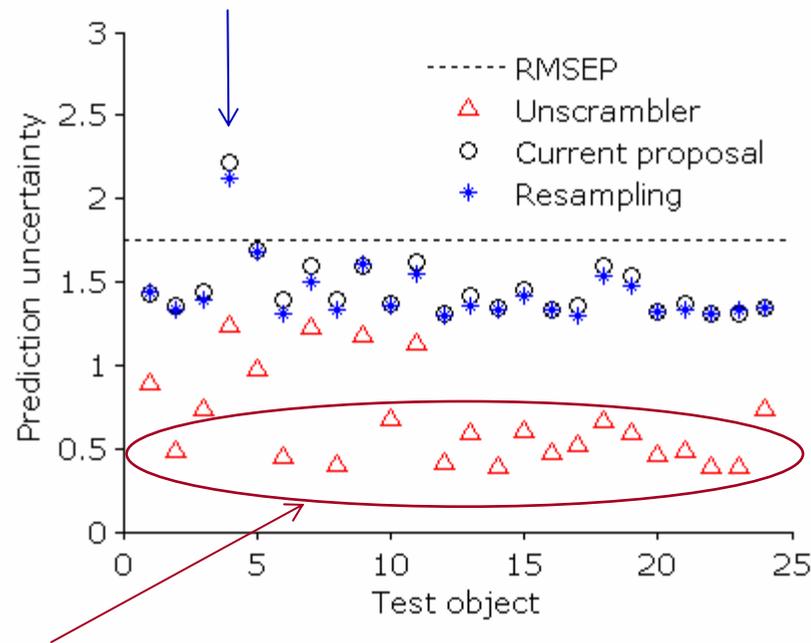
excellent agreement with resampling

poor agreement with resampling

overly optimistic: too low by a factor 3

J.A. Fernández Pierna, L. Jin, F. Wahl, N.M. Faber and D.L. Massart, *Chemometrics and Intelligent Laboratory Systems*, **65** (2003) 281-291

Example 3: multiway PLS

Standard deviation = 1.08

All objects

Standard deviation = 1.19

Zero-content objects only

Promising result: to enable the construction of intervals with proper coverage probability, the studentized prediction residuals should be distributed as Student's *t* with standard deviation 1.04.

R. Bro, A. Rinnan and N.M. Faber, *Chemometrics and Intelligent Laboratory Systems*, **75** (2005) 69-76

Extensive review article (166 references) in the official literature

A.C. Olivieri[*], N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas and H. Mark

Guidelines for calibration in analytical chemistry.

Part 3. Uncertainty estimation and figures of merit for multivariate calibration

[*] Corresponding author: aolivier@fbioyf.unr.edu.ar

IUPAC

The two main subjects

- Uncertainty estimation:
  - parameters, i.e. regression coefficients, loadings and scores
  - predictions, usually obtained from regression coefficients

This enables a better assessment of uncertainty than found in current practice.

- Analytical figures of merit:
  - limit-of-detection, sensitivity, selectivity, signal-to-noise ratio, etc.

Analytical figures of merit are performance characteristics of an analytical determination.

The key role of analytical figures of merit

The authoritative voice of the analytical community

"*Analytical Chemistry* is a peer-reviewed research journal that explores the latest concepts in analytical measurements and the best new ways to increase accuracy, selectivity, sensitivity, and reproducibility."

General remarks concerning analytical figures of merit

- They show how characteristics of the predictor data (e.g. spectra) affect the uncertainty in model results.

- They can be used to select a method or optimize a method that is already in use to meet specific needs.

- The terminology is borrowed from analytical chemistry, but the applicability is much wider.

Concluding remarks (1)

Detailed uncertainty estimates can be useful for:

- construction of confidence and prediction intervals
- limit of detection (LOD) estimation
- enhancing plots (e.g. PC scores)
- etc.

Concluding remarks (2)

- Unfortunately, commercial software lags behind or, which is even worse, has incorrect methodology implemented. A good example of the latter is the Unscrambler (e.g. errors in derivations).

- Conjecture: the not-invented-here syndrome is responsible for scientific progress not reaching end-users.

- Recommendations:
  - check the references and downloads on www.chemometry.com,
  - perhaps take a course, and
  - try it out for PCA, MLR, PCR, PLS, PARAFAC, BLLS, GRAM, …

Acknowledgements, in alphabetical order

- R. Boqué
- R. Bro
- D.L. Duewer
- J.A. Fernández Pierna
- J. Ferré
- P.K. Hopke
- J.H. Kalivas

- B.R. Kowalski
- A. Lorber
- D.L. Massart
- A.C. Olivieri
- Å. Rinnan
- F.X. Rius
- X.-H. Song